UNCERTAINTY

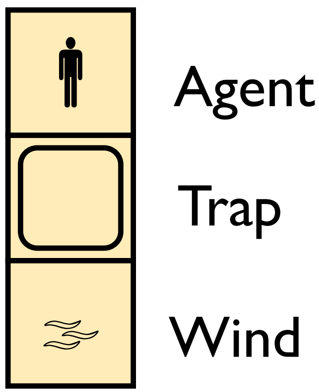
In which we see what an agent should do when not all is crystal-clear.

Outline

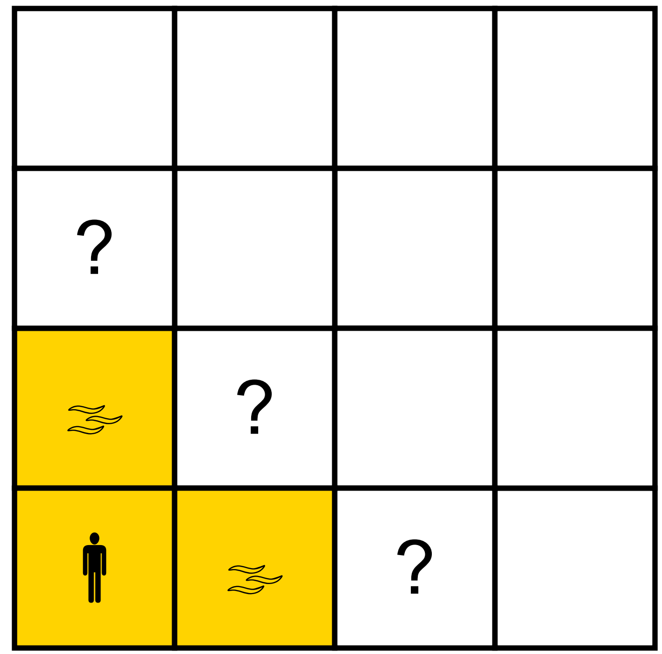
- Uncertainty
- Probabilistic Theory
- Axioms of Probability
- Probabilistic Reasoning
- Independency
- Bayes' Rule
- Summary

Uncertainty

WUMPUS-World

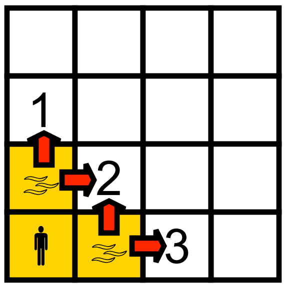


Where are the traps?

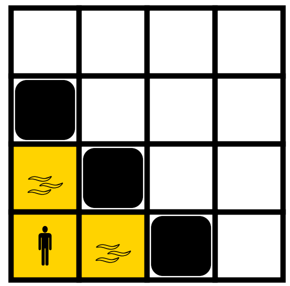


There are no secure actions, but which one is best?

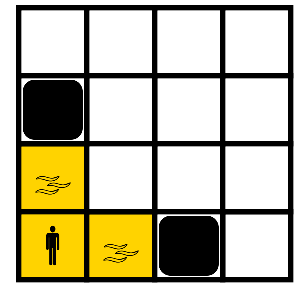
Reasoning under Uncertainty



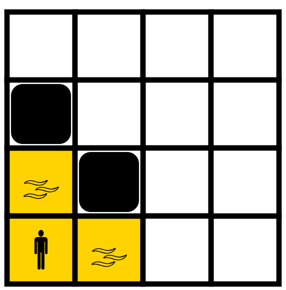
Actions 1,2, 3



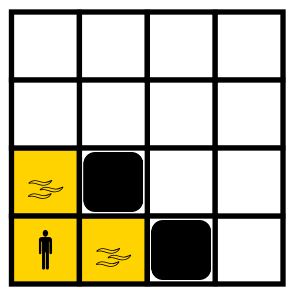
Successful:
Fails: 1,2, 3



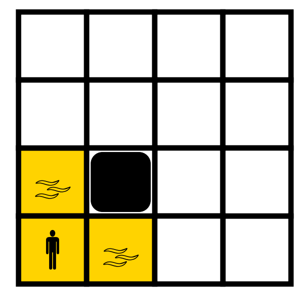
Successful: 2
Fails: 1,3



Successful: 3
Fails: 1,2



Successful: 1
Fails: 1,2



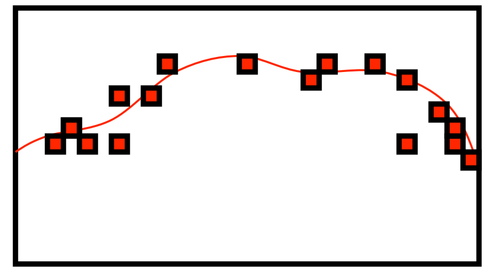
Successful: 1,3
Fails: 2

Causes of Uncertainty

- Incomplete knowledge
- Agents unlikely to have complete knowledge about their environment
- Practical uncertainty: not enough data, e.g. noise
- Theoretical uncertainty: no complete theory exists, e.g. medical diagnosis

