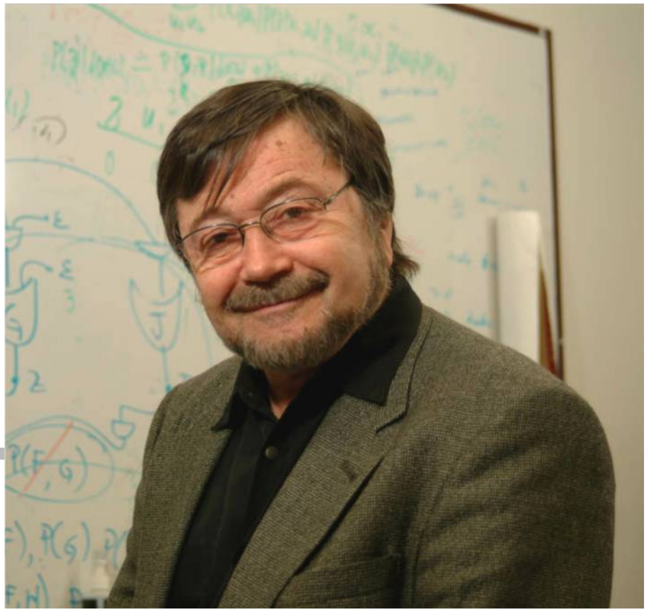
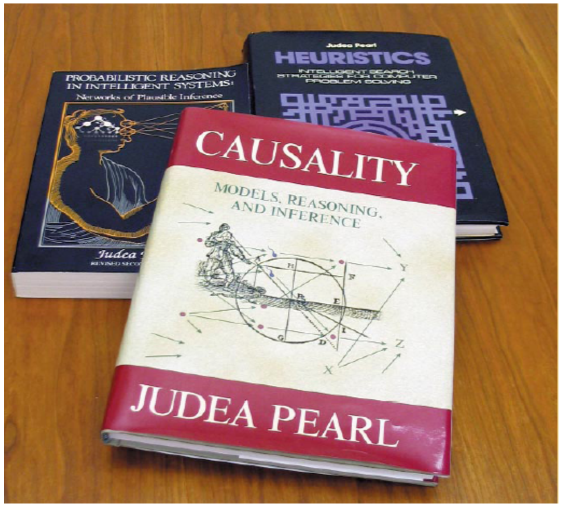
Judea Pearl: Turing Award, 2011



Rina Dechter
Donald Bren School of Computer Science
University of California, Irvine, USA



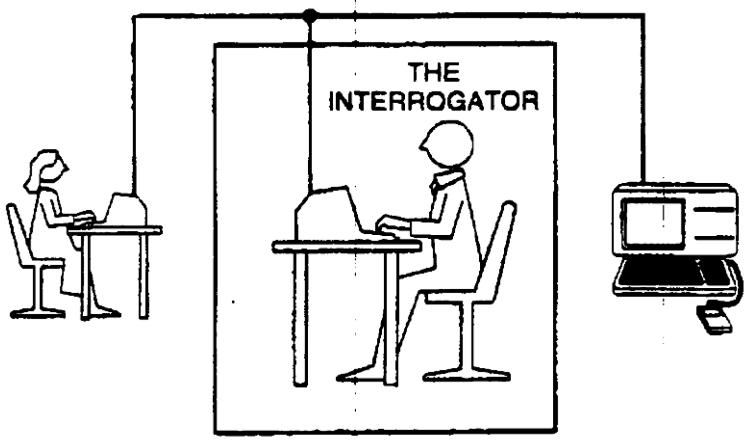
Can Machines Think?

Alan M. Turing (1912-1954)

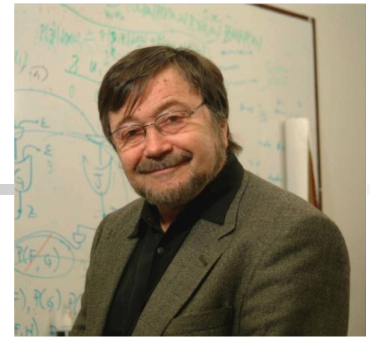
- The Turing Test
  "Computing Machinery
  and Intelligence" (1950)

- Turing answer: "Yes. If it acts like it thinks"

- If it can answer questions (about a story, solve a puzzle...)



Some Biographical Notes



- Born 1936 in Bnei-Brak
- Technion 1960, Electrical Engineering
- 1961: in Electronics (Newark College)
- 1965: Master in Physics (Rutgers)
- 1965: Phd in Electrical Engineering (Polytechnic Institute of Brooklyn and Phd (Brookleen Polytechnic Institute).
- Until 1969 at RCA labs.
- Joined UCLA at 1969

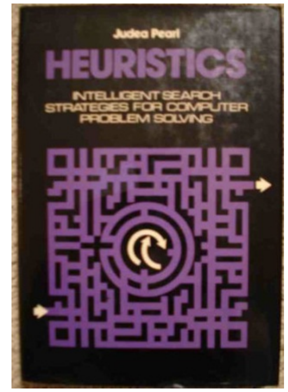
Pearl Turing Award , 2011

- For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning (Citation, Turing award 2011)
- He invented *Bayesian networks*, a mathematical formalism for defining complex probability models
- as well as the principal algorithms used for inference in these models.
- This work revolutionized the field of artificial intelligence and became an important tool for many other branches of engineering and the natural sciences.
- He later created a mathematical framework for causal inference that has had significant impact in the social sciences.

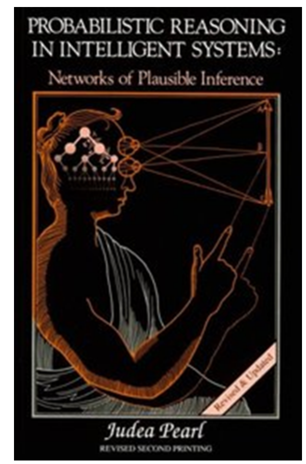
(ACM, 2011)

Pearl’s Main Contributions

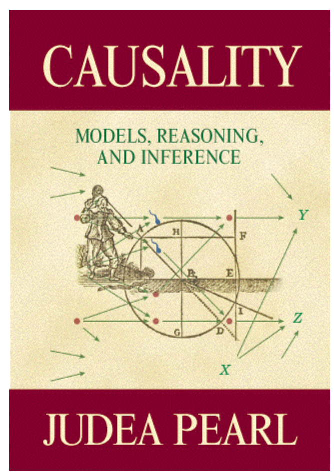
- Heuristic Search (1984)



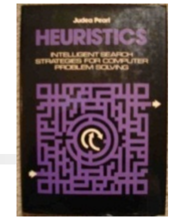
- Probabilistic Reasoning (1988)



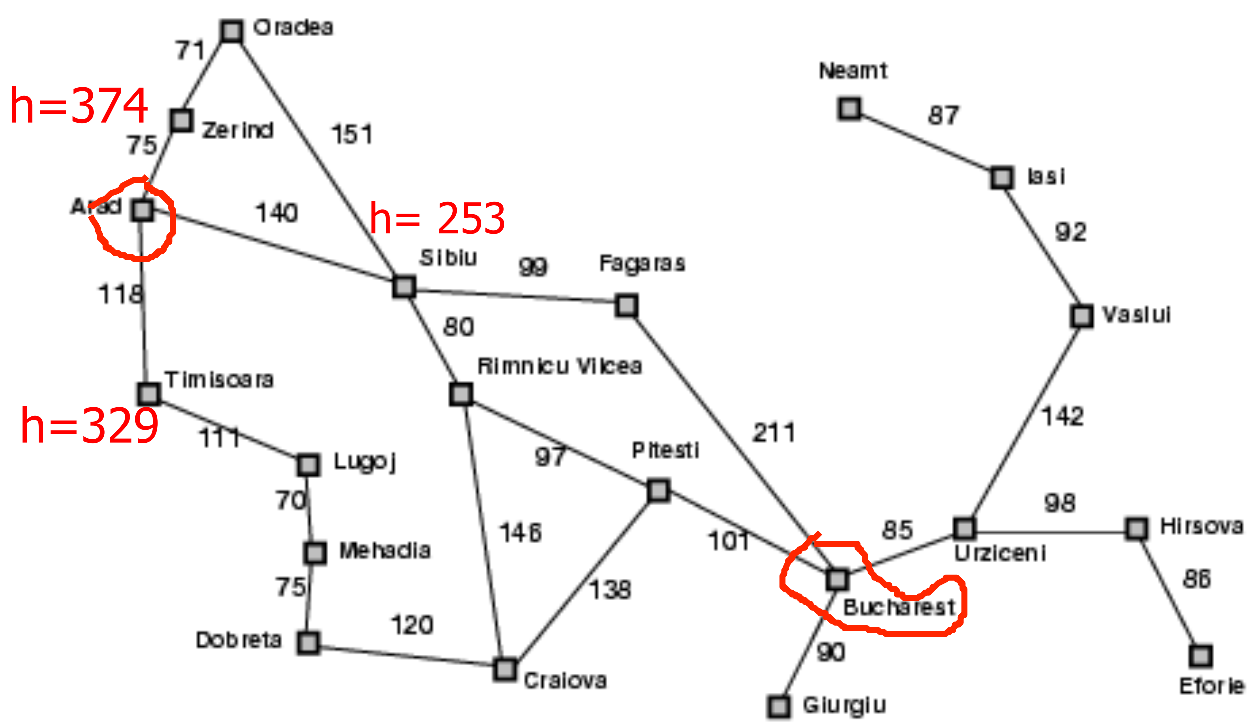
- Causality (2000)



Heuristic Search



- State-Space Search: every problem is like search of a map
- A problem solving robot finds a path in a state-space graph from start state to goal state, using heuristics

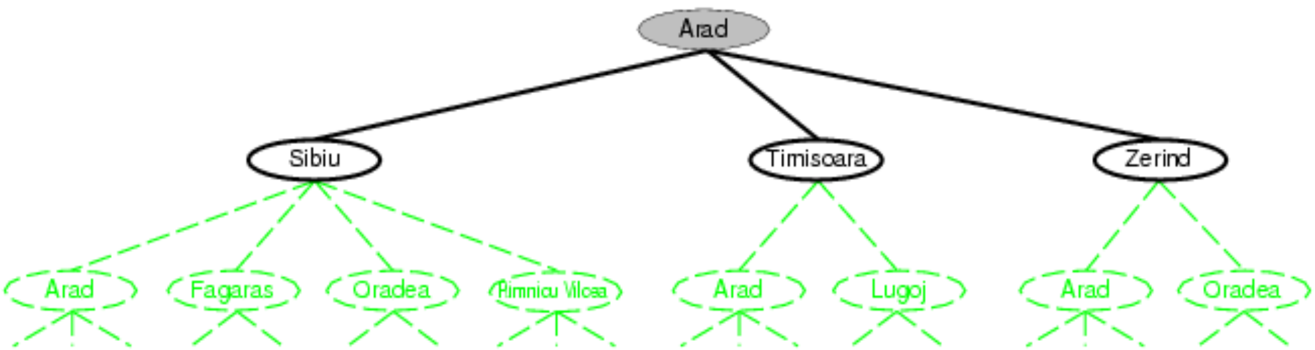


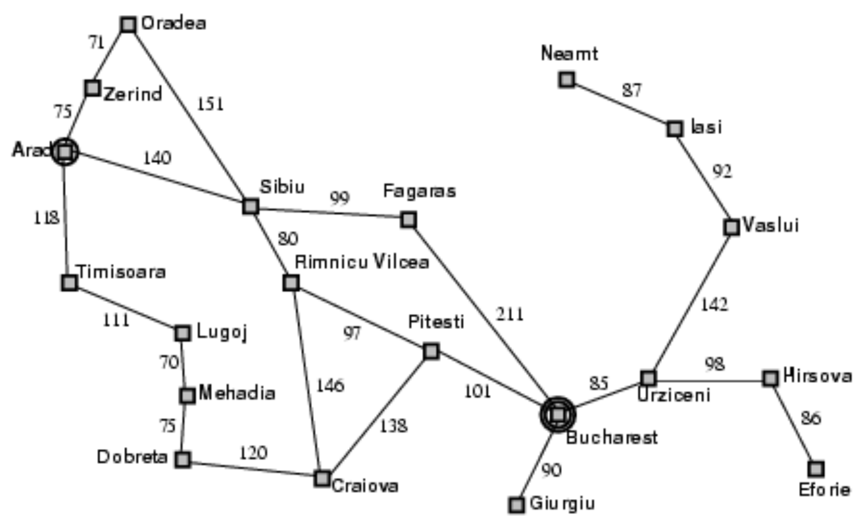
Straight–line distance to Bucharest

Arad	366
Bucharest	0
Craiova	160
Dobreta	242
Eforie	161
Fagaras	176
Giurgiu	77
Hirsova	151
Iasi	226
Lugoj	244
Mehadia	241
Neamt	234
Oradea	380
Pitesti	10
Rimnicu Vilcea	193
Sibiu	253
Timisoara	329
Urziceni	80
Vaslui	199
Zerind	374