Uncertainty and Vagueness

- Example:
  Boundaries of spatial objects
  - uncertain boundaries: not known where exactly boundary is
  - vague boundary: boundary is not sharp but localized



Uncertain knowledge:
sonar records of a wall

Vague knowledge:
transition between soil types

Causes of Uncertainty

- Complex knowledge
  - A complete formalization in FOL is to complex to describe
- Qualification problem
  - Rules about application area are incomplete because there are too many preconditions

Boundaries of Logics

- Technical Diagnosis
  - Input: Symptoms
  - Output: Sources of errors

- Example
  - Symptom: no sound from radio
  - Sources of errors:
    On/Off switch, batteries, ...

- Diagnostic Reasoning

$$\forall p\, Symptom(p, no\_sound) \implies Failure(p, switch) \lor Failure(p, battery) \lor \ldots$$

- Causal Reasoning

$$Failure(p, switch) \lor Failure(p, battery) \lor \implies \forall p\, Symptom(p, no\_sound)$$

Logics and Uncertainty

- Default logic
  - Rules are valid until contradicted
    $\forall p \, Symptom(p, no\_sound) \implies Failure(p, switch)$
  - Additional knowledge can invalid former derivations
- – Non-Monotonicity

- Fuzzy logic
  - – Describes the degree of validity of a statement
  - – rocky(location) = 0.43
    i.e. 43% rock
  - – Approach for vagueness, not uncertainty
  - – Description of natural language

Probability

- Probabilistic Statement
  - *"There is a 70% chance of an empty battery if the portable Bluetooth player does not give a sound."*
- Combined uncertainty
  - From two different sources
  - "Unknown": non-existing knowledge
  - "Incomplete": existing knowledge too complex

Question

- True or False?
  - Probabilistic Theory has the same ontological commitment as logics:
  - Facts hold or do not hold in the world

Different AI Logics

- Logic languages
  - "Logics"
  - Characterized through language elements (logic constants)
  - Facts? Objects?
  - Time? Belief?
  - Ontological commitment (wrt. reality, what exists in world)
  - Epistemological commitment (what an agent believes about facts)

Logic language	Ontological commitment	Epistemological commitment
Propositional logic	Facts	True/False/ Unknown
FOL	Facts Objects Relations	True/False/ Unknown
Temporal logic	Facts Objects Relations, Times	True/False/ Unknown
Probabilistic theory	Facts	Degree of belief 0..1
Fuzzy-Logic	Degree of truth	Degree of belief 0..1

Uncertainty and Decisions

- Action A_t
  - Go to airport *t* minutes before flight
- Probabilistic verdicts through agent
  - A_{30} is the probability that I get the flight, 1%
  - A_{360} is the probability that I get the flight, 99%

- Which action does the agent select?
  - Does not depend only on probability, also on preferences (→ decision theory)
  - Decision theory = Probabilistic theory + Utility theory

Uncertainty and Decisions

function DT-AGENT(*percept*) **returns** an *action*
  persistent: *belief_state*, probabilistic beliefs about the current state of the world
    action, the agent's action

  update *belief_state* based on *action* and *percept*
  calculate outcome probabilities for actions,
    given action descriptions and current *belief_state*
  select *action* with highest expected utility
    given probabilities of outcomes and utility information
  return *action*

Figure 13.1 A decision-theoretic agent that selects rational actions.

Probabilistic Theory

- How we see it in class
  - Probabilistic theory as an extension of propositional logic
  - Extension: the truth values are labeled with probabilities
- Note
  - Probability theory makes the same ontological commitment as logic, namely, that facts either do or do not hold in the world

Random Variable

- Basic Idea
  - *P(X=a)* quantifies the probability that random variable X takes the value *a*

- Variables
  - Boolean
    - Range [TRUE,FALSE]
  - discrete
    - Range of finite set of boolean variables, e.g. weather [sunny, rainy, cloudy, snow]
  - continuous
    - Real values or subsets, e.g. [0,1]

Atomic Events

- Complete specification of the world state about which the agent is uncertain.

- Example:
  Cavity (c) and toothache (t) has four atomic events (aE)

1. Exclusive: only one statement true, ($c \wedge t$) and ($c \wedge \neg t$) is mutually exclusive
2. Exhaustive: at least one of set of all aE is the case, i.e. disjunction of all aE is logically equivalent to TRUE
3. Entailment: from aE we can entail the truth or falsehood of every proposition, e.g.
   ($c \wedge \neg t$) entails c = TRUE, $c \Rightarrow t$ = FALSE
4. Propositions are logically equivalent to disjunction of all aE that entail the truth of the proposition. E.g. the proposition *cavity* is equivalent to the disjunction of the aE $c \wedge t$ and $c \wedge \neg t$.