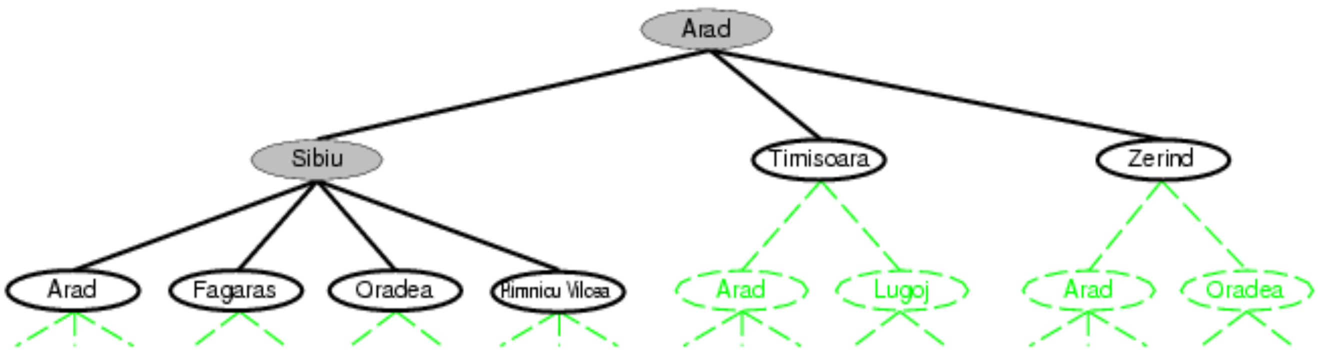
Heuristic = air distance

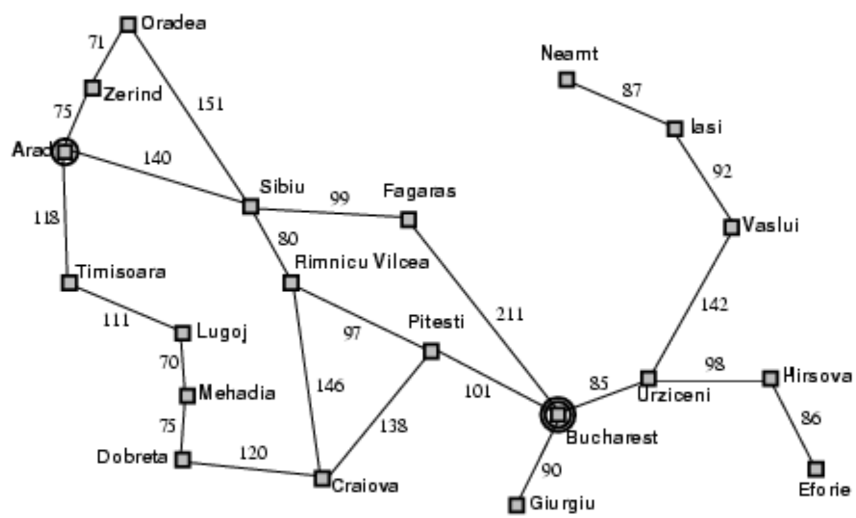
State Space for Path Finding in a Map



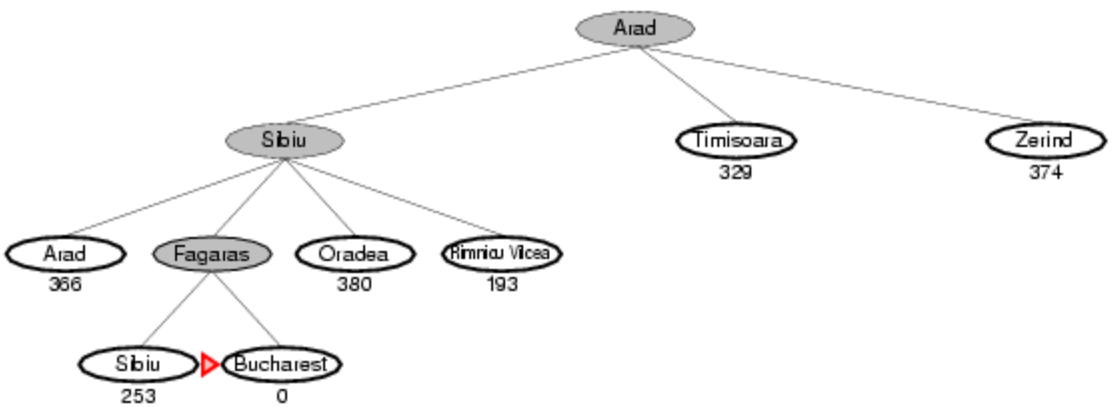


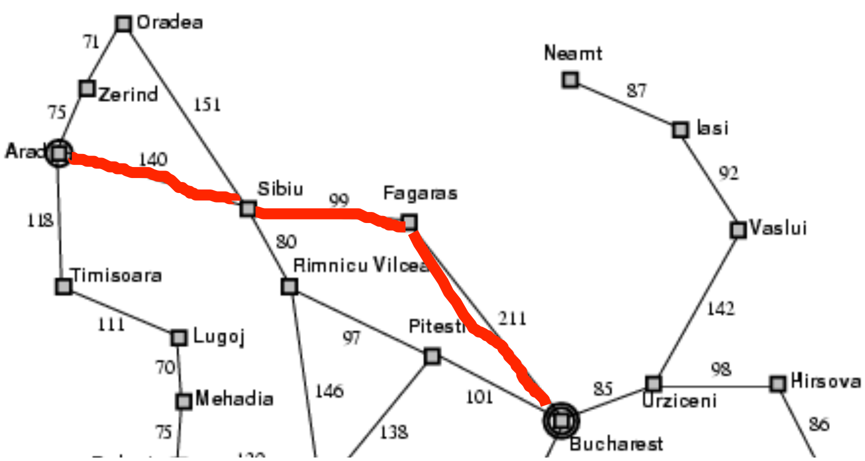
State Space for Path Finding in a Map





Greedy Search Example





The Sliding Tile (8 Puzzle) Problem

2	8	3
1	6	4
7		5

1	2	3
8		4
7	6	5

Start and Goal Configurations for the Eight–Puzzle

Up
Down
Left
Right

State Space of the 8 Puzzle Problem

8-puzzle: 181,440 states
15-puzzle: 1.3 trilion
24-puzzle: 10^25

Search space exponential

Use Heuristics as people do

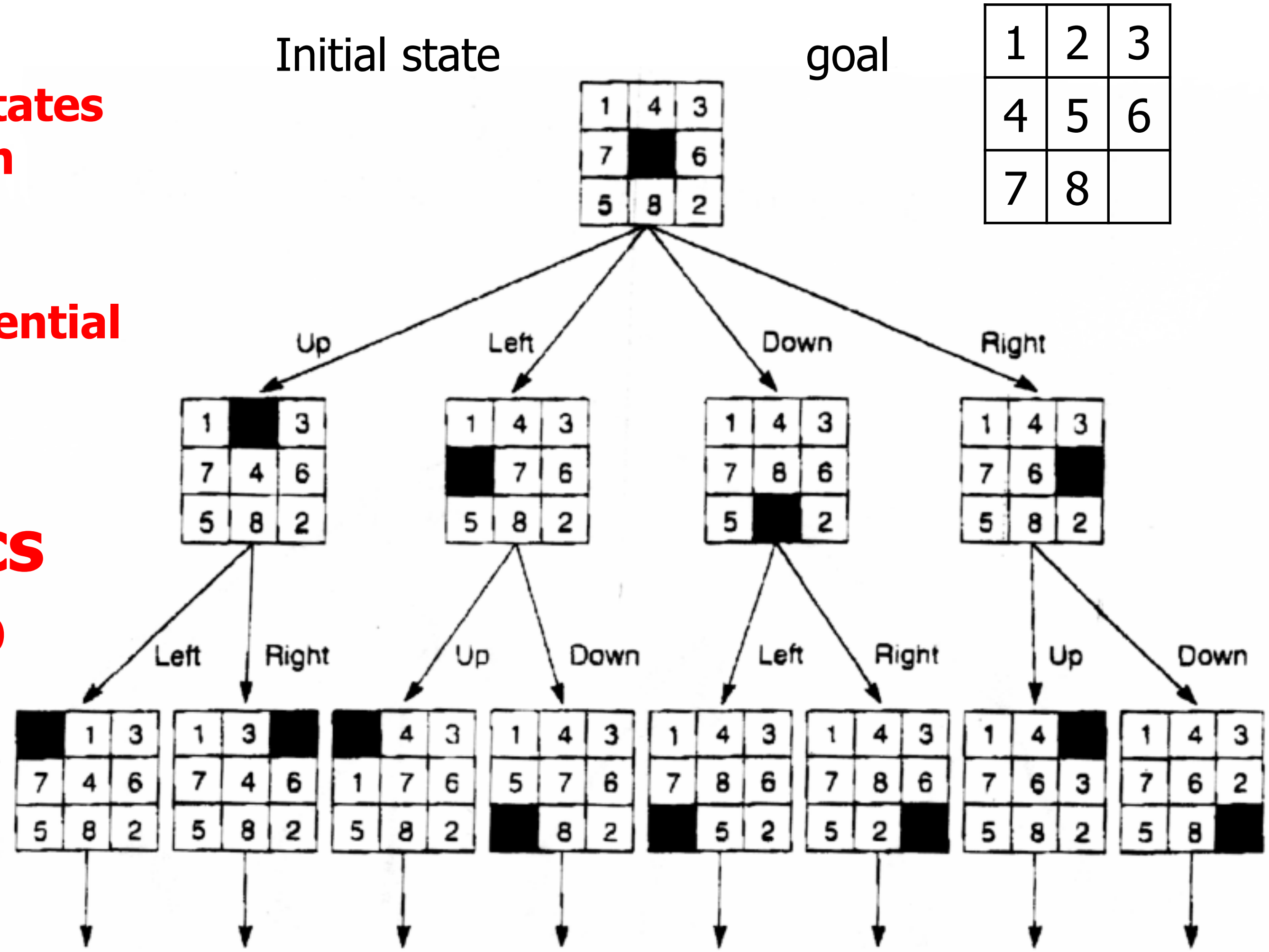


Figure 3.6 State space of the 8-puzzle generated by "move blank" operations.

State Space of the 8 Puzzle Problem

1	2	3
4	5	6
7	8	

h1 = number of misplaced tiles

h2 = Manhattan distance

Use Heuristics as people do

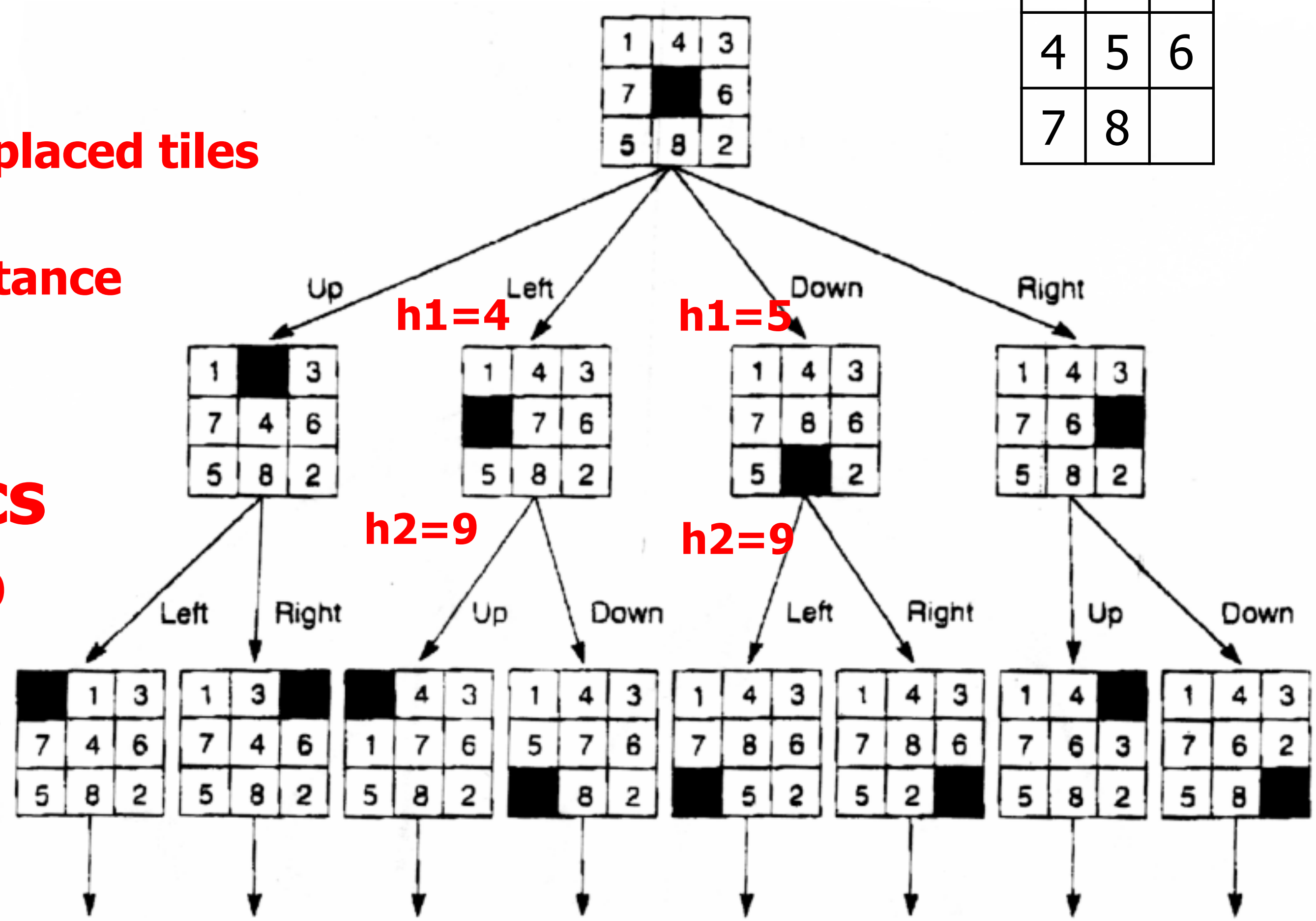
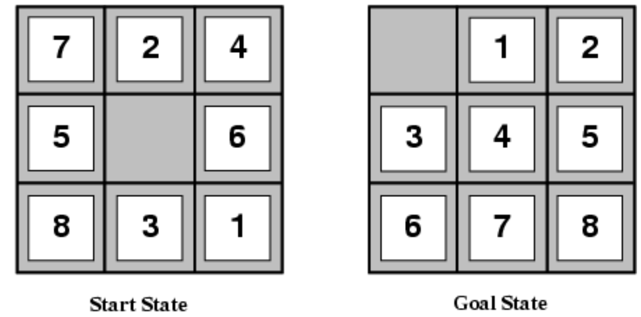


Figure 3.6 State space of the 8-puzzle generated by "move blank" operations.

What are Heuristics

- **Rule of thumb, intuition**
- **A quick way to estimate how close we are to the goal. How close is a state to the goal..**
- Pearl: “the ever-amazing observation of how much people can accomplish with that simplistic, unreliable information source known as *intuition*.”



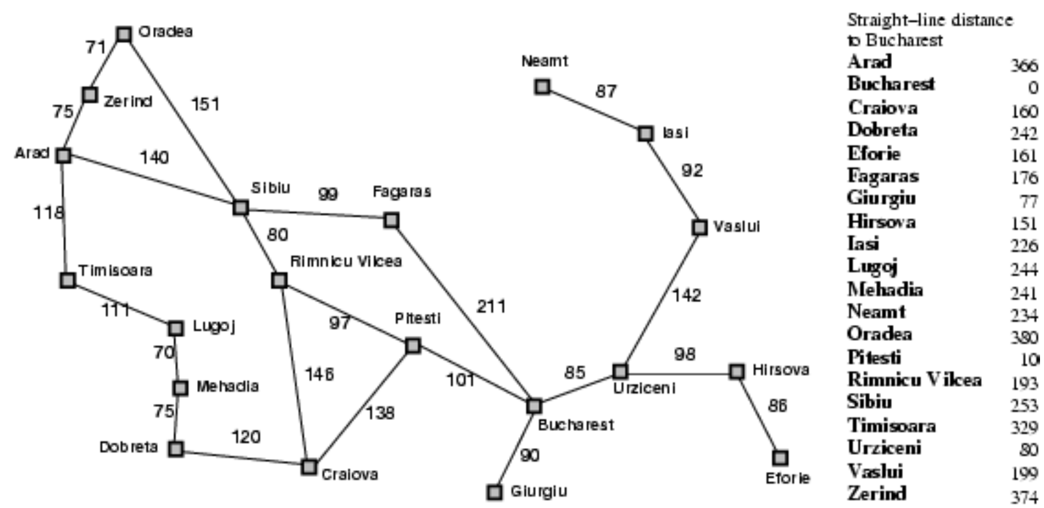
8-puzzle

- **h1(n): number of misplaced tiles**
- **h2(n): Manhattan distance**

$h_1(S) = ?$ 8

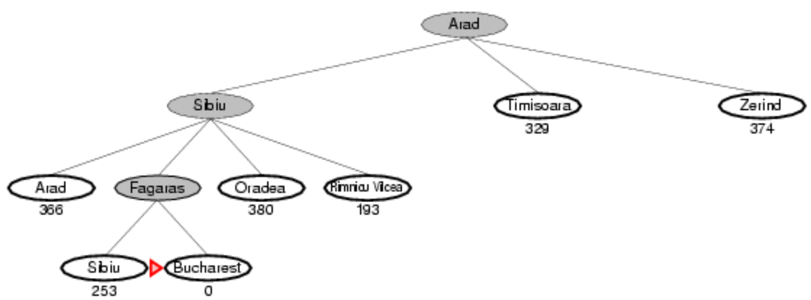
$h_2(S) = ?$ 3+1+2+2+2+3+3+2 = 18

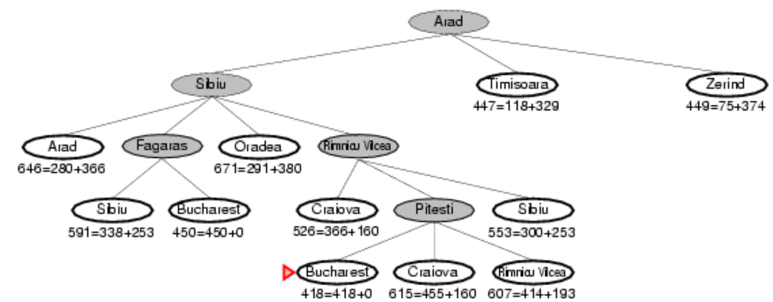
- Path-finding on a map
  - Euclidean distance

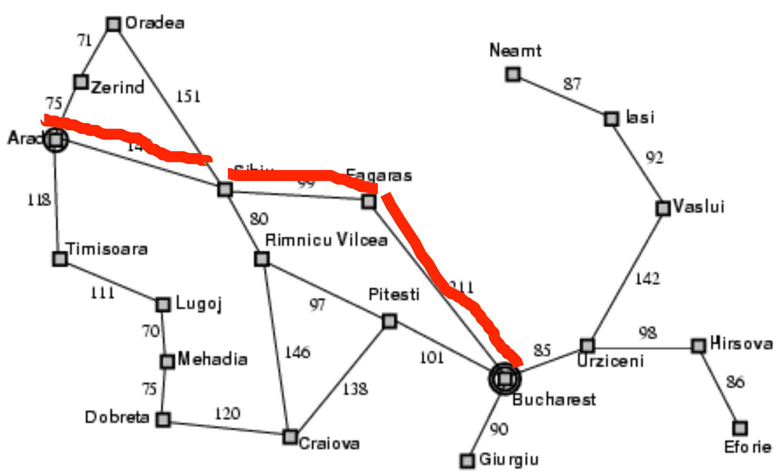


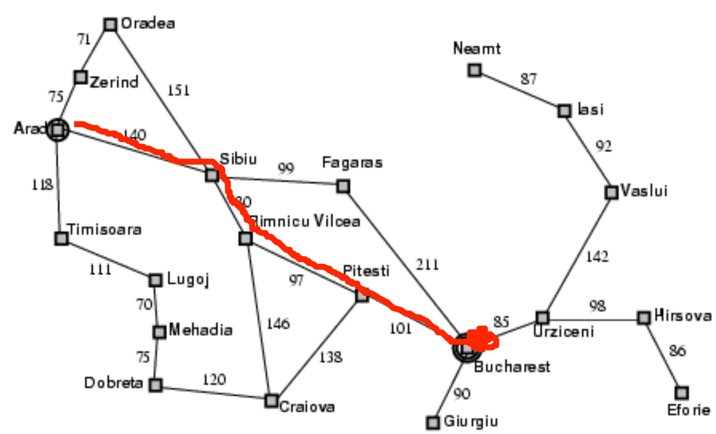
Search Examples

- **Greedy search**
- A* Search









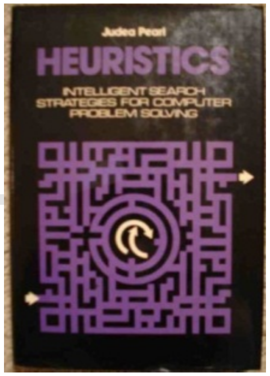
Effectiveness of A* Search Algorithm

Average number of nodes expanded

d	IDS	A*(h1)	A*(h2)
2	10	6	6
4	112	13	12
8	6384	39	25
12	364404	227	73
14	3473941	539	113
20	------------	7276	676

Average over 100 randomly generated 8-puzzle problems
h1 = number of tiles in the wrong position
h2 = sum of Manhattan distances

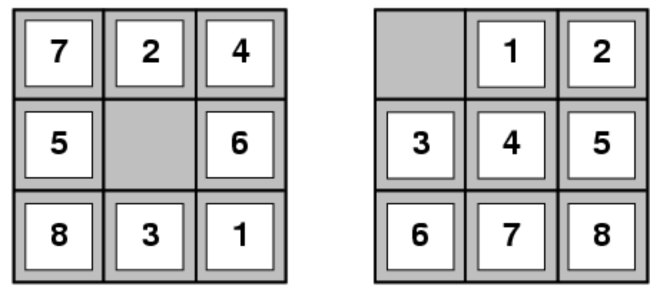
Pearl’s Research on Heuristics



- His work included many new results on traditional search algorithms and on game-playing algorithms, raising AI research to a new level of rigor and depth.
- Provided new methods for analysis: Complexity vs precision of heuristics, probabilistic approach
- Ideas of how to generate admissible heuristics automatically from relaxed problem definitions,
  - This approach that has led to dramatic advances in planning systems automatically, and problem solving in general

The Simplified Model Paradigm

Pearl 1983 (*On the discovery and generation of certain Heuristics, 1983, AI Magazine, 22-23*) : “knowledge about easy problems could serve as a heuristic in the solution of difficult problems, i.e., that it should be possible to manipulate the representation of a difficult problem until it is approximated by an easy one, solve the easy problem, and then use the solution to guide the search process in the original problem.”



Start State

Goal State

Move(x,c1,c2)

Precond list: on(x1,c1), clear(c2), adj(c1,c2)

Add-list: on(x1,c2), clear(c1)

Delete-list: on(x1,c1), clear(c2)

Move*(x,c1,c2)

Precond list: on(x1,c1), ~~clear(c2)~~, ~~adj(c1,c2)~~

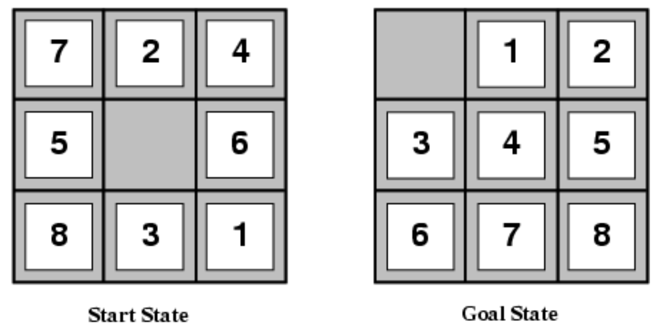
Add-list: on(x1,c2), clear(c1)

Delete-list: on(x1,c1), clear(c2)

h1 = number of misplaced tiles

The Simplified Model Paradigm

Pearl 1983 (*On the discovery and generation of certain Heuristics, 1983, AI Magazine, 22-23*) : "knowledge about easy problems could serve as a heuristic in the solution of difficult problems, i.e., that it should be possible to manipulate the representation of a difficult problem until it is approximated by an easy one, solve the easy problem, and then use the solution to guide the search process in the original problem."



Move(x,c1,c2)

Precond list: on(x1,c1), clear(c2), adj(c1,c2)

Add-list: on(x1,c2), clear(c1)

Delete-list: on(x1,c1), clear(c2)

Move(x,c1,c2)**

Precond list: on(x1,c1), ~~clear(c2)~~, adj(c1,c2)

Add-list: on(x1,c2), clear(c1)

Delete-list: on(x1,c1), clear(c2)

h2 = Mahattan distance

Summary on Heuristic Search

Pearl's early work on heuristic search – a trial-and-error method of problem-solving – propelled the evolution of AI into a mature field with sound scientific foundations.

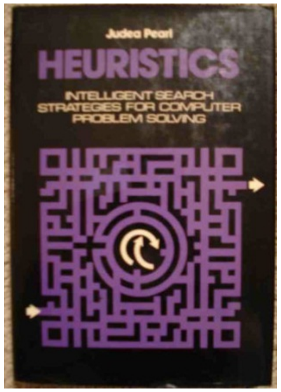
In his 1984 book *Heuristics: Intelligent Search Strategies for Computer Problem Solving*, he set a new standard where algorithms, even heuristic ones, had to be analyzed rigorously in terms of their correctness and performance.

He subsequently devised ways of programming machines to discover their own heuristics.

Heuristic search and n particular the mechanical generation of heuristics Have impact the planning field dramatically over the past 15 years

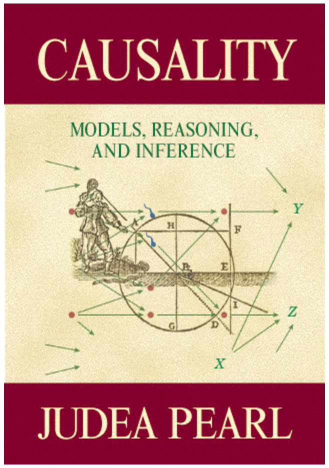
Pearl’s Main Contributions

- Heuristic Search (1984)



- Probabilistic Reasoning (1988)

- Causality (2000)



Rumelhart 1976:
Towards an Interactive Model of Reading

TOWARD AN

CENT

UNIVERSITY OF CALIFOR
LA JOLLA, CALIFORNIA 9

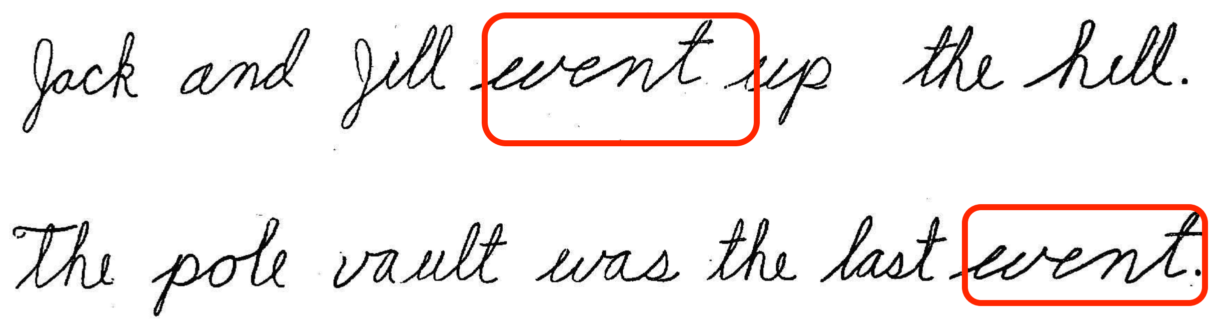


Figure 3 The dependence of letter perception of context. (After Nash-Weber, 1975.)