Prior Probability

- Probability
  - also: a priori probability, unconditional probability
  - P(a) = 0.4

    means: probability associated with the proposition *a* is the degree of belief accorded to it in the absence of any other information

- P-Distribution

$$\mathbf{P}(Switch\_on) = (0.4, 0.6)$$
$$P(Switch\_on = 1) = 0.4$$
$$P(Switch\_on = 0) = 0.6$$

$$\mathbf{P}(Weather) = (0.4, 0.29, 0.3, 0.01)$$
$$P(Weather = sunny) = 0.4$$
$$P(Weather = rainy) = 0.29$$
$$P(Weather = cloudy) = 0.3$$
$$P(Weather = snowy) = 0.01$$

$P(Cavity = true) = 0.1$ or $P(cavity) = 0.1$

Probability Distribution

- A boolean random variable A
  - For all possible values a value of truth
- Probability distribution **P**(A)
  - Denotes probability for all possible values for random variable *A*.

Value	Probability
A	**P**(A)
0	0.3 = P(¬A)
1	0.7 = P(A)

Note: similar notation!

P(A) short for *P(A=1)*

P*(A)* P-Distribution of A

Joint Probability Distribution

- Multiple boolean random variables
  - Cavity, toothache, catch
  - For each possibility a defined probability

- Joint distribution **P**(A,B,C)
  - Gives probabilities of all possible value combinations of random variables
  - e.g.: ***P****(Cavity, Toothache, Weather)* can be represented by a 2x2x4 table with 16 entries

$$\mathbf{P}(\neg \mathbf{cavity} \wedge \neg \mathbf{toothache} \wedge \neg \mathbf{catch}) = \mathbf{0.576}$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Conditional Probability

- Conditional probability
  - Agent has evidence (information) about former unknown variables
  - A priori probabilities no longer applicable
  - Notation: *P(A|B)*, *A* and *B* are any propositions
  - This is read as "the probability of *A* given that all we know is *B*."
  - Example.:

$$P(cavity|toothache) = 0.8$$

- **Product rule**
  - conditional probabilities can be defined as unconditional probabilities:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)}$$

  - which holds for all P(b)>0
  - Can also be written as **product rule**:

$$P(a \wedge b) = P(a|b)\,P(b)$$
$$P(a \wedge b) = P(b|a)\,P(a)$$

  - $\mathbf{P}(X,Y) = \mathbf{P}(X|Y)\,\mathbf{P}(Y)$

Kolmogorov's axioms of probability

- Axiom 1

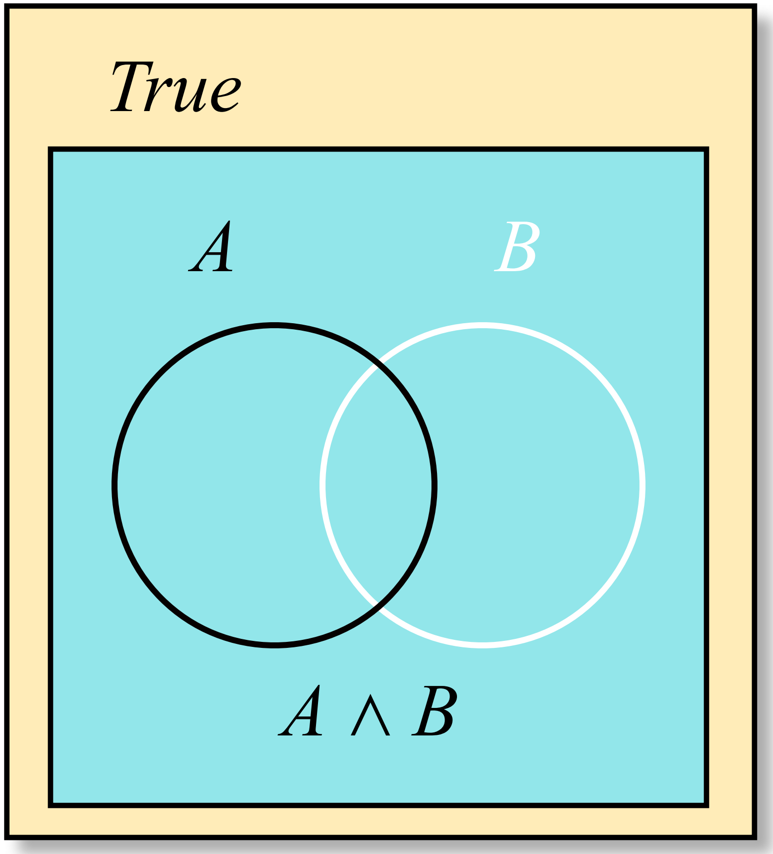
$$0 \leq P(a) \leq 1$$

- Axiom 2

$$P(Tautology) = 1$$
$$P(Contradiction) = 0$$

- Axiom 3

$$P(a \vee b) =$$
$$P(a) + P(b) - P(a \wedge b)$$



Using axioms of probability

Derivation of a variety of useful facts from the basic axioms. E.g., the familiar rule for negation follows by substituting ¬*a* for *b* in axiom 3:

$$
\begin{aligned}
P(a \vee \neg a) &= P(a) + P(\neg a) - P(a \wedge \neg a) && (by\ axiom\ 3\ with\ b = \neg a) \\
P(true) &= P(a) + P(\neg a) - P(false) && (by\ logical\ equivalence) \\
1 &= P(a) + P(\neg a) && (by\ axiom\ 2) \\
P(\neg a) &= 1 - P(a) && (by\ algebra)
\end{aligned}
$$

Sum of all probabilities is always 1.

Are these axioms reasonable?

- Example

$$P(a \wedge b) = 0.0$$

  - Agent 2 chooses to bet $4 on a, $3 on b and $2 on ¬(a ∨ b)

- Theorem
  - If Agent 1 expresses a set of degrees of belief that violate the axioms of probability theory then there is a betting strategy for Agent 2 that guarantees that Agent 1 will lose money on every bet.
  - Proven by Finetti (1931)

Agent 1

prop.	belief
a	0.4
b	0.3
a ∨ b	0.8
a ∧ b	0.0

Agent 2

bet	stakes
a	4 to 6
b	3 to 7
¬(a ∨ b)	2 to 8

Output for Agent 1

a, b	a, ¬b	¬a, b	¬a,¬b
-6	-6	4	4
-7	3	-7	3
2	2	2	-8
-11	-1	-1	-1

Probabilistic Inference

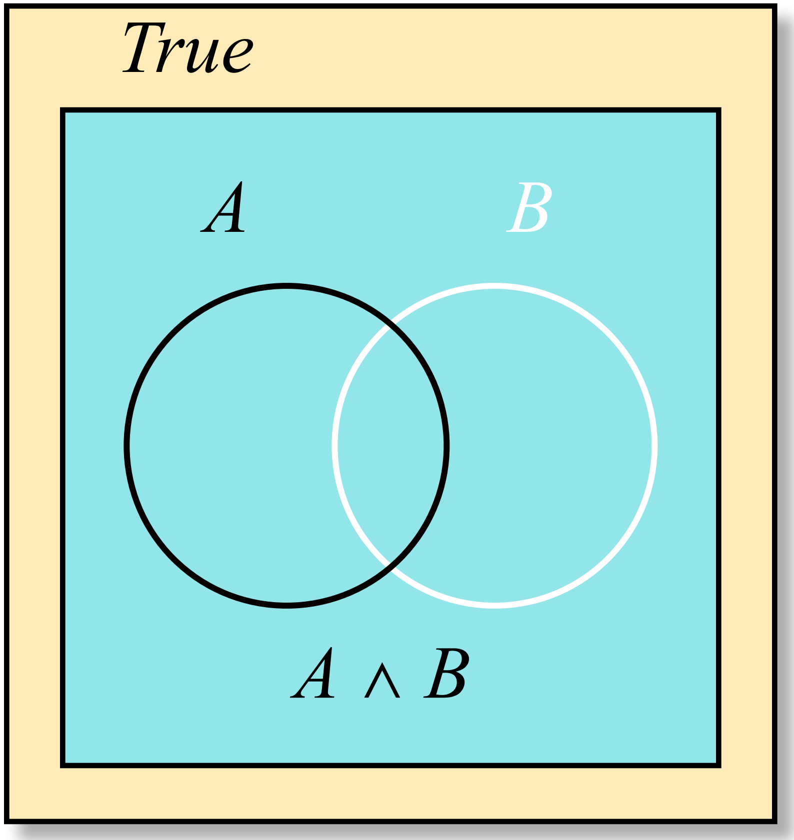
- Conditional probabilities

$$P(A|B) = P(A \wedge B)/P(B)$$
$$if\ P(B) > 0$$

- As product rule

$$P(A \wedge B) = P(A|B)\, P(B)$$
$$if\ P(B) > 0$$

$$P(A \wedge B) = P(B|A)\, P(A)$$
$$if\ P(A) > 0$$



Inferences with joint probabilistic distributions

- Inferences
  - Calculate posterior probabilities for given evidences
  - aka "knowledge base"
- Probabilities
  - Sum is 1
  - Helps to calculate simple and complex propositions
  - Take atomic events where proposition is true and then add probabilities