Rumelhart's Proposed Solution

Rumelhart (1976)
Figure 10

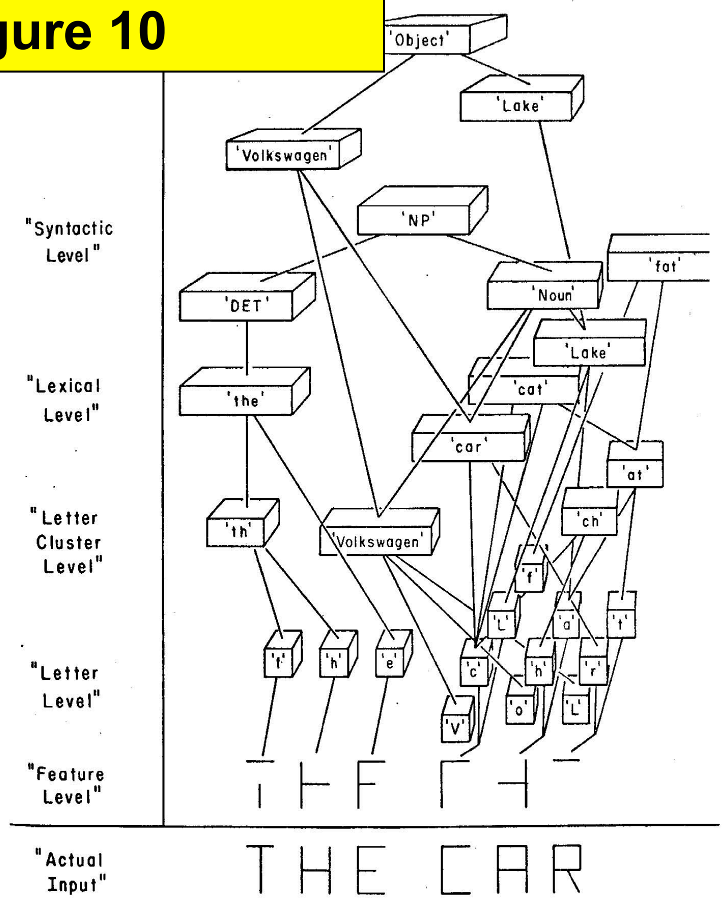


Figure 10 The message center well into the processing sequence.

and sisters of h_i. Equation (2) gives the value of the contextual strength of h_i:

$$(2) \qquad \beta_i = \begin{cases} Pr(h_i) & P_i = L_i = \phi \\ \dfrac{\Sigma s_k \cdot Pr(h_i \mid h_k)}{\nu_i} & \text{otherwise,} \end{cases}$$

where the sum is over all $h_k \epsilon P_i$ or L_i. Thus, when h_i has no parents or left sisters, its contextual strength is given by its a priori probability. Otherwise, its contextual strength is given by the sum, over all of its left sisters and parents of the strength of the left sister or parent, h_k, times the conditional probability of the hypothesis given h_k. This sum is then

Pearl: so we have a combination of a top down and a bottom up modes of reasoning which somehow coordinate their actions resulting in a friendly handshaking."

Pearl 1982:
Reverend Bayes on Inference Engines

REVEREND BAYES ON INFERENCE ENGINES: A DISTRIBUTED HIERARCHICAL APPROACH(*)(**)

Judea Pearl
Cognitive Systems Laboratory
School of Engineering and Applied Science
University of California, Los Angeles
90024

ABSTRACT

This paper presents generalizations of Bayes likelihood-ratio updating rule which facilitate an asynchronous propagation of the impacts of new beliefs and/or new evidence in hierarchically organized inference structures with multi-hypotheses variables. The computational scheme proposed specifies a set of belief parameters, communica-

feature of hierarchical inference systems is th the relation P(D|H) is computable as a cascade local, more elementary probability relations in volving intervening variables. Intervening var ables, (e.g., organisms causing a disease) may may not be directly observable. Their computat al role, however, is to provide a conceptual summarization for loosely coupled subsets of ob vational data so that the computation of P(H|D

Bayesian Network for a Simple Conversation

Q1: If the season is dry, and the pavement is slippery, did it rain?

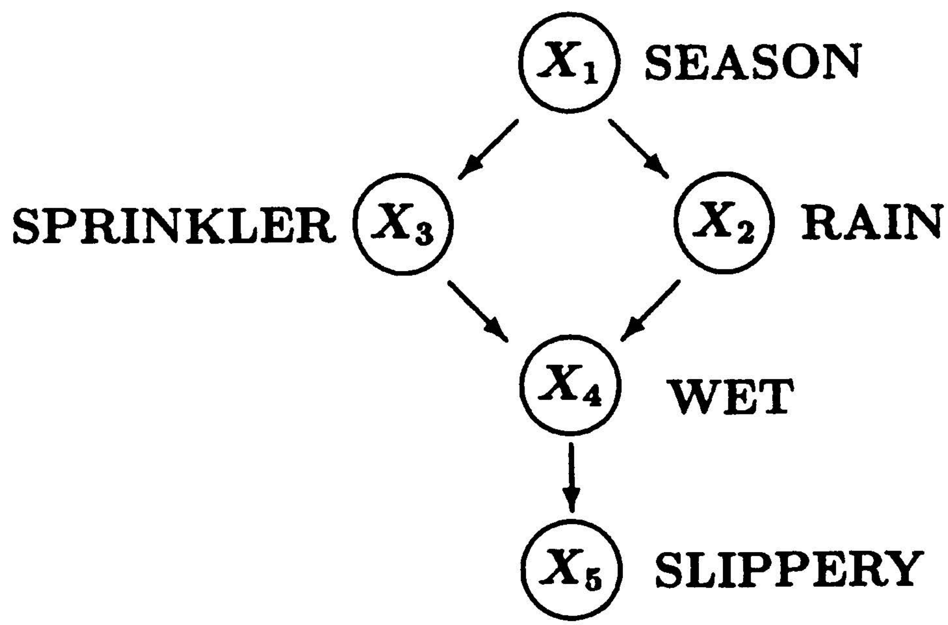
A1: Unlikely, it is more likely the sprinkler was ON.

Q2: But what if we SEE that the sprinkler is OFF?

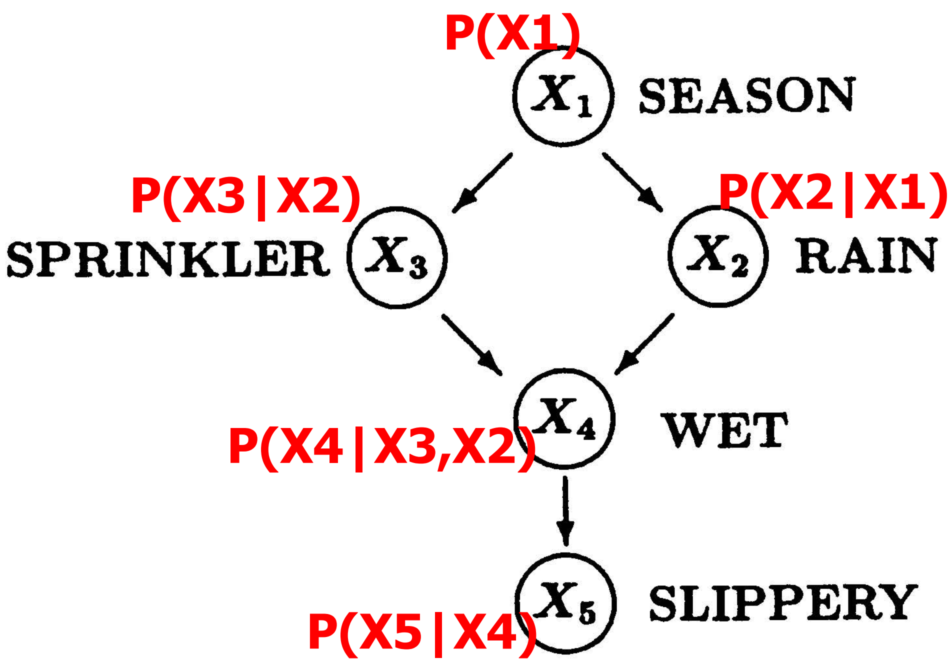
A2: Then it is more likely that it rained

.

The Story



Bayesian Network for a Simple Conversation



CPD:

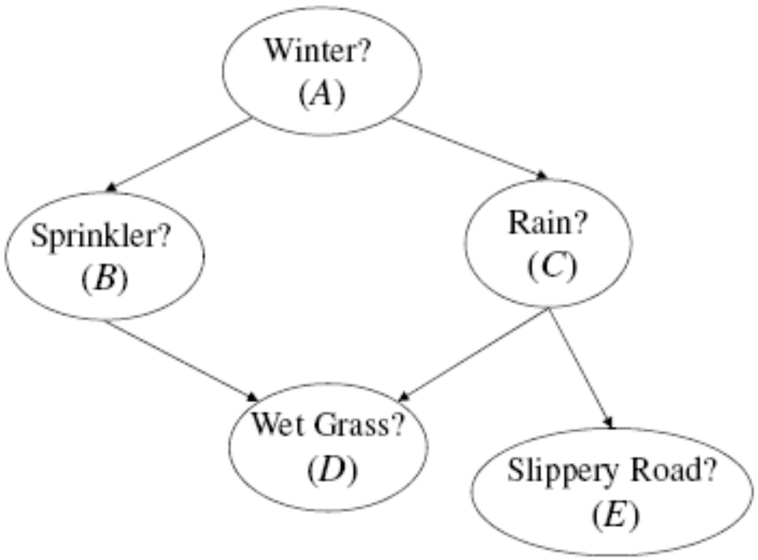
X3	X2	wet=0,	wet=1
0	0	0.9	0.1
0	1	0.1	0.9
1	0	0.2	0.8
1	1	0	1

P(X1,X2,X3,X4,X5))= *P(X1) P(X2|X3) P(X3|X1) P(X4|X3,X2) P(X5|X4)*

Conditional Independencies ➡ Efficient Representation

A Bayesian Network

Bayesian Networks encode independencies



A	Θ_A
true	.6
false	.4

A	B	$\Theta_{B\|A}$
true	true	.2
true	false	.8
false	true	.75
false	false	.25

A	C	$\Theta_{C\|A}$
true	true	.8
true	false	.2
false	true	.1
false	false	.9

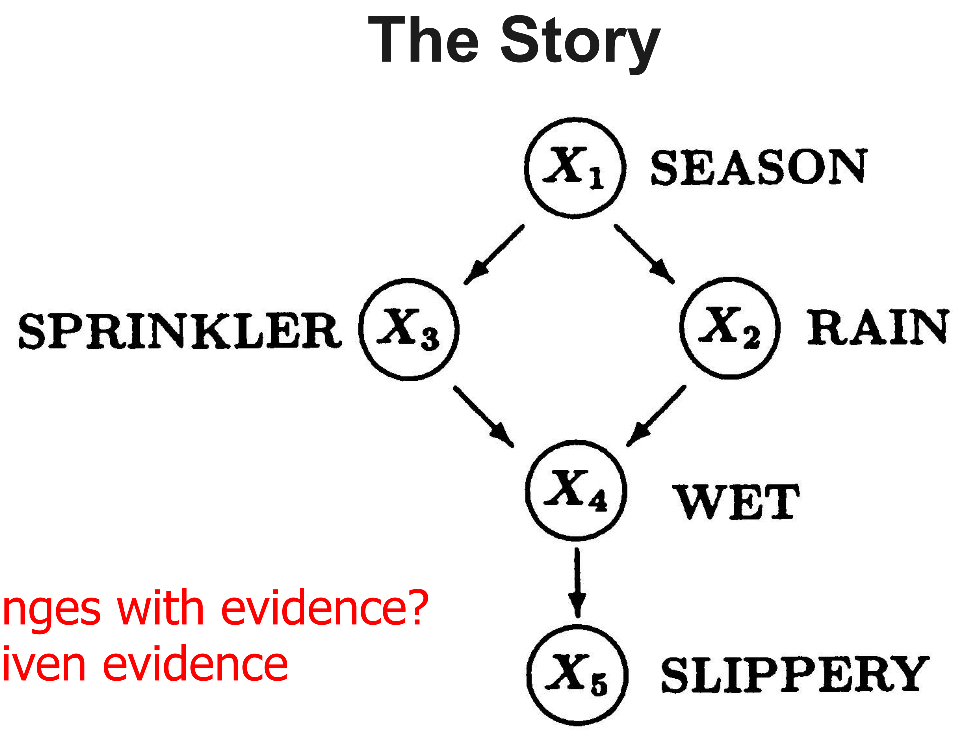
B	C	D	$\Theta_{D\|BC}$
true	true	true	.95
true	true	false	.05
true	false	true	.9
true	false	false	.1
false	true	true	.8
false	true	false	.2
false	false	true	0
false	false	false	1

C	E	$\Theta_{E\|C}$
true	true	.7
true	false	.3
false	true	0
false	false	1

Bayesian Network for a Simple Conversation

Q1: If the season is dry, and the pavement is slippery, did it rain?

Q2: But what if we SEE that the sprinkler is OFF?



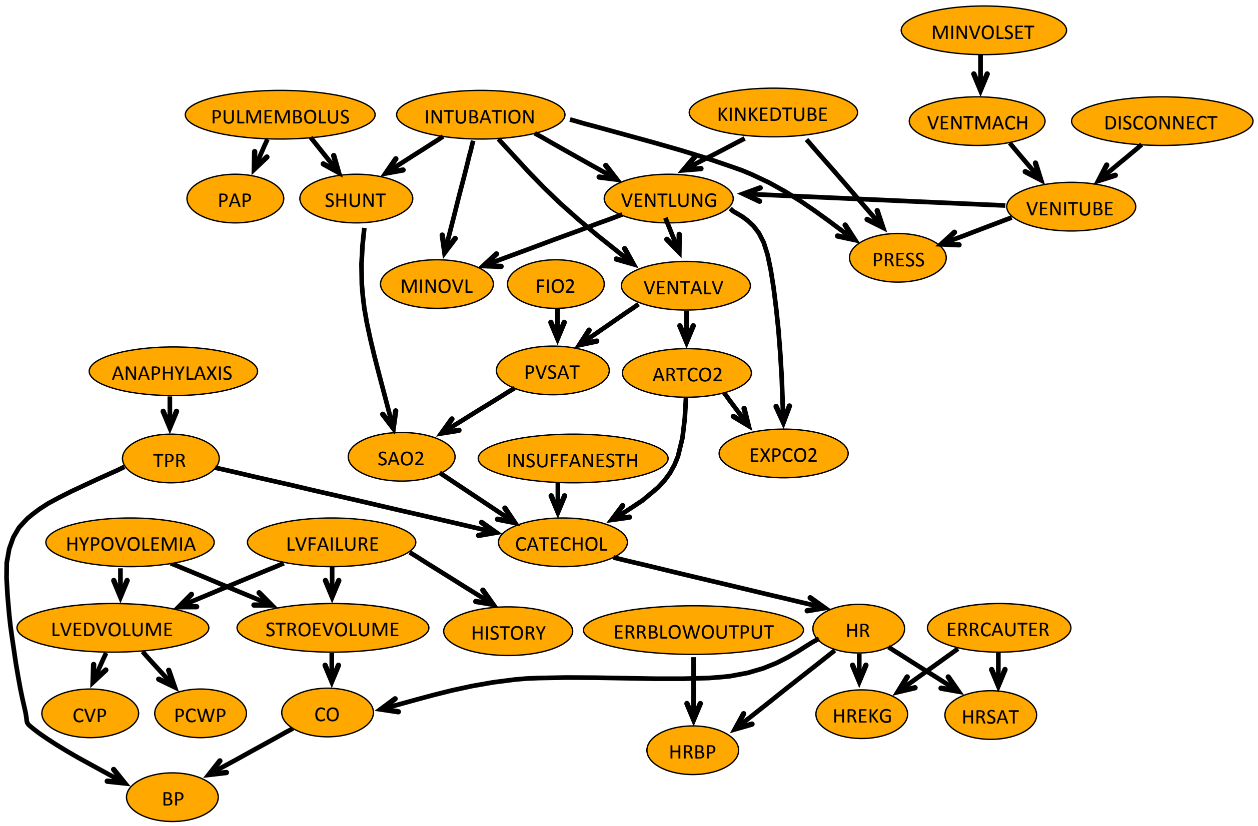
- **Belief updating**: how probability changes with evidence? What is more likely? Rain or not rain given evidence

$P(X1,X2,X3,X4,X5)) = P(X1)\ P(X2|X3)\ P(X3|X1)\ P(X4|X3,X2)\ P(X5|X4)$

Q1: Pr(rain=on | Slippery=yes, season= summer)?