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Inferences with joint probabilistic distributions

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- For example, there are six atomic events in which *cavity* ∨ *toothache* holds:

  $P(cavity \lor toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$

- **Marginal probability**
  - Extract the distribution over some subset of variables, z.B. cavity
  - $P(cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Conditional Probabilities

- Example 1
  - Use product rule and take the probability distribution. Here: probability for having cavity given toothache:

$$P(cavity|toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

- Example 2
  – Compute the probability that there is no cavity, given toothache:

$$P(\neg cavity|toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Normalization

- Constant
  - *1/P(toothache)* always constant, no matter what value exists for *Cavity*
  - It can be viewed as a normalization constant for the distribution **P**(*Cavity*|*toothache*), ensuring that it adds up to 1
  - We will use α to denote such constants
  - We can then write the two preceding equations in one

$$\mathbf{P}(Cavity|toothache) = \alpha \mathbf{P}(Cavity, toothache)$$
$$= \alpha[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]$$
$$= \alpha[\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle]$$
$$= \alpha \langle 0.12, 0.08 \rangle$$
$$= \langle 0.6, 0.4 \rangle$$

Algorithmic Analysis

- Joint probability distribution
  - n variables with max. k values
- Calculation complexity
  - Table size $O(k^n)$ is exponential in n
  - In the worst case: $O(k^n)$ calculation steps
- Problem in practice
  - We need a lot! of observations in order to get valid and reliable table entries!

Independence

- New variable: Weather
  - Joint probability distribution $\mathbf{P}(Toothache, Catch, Cavity, Weather)$
  - Weather has 4 values = 32 entries

- Cavity is independent from weather (also marginal or absolute independency)

$$\mathbf{P}(Toothache, Catch, Cavity, Weather) = \mathbf{P}(Toothache, Catch, Cavity)\,\mathbf{P}(Weather)$$

→ Reduction of entries: 8 + 4 instead of 32!

- In general

$$\mathbf{P}(X|Y) = \mathbf{P}(X)$$
$$\mathbf{P}(Y|X) = \mathbf{P}(Y)$$
$$\mathbf{P}(X, Y) = \mathbf{P}(X)\,\mathbf{P}(Y)$$

Decomposition of Joint Distributions

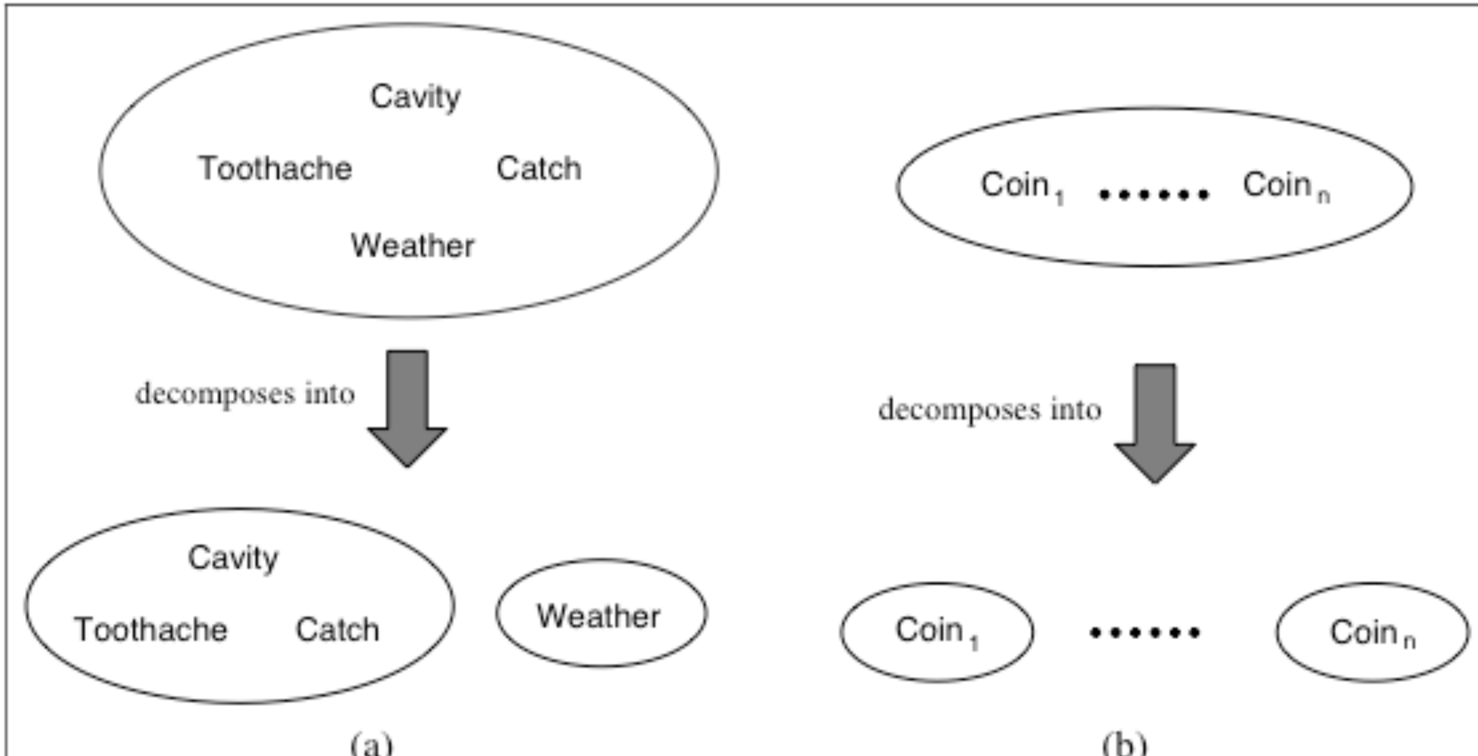


Figure 13.5 Two examples of factoring a large joint distribution into smaller distributions, using absolute independence. (a) Weather and dental problems are independent. (b) Coin flips are independent.

Bayes'-Rule

- Fundamental idea
  - Get around joint distributions
  - Calculate directly with conditional probabilities

- Recall the two forms of the product rule

$$P(a \vee b) = P(a|b)\,P(b) \quad for\ P(b) > 0$$
$$P(a \vee b) = P(b|a)\,P(a) \quad for\ P(a) > 0$$

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)} \qquad \mathbf{P}(Y|X) = \frac{\mathbf{P}(X|Y)\mathbf{P}(Y)}{\mathbf{P}(X)}$$

- Known as Bayes'-Rule

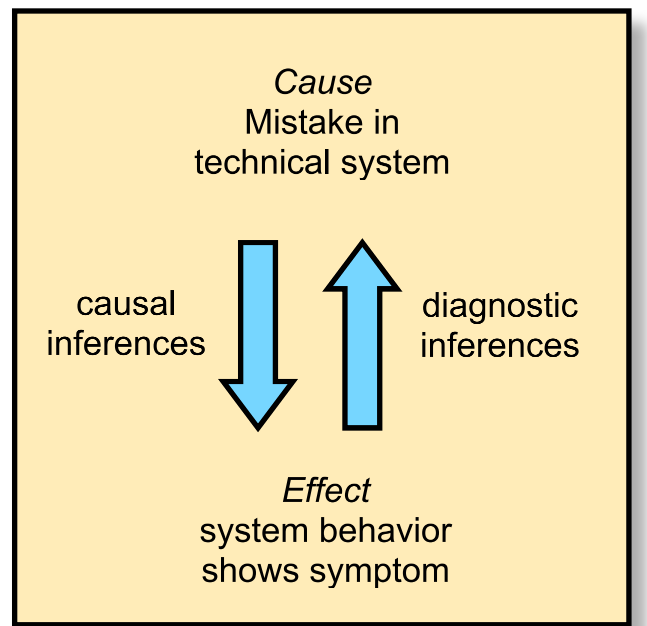
Bayes'-Rule

Why is this rule useful?

- Causal experiences
  C: cause, E: effect
- Diagnostic Inference

$$P(C|E) = \frac{P(E|C)P(C)}{P(E)}$$

This simple equation underlies all modern AI systems for probabilistic inference!



Technical Diagnosis

- Example
  - *C = power_switch_off, E = no_sound_heard*
  - Bayes'-Rule:

$$P(C|E) = \frac{P(E|C)P(C)}{P(E)}$$

  - Knowledge from causal experience

$$P(E|C) = 0.98$$
$$P(C) = 0.01 \quad (mostly\ on)$$
$$P(E) = 0.2 \quad (often\ defect)$$

Technical Diagnosis

- Conditional probability

$$P(C|E) = \frac{P(E|C)P(C)}{P(E)} = \frac{0.98 * 0.01}{0.2} = 0.049$$

- Other independent probabilities
P(C) = 0.2 (more often switched off)

$$P(C|E) = \frac{P(E|C)P(C)}{P(E)} = \frac{0.98 * 0.2}{0.2} = 0.98$$

- Depends highly on a priori probability *P(C)*!

Combining evidences

- So far
  - an evidence of the form $P(Effect|Cause)$
  - What if there is more than one evidence? $\mathbf{P}(Cavity|toothache \wedge catch)$
  - Possible with joint distributions, but what if we have large problems?