Q2: Pr(rain=on | Slippery=off, season=winter)?

Monitoring Intensive Care Patients



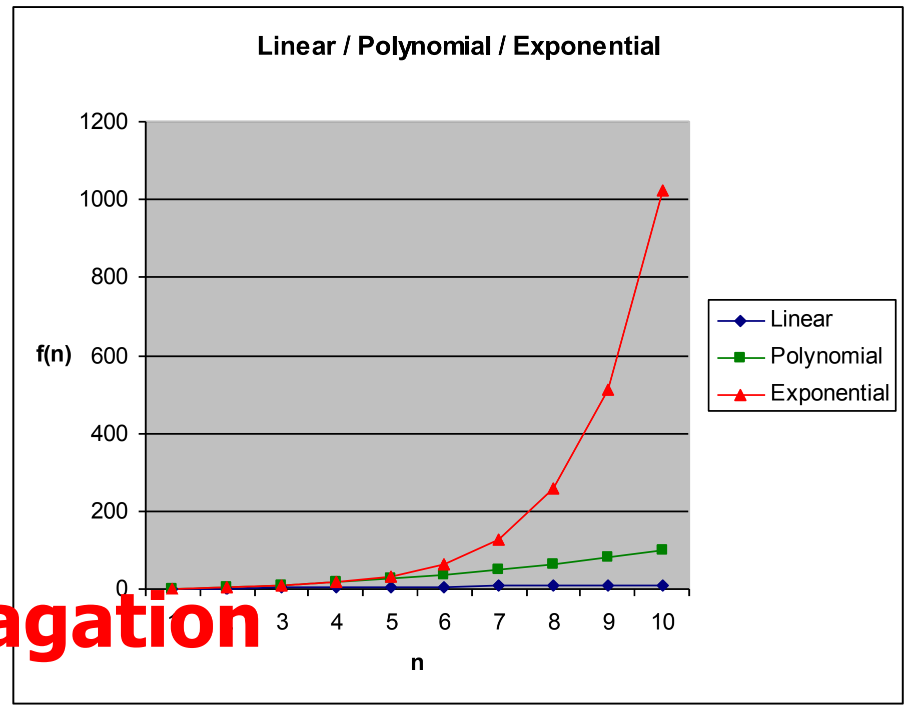
Alarm network

37 variables
509 parameters

<< 2^{37}

Complexity of Reasoning Tasks

- Belief updating
- Most probable explanation
- Decision-theoretic planning



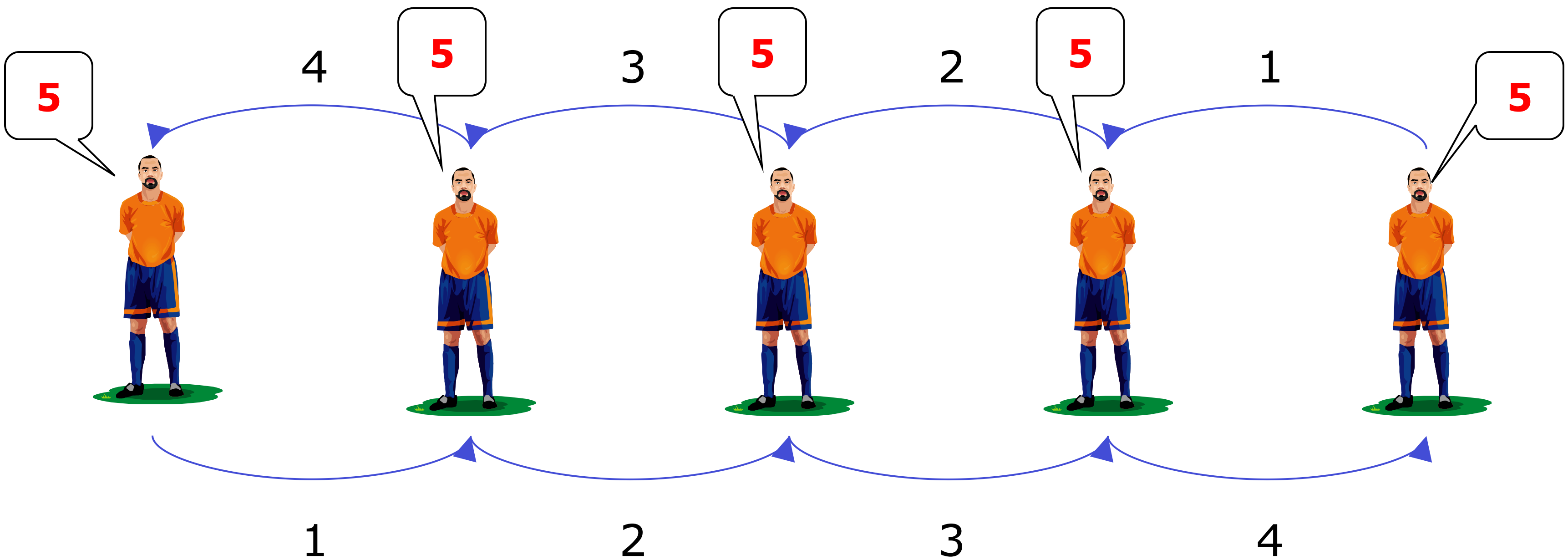
Reasoning is computationally hard

Approach: Belief Propagation (Pearl 1982)

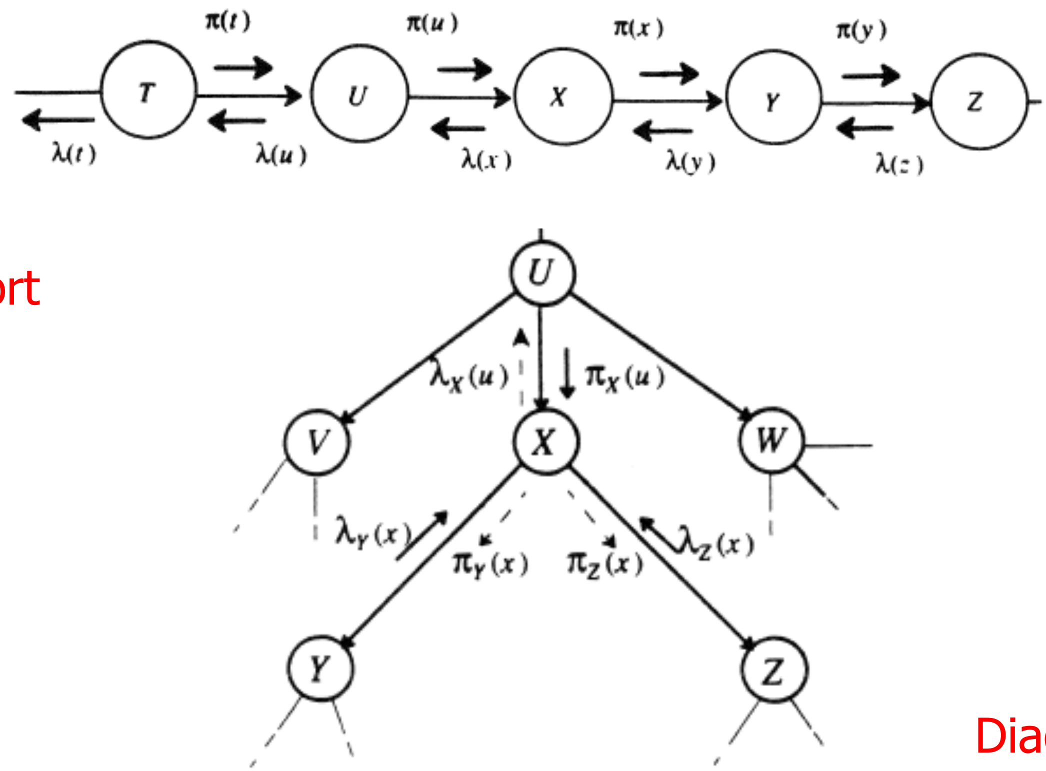
Distributed Belief Propagation

The essence of belief propagation is to make global information be shared locally by every entity

How many people?



Distributed Belief Propagation



Z=1

Causal support ↓

↑ Diagnostic support

Pearl (1982), (Belief Propagation in trees)

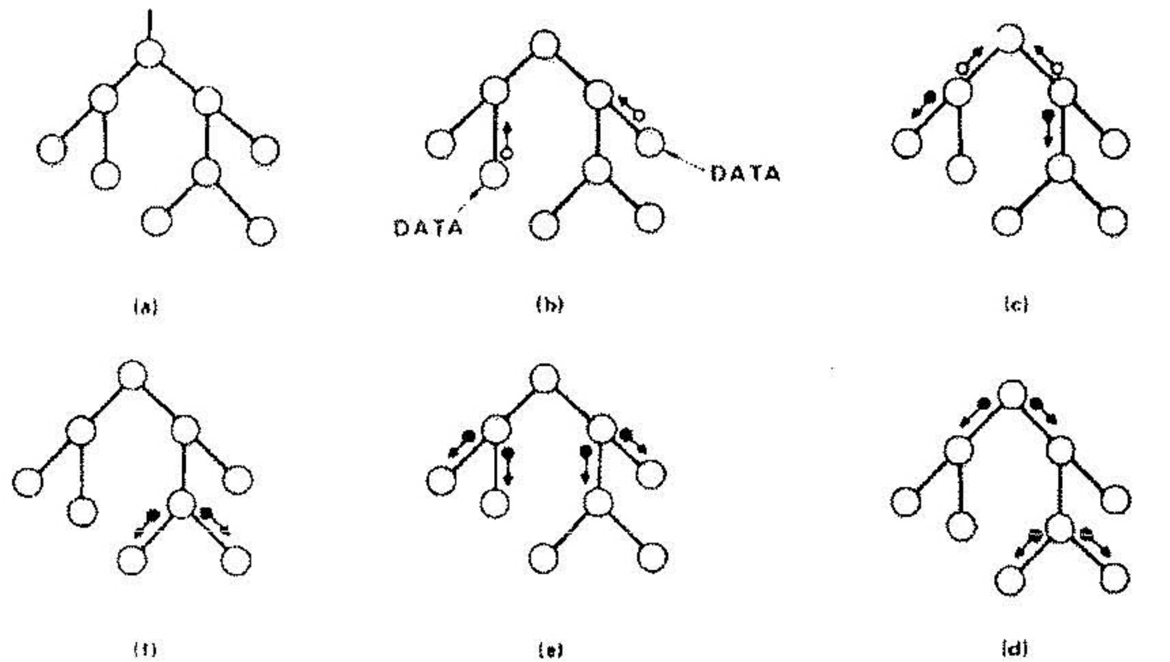


Figure 2

<u>Properties of the Updating Scheme</u>

1. The local computations required by the proposed scheme are efficient in both storage and time. For an m-ary tree with n states per node, each processor should store $n^2+mn+2n$ real numbers, and perform $2n^2+mn+2n$ multiplications per update. These expressions are on the order of the number of rules which each variable invokes.

Kim & Pearl (1983)
Explaining a way

data which we shall call D_R and ... vely. These data are defined as the observations and prior beliefs obtained only at the boundaries of network. Likewise, every node A partitions the graph into two parts: above A, G ., and below A, G_A , representing the data set D and D respectively. Figure 1 shows the causal network representing Mr. Holmes' belief structure.