- **Conditional Independence**
  - *Toothache* and *Catch* are not really independent
  - Probe goes into tooth that has cavity and catches the tooth
  - They are not directly dependent: but related but via cavity

$$\mathbf{P}(toothache \wedge catch|Cavity) = \mathbf{P}(toothache|Cavity)\,\mathbf{P}(catch|Cavity)$$

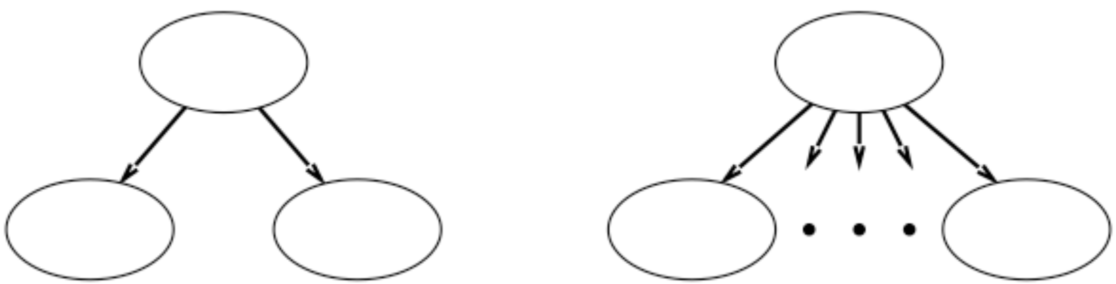
- Here also: significant reduction of algorithmic complexity O(n) instead O(2^n)

Bayes' Rule and Conditional Independence

$$\mathbf{P}(Cavity|toothache \land catch)$$
$$= \alpha\mathbf{P}(toothache \land catch|Cavity)\,\mathbf{P}(Cavity)$$
$$= \alpha\mathbf{P}(toothache|Cavity)\,\mathbf{P}(catch|Cavity)\,\mathbf{P}(Cavity)$$

This is an example of a naive Bayes model:

$$\mathbf{P}(Cause|Effect_1, ..., Effect_n) = \mathbf{P}(Cause)\prod \mathbf{P}(Effect_i|Cause)$$



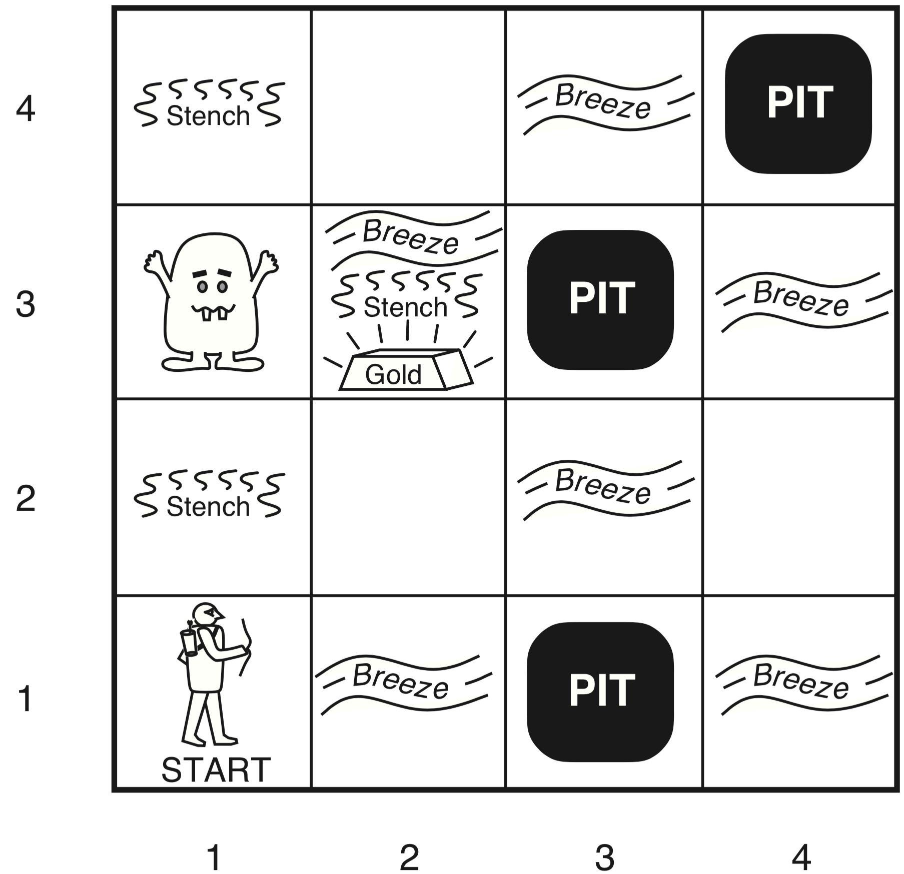
The total number of parameters is linear in *n*.