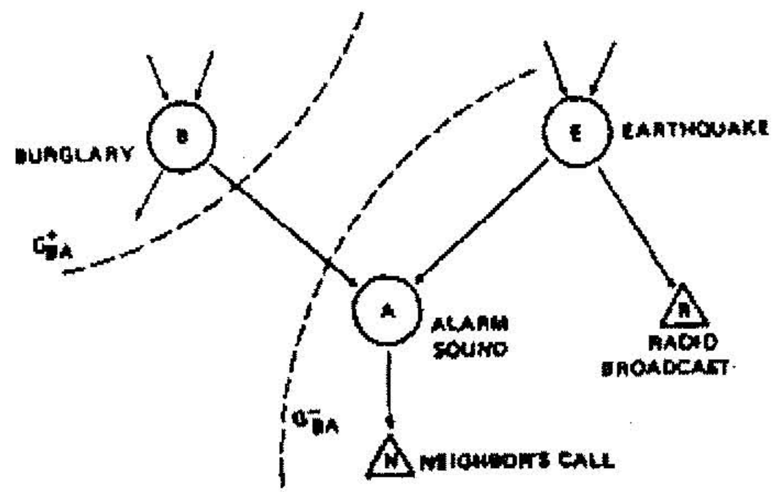


Figure 1 : Mr. Holmes' Belief Structure

III STRUCTURAL ASSUMPTIONS OF INDEPENDENCE

The likelihood of the various states of a variable X would, in general, depend on the entire data observed so far. However, the existence of only one path from G to X implies that the

which n
priori.
certaint
this
(See Fi

The i
to work
are vie
their
coupled
burglari
earthqu
likeliho
the occ
relation
2-b), a

(5) P

BAYESIAN NETWORKS: A MODEL OF SELF-ACTIVATED MEMORY FOR EVIDENTIAL REASONING*

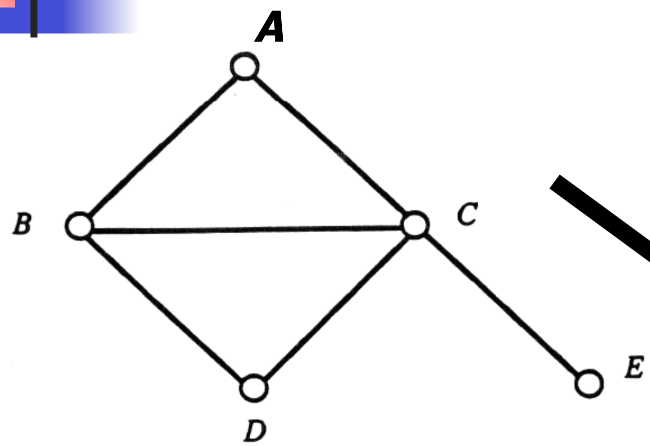
Bayes Net (1985)

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California
Los Angeles, CA 90024
(judea@UCLA-locus)
(213) 825-3243

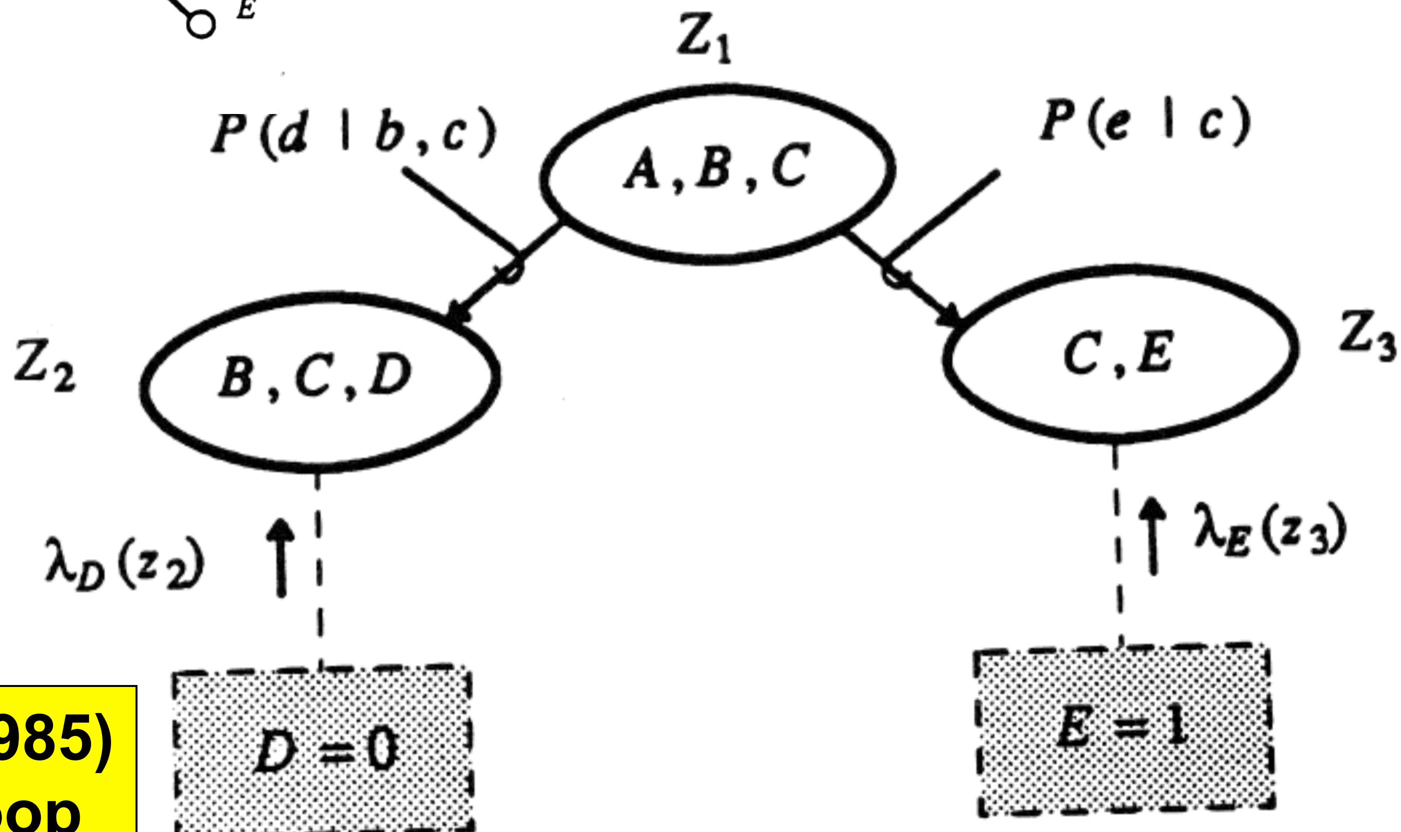
Topics: Memory Models
Belief Systems
Inference Mechanisms
Knowledge Representation

To be presented at
the 7th Conference of
the Cognitive Science Society
University of California, Irvine
August 15-17, 1985

A Loopy Bayesian Network

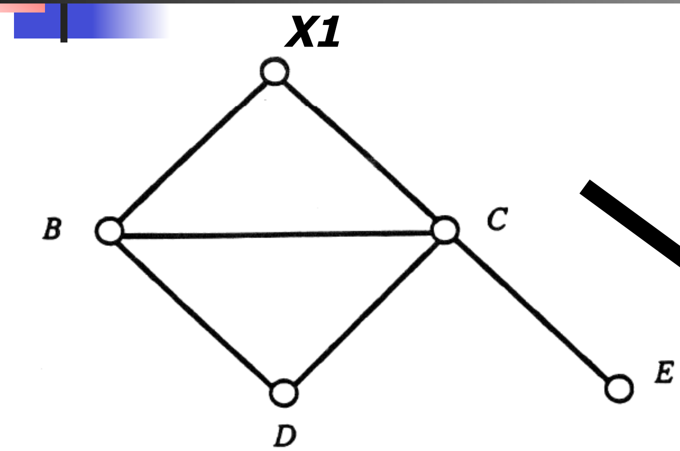


Tree clustering



Bayes Net (1985)
Breaking a loop

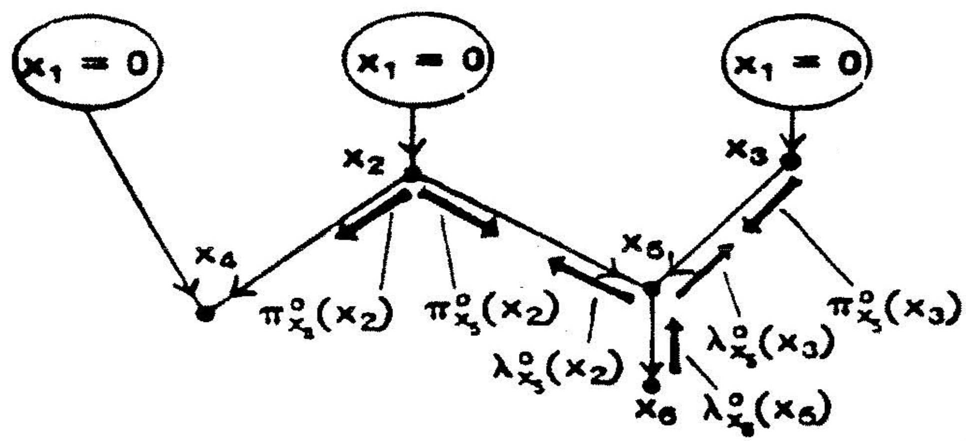
A Loopy Bayesian Network



Cutset-Conditioning

$$\pi^{0}_{x_6}(x_5) = \sum_{x_2,x_3=0,1} P(x_5 \mid x_2, x_3)\,\pi^{0}_{x_5}(x_2)\,\pi^{0}_{x_5}(x_3)$$

$$\pi^{1}_{x_6}(x_5) = \sum_{x_2,x_3=0,1} P(x_5 \mid x_2, x_3)\,\pi^{1}_{x_5}(x_2)\,\pi^{1}_{x_5}(x_3)$$



Bayes Net (1985)
Breaking a loop

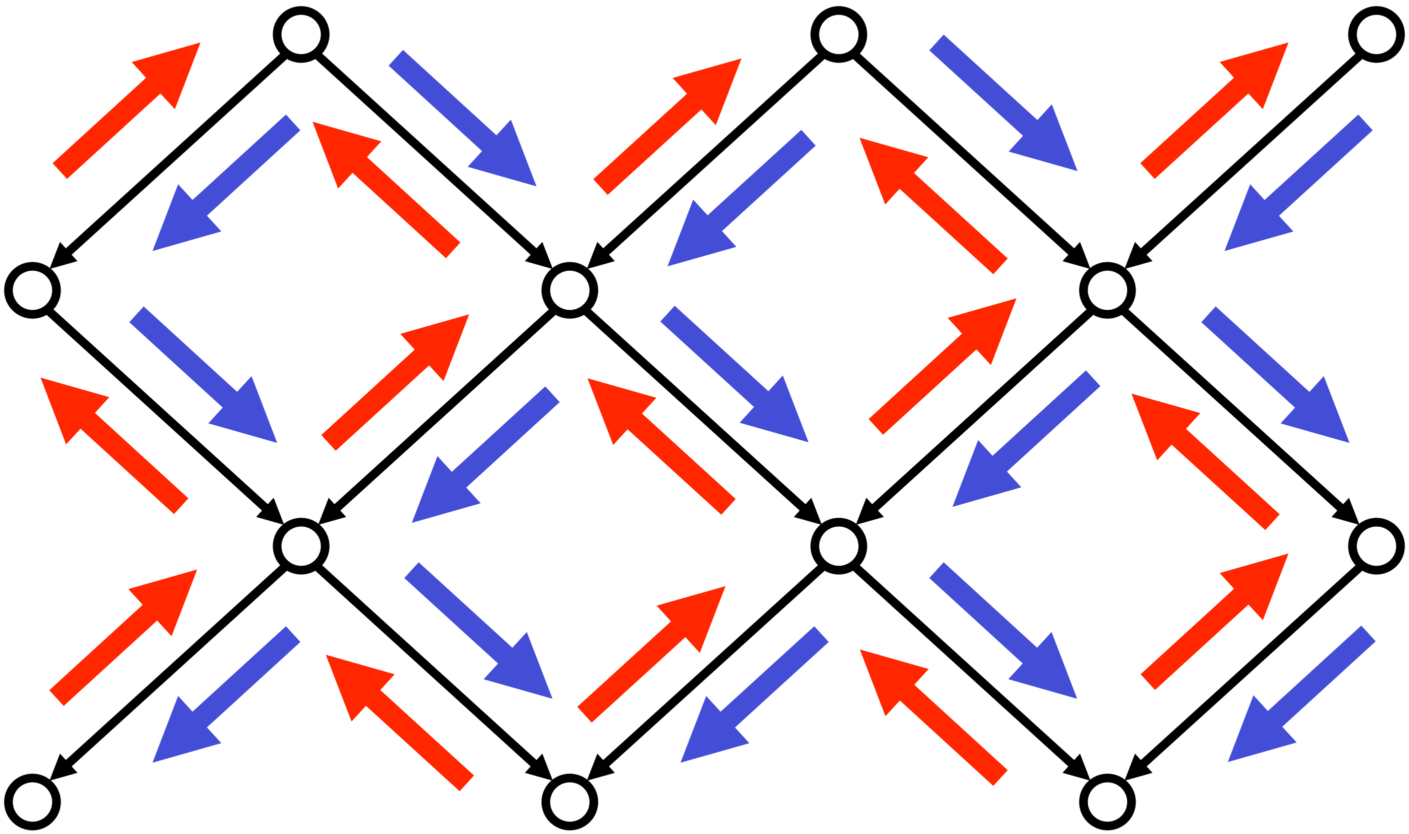
Loop-Breaking Techniques (1985 – 1990)

1. Conditioning (1985)
2. Stochastic simulation (1987)
3. Tree clustering (Spiegelhalter & Lauritzen 1986)
4. Node elimination (Shachter 1986)

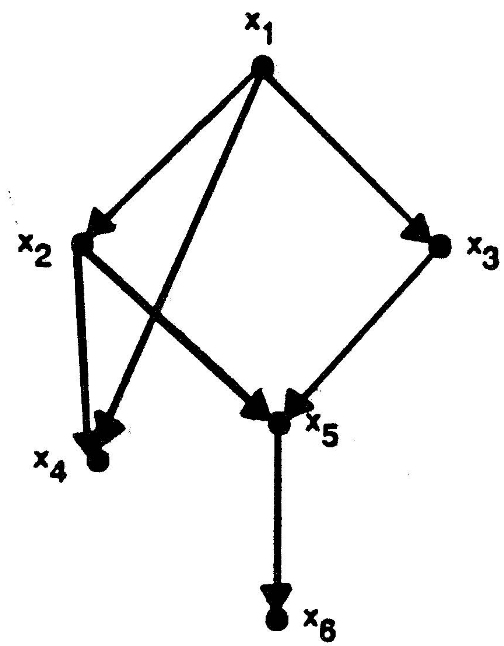
Problems:

- Time exponential in tree-width (Dechter 1996)
- Autonomy is lost

Loopy Belief Propagation



Bayesian Networks – Construction and d-Separation



Simple construction:

$$P(x_1,\ldots,x_n) = \prod_{i=1,\ldots,n} P(x_i \mid x_1,\ldots,x_{i-1})$$

$$= \prod_{i=1,\ldots,n} P(x_i \mid pa_i)$$

Qualitative judgment:
conditional independence

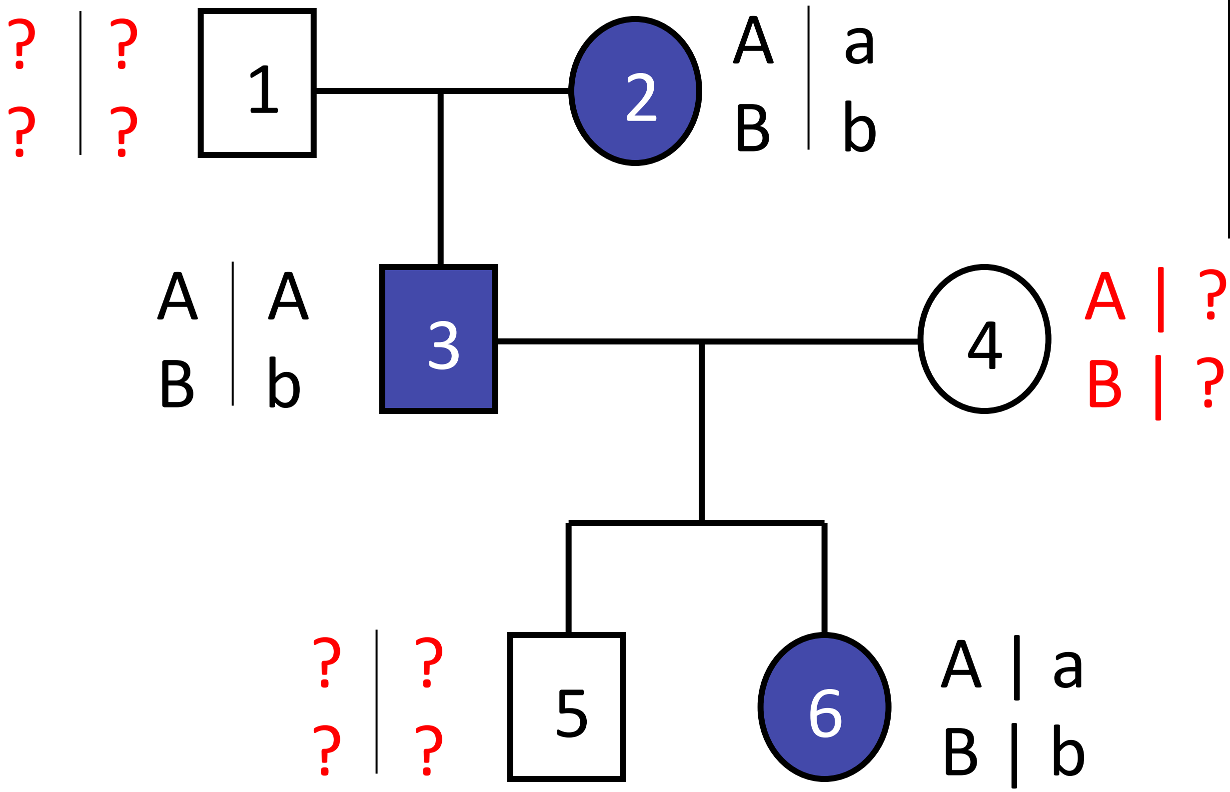
$$(X_i \perp\!\!\!\perp \mathrm{PRED}_i \mid pa_i)$$

$$P(x_1,x_2,x_3,x_4,x_5,x_6) = P(x_6 \mid x_5)P(x_5 \mid x_2,x_3)$$
$$P(x_4 \mid x_1,x_2)P(x_3 \mid x_1)P(x_2 \mid x_1)P(x_1)$$

Applications of Bayesian Networks

1. Medical Diagnosis
2. Clinical Decision Support
3. Complex Genetic Models
4. Crime Risk Factors Analysis
5. Spatial Dynamics in Geography
6. Inference Problems in Forensic Science
7. Conservation of a Threatened Bird
8. Classifiers for Modelling of Mineral Potential
9. Student Modelling
10. Sensor Validation
11. An Information Retrieval System
12. Reliability Analysis of Systems
13. Terrorism Risk Management
14. Credit-Rating of Companies
15. Classification of Wines
16. Pavement and Bridge Management
17. Complex Industrial Process Operation
18. Probability of Default for Large Corporates
19. Risk Management in Robotics

Linkage Analysis

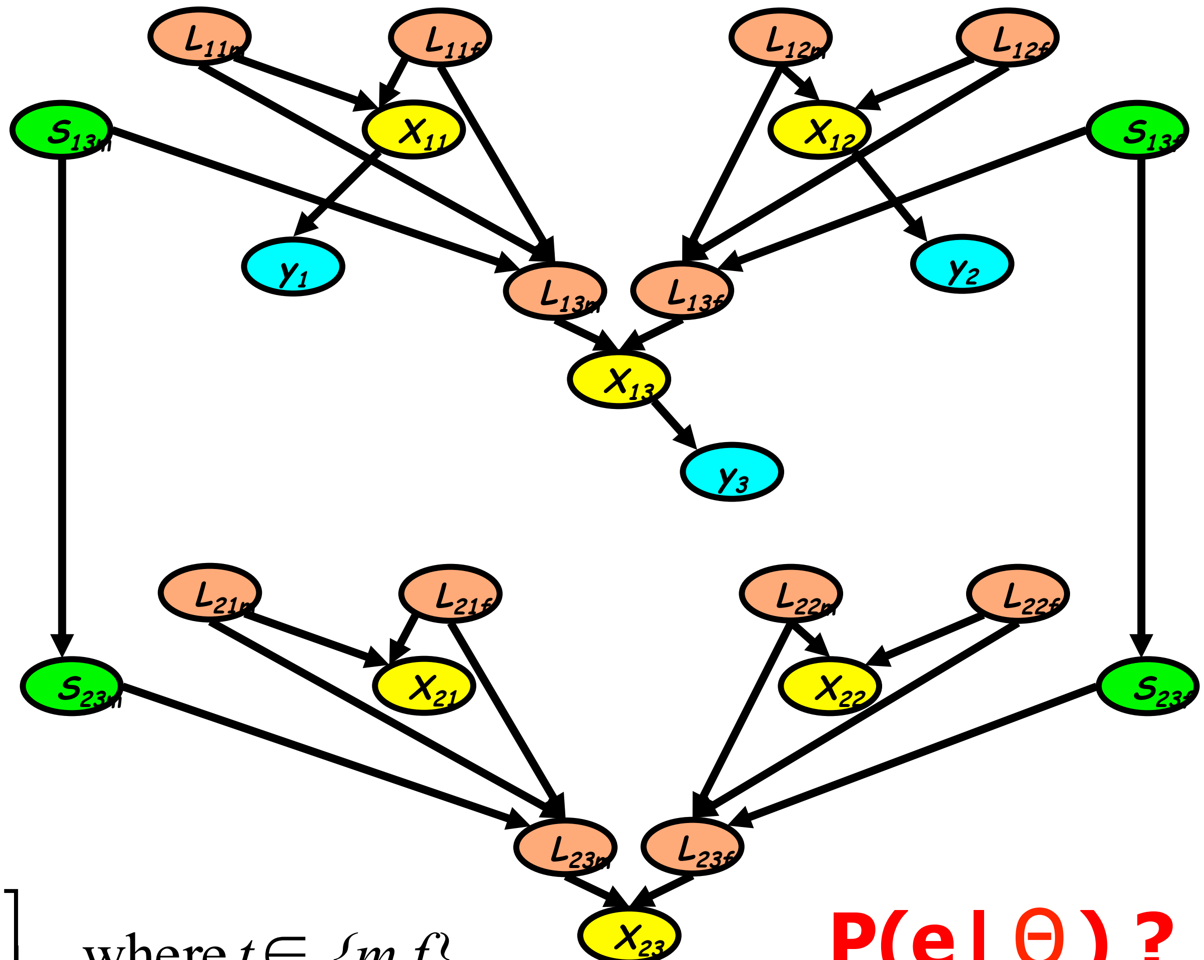


- 6 individuals
- Haplotype: {2, 3}
- Genotype: {6}
- Unknown

Bayesian Network for Recombination

Locus 1

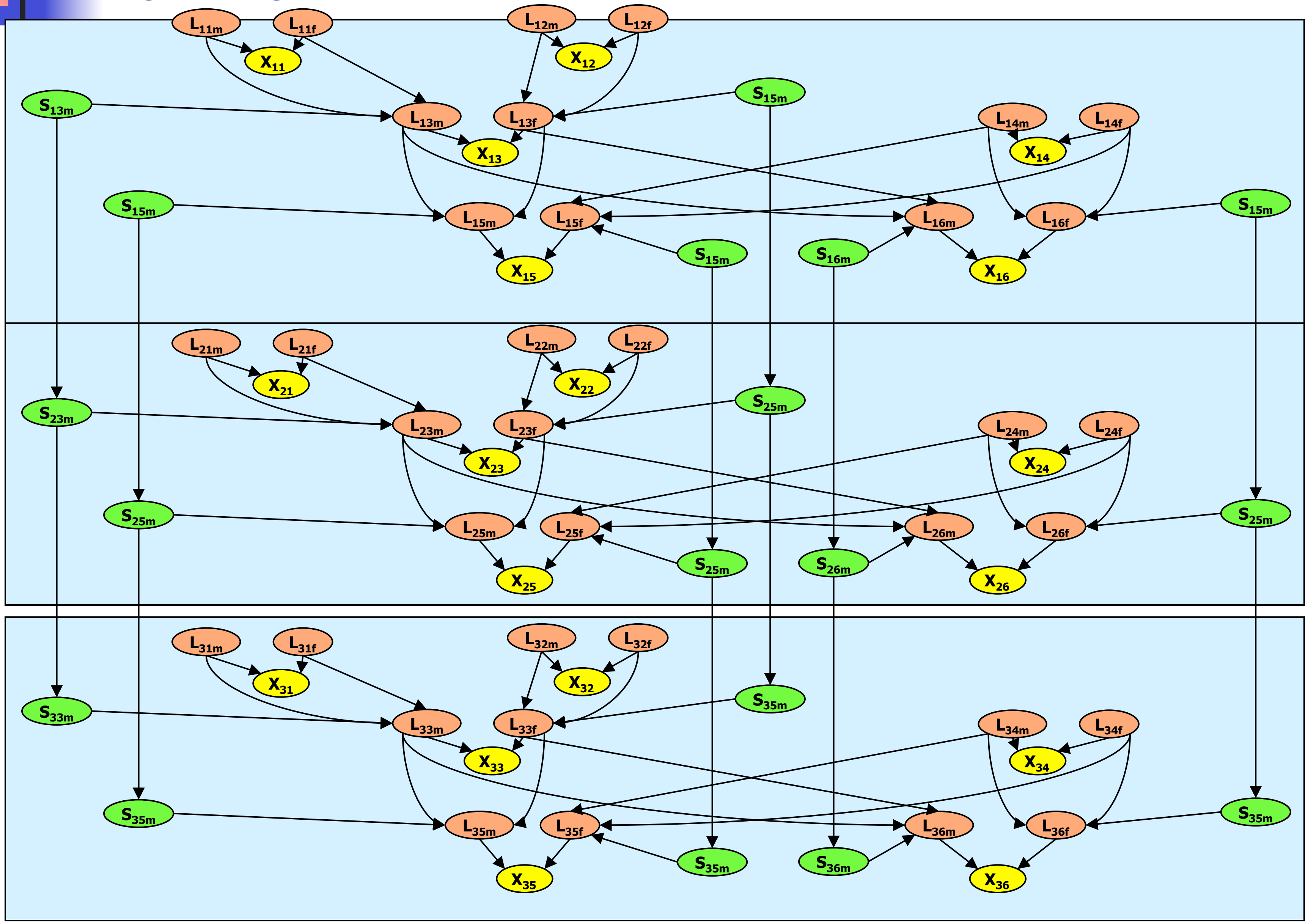
Locus 2



$$P(s_{23_t} \mid s_{13_t}, \theta) = \begin{bmatrix} 1-\theta & \theta \\ \theta & 1-\theta \end{bmatrix} \quad \text{where } t \in \{m,f\}$$

P(e| Θ) ?

6 people, 3 markers



Summary of Bayesian Networks

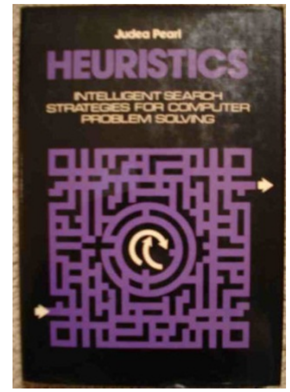
- The framework of Bayesian networks revolutionized AI: Pearl showed how Bayesian networks and their belief-updating algorithms provide an intuitive, elegant characterization of complex probability distributions, and the way they track new evidence.

- Probabilistic Networks remains the most successful approach to solving problems of representing, organizing, and exploiting information. His approach has changed the face of research in machine learning, which relies fundamentally on probabilistic and statistical inference.

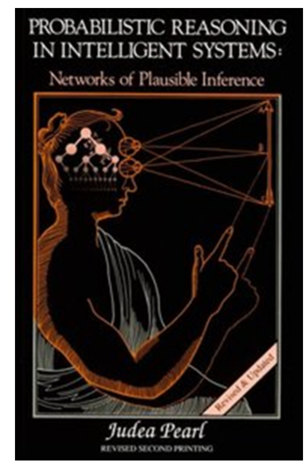
- Bayesian networks have also altered the analysis of biological data, with applications in medicine ranging from the design of HIV vaccines to the search for genetic causes of disease. They also underlie most systems for speech recognition, fault diagnosis, and machine translation. His 1988 book *Probabilistic Reasoning in Intelligent Systems* offers techniques based on belief networks that provide a mechanism for making semantics-based systems operational.

Pearl’s Main Contributions

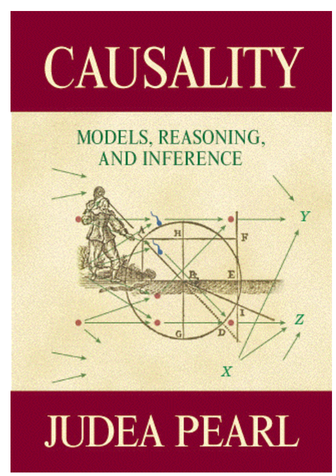
- Heuristic Search (1984)



- Probabilistic Reasoning (1988)



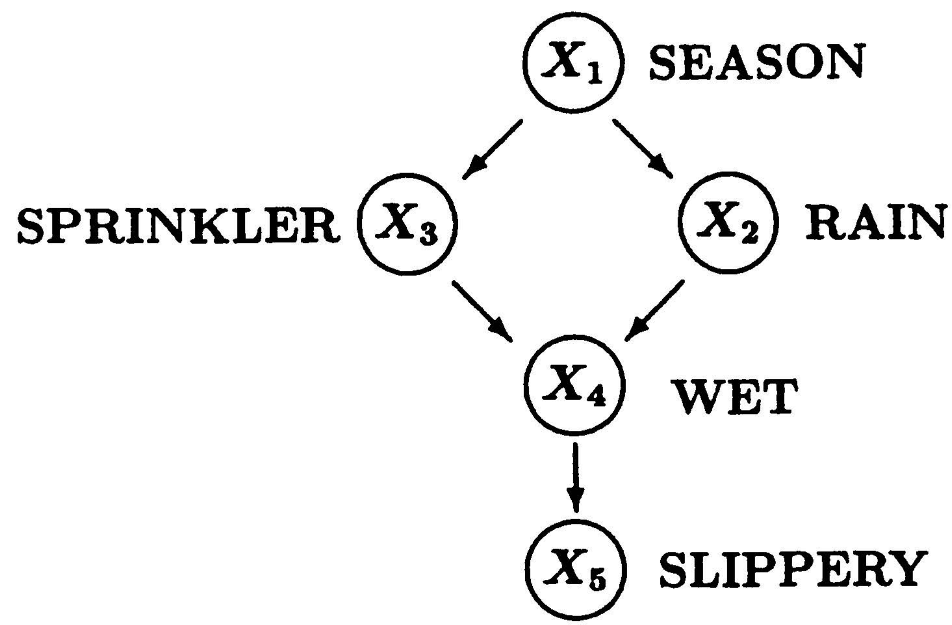
- Causality (2000)



Bayesian Network for a Causal Conversation

Q2: But what if we SEE that the sprinkler is OFF?
A2: Then it is more likely that it rained
Q3: Do you mean that if we actually turn the sprinkler OFF, the rain will be more likely?
A3: No, the likelihood of rain would remain the same

The Story



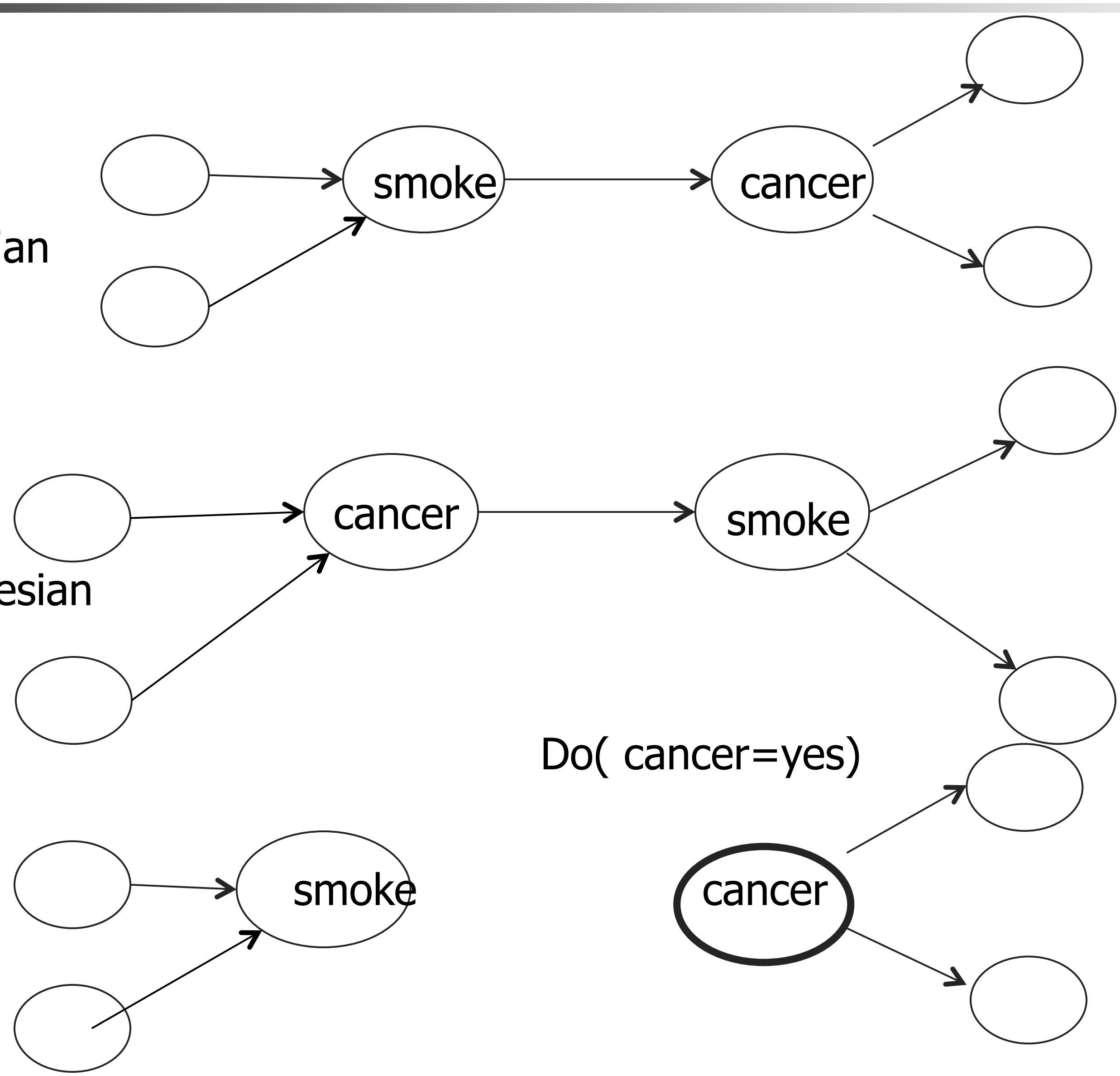
.

Observing (sprinkler=on) $\neq$ **Doing (sprinkler=on)**

Smoking and Cancer

Casual Bayesian Network

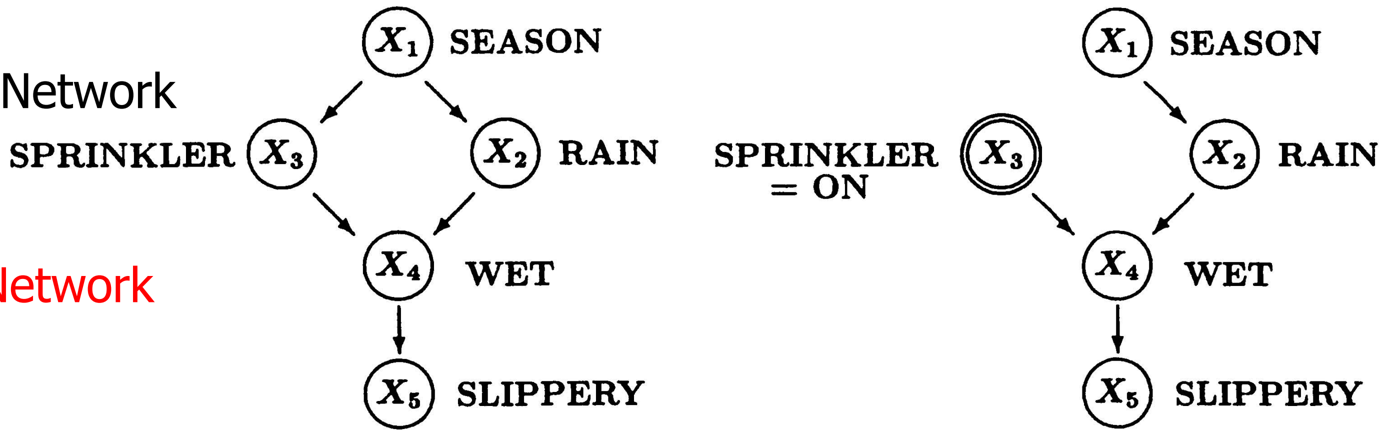
Non-causal Bayesian Network



Seeing vs. Doing

Bayesian Network

Causal Network



Effect of turning the sprinkler ON: P(season | do(sprinkler=on))

Pearl *do-calculus* leads to a complete mathematical framework for formulating causal models and for analyzing data to determine causal relationships.

Rules For do-Calculus

Rule 1: Ignoring observations

$$P(y \mid do\{x\}, z, w) = P(y \mid do\{x\}, w)$$

if $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}$

Rule 2: Action/observation exchange

$$P(y \mid do\{x\}, do\{z\}, w) = P(y \mid do\{x\}, z, w)$$

if $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\underline{Z}}}$

Rule 3: Ignoring actions

$$P(y \mid do\{x\}, do\{z\}, w) = P(y \mid do\{x\}, w)$$

if $(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\,\overline{Z(W)}}}$

Causality

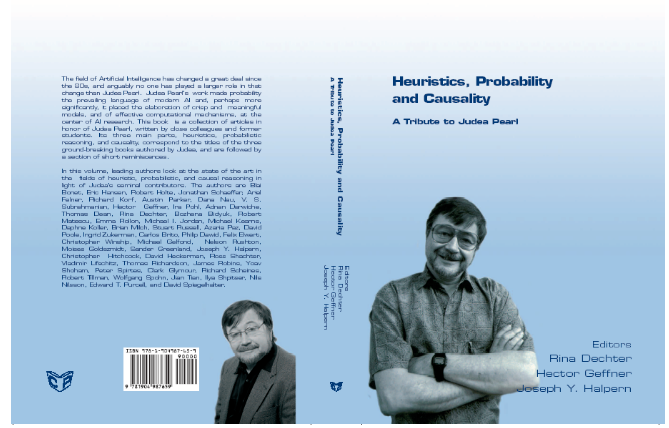
- *causal network* captures the potential effect of exogenous intervention
- *Smoking* → *Cancer* as a causal network captures our beliefs about how the world works. Namely, intervening on cancer does not change the likelihood a person smoke, but intervening on smoking does.
- Pearl developed the *do-calculus*, leads to a complete mathematical framework for formulating causal models and for analyzing data to determine causal relationships.
- This work changed long-held belief in statistics that causality can be determined only from controlled random trials – which are impossible in areas such as the biological and social sciences.

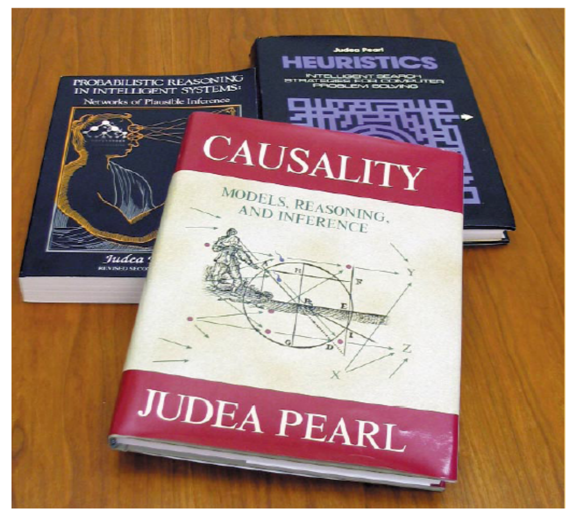
Summary: Pearl's Turing Award

Pearl's Bayesian networks provided a syntax and a calculus for probability models, in much the same way that George Boole provided a syntax and a calculus for logical models.

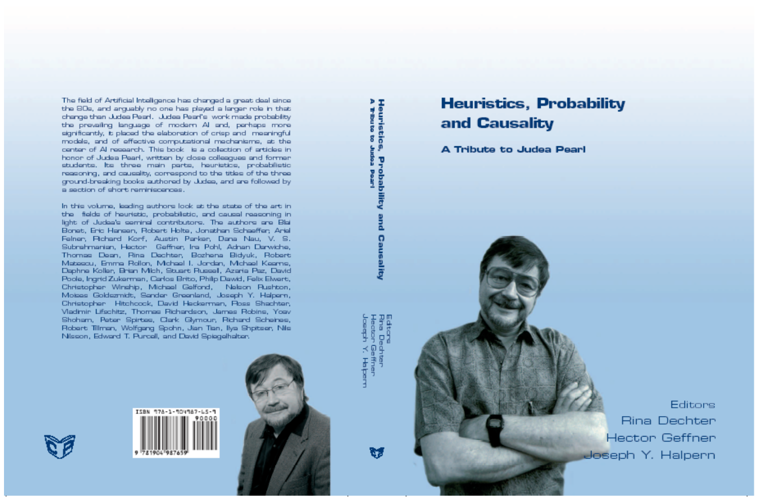
Theoretical and algorithmic questions associated with Bayesian networks form a significant part of the modern research agenda for machine learning and statistics,

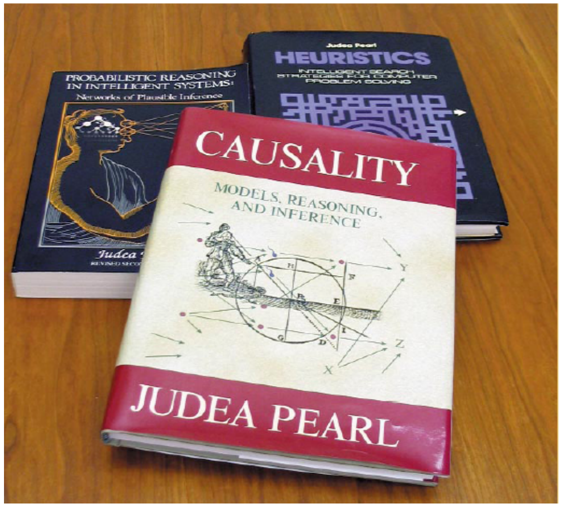
Their use has also permeated other areas, such as natural language processing, computer vision, robotics, computational biology, and cognitive science. As of 2012, some 50,000 publications have appeared with Bayesian networks as a primary focus. (Acm, 2011)





Thank you


Heuristics, Probability and Causality
A Tribute to Judea Pearl
Editors
Rina Dechter
Hector Geffner
Joseph Y. Halpern


PROBABILISTIC REASONING IN INTELLIGENT SYSTEMS:
Networks of Plausible Inference
Judea Pearl
HEURISTICS
INTELLIGENT SEARCH STRATEGIES FOR COMPUTER PROBLEM SOLVING
CAUSALITY
MODELS, REASONING, AND INFERENCE
JUDEA PEARL