Notes

- Conditional independence
  - allows "large" systems that rely on probabilities
  - analogue to independence
- Decomposition
  - ...of large probabilistic application areas important in AI
- Dentist domain
  - shows that a single cause can have a number of effects
  - theses are conditional independent (given cause)
  - this pattern is often seen
- Joint Probability Distribution

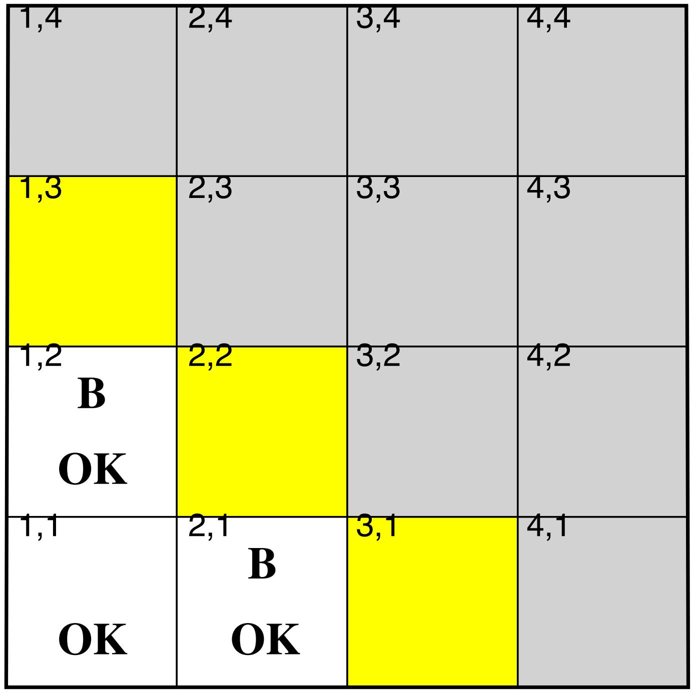
$$P(Cause, Effect_1, ..., Effect_n) = P(Cause) \prod P(Effect_i | Cause)$$

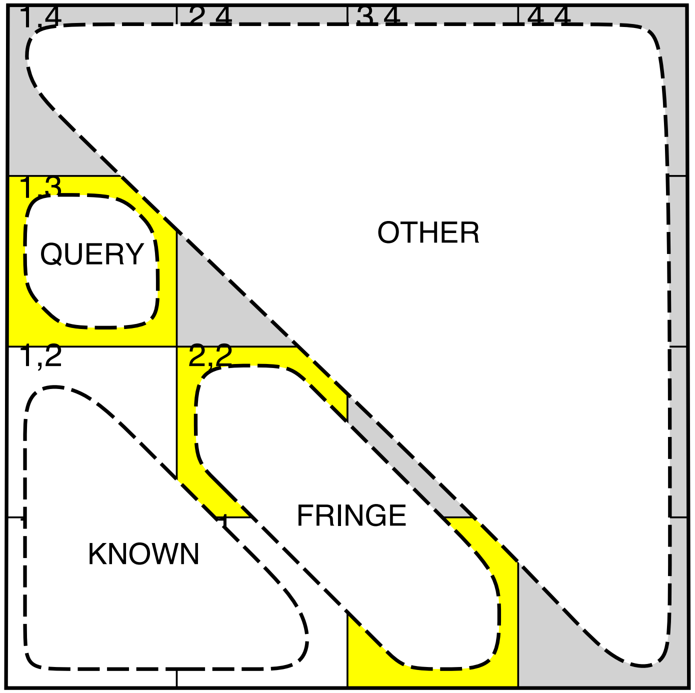
- These distributions are also called naive Bayes' Models (also Bayes' classifier)



- Five sensors
  - Square containing the wumpus and the adjacent squares: *Stench*
  - Adjacent to a pit: *Breeze*
  - Square with gold: *Glitter*
  - Agent walks into a wall: *Bump*
  - If rumpus is killed: *Scream*

Wumpus World





Boolean variable for each square P_{ij}, which is true iff square $[i, j]$ actually contains a pit

We also have Boolean variables B_{ij} which are true iff square $[i, j]$ is breezy; we include these variables only for the observed squares—in this case,$[1, 1]$, $[1, 2]$, and $[2, 1]$.

Specifying the Probability Model

The full joint distribution is: $\mathbf{P}(P_{1,1}, ..., P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$

Apply the product rule: $\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} | P_{1,1}, ..., P_{4,4})\,\mathbf{P}(P_{1,1}, ..., P_{4,4})$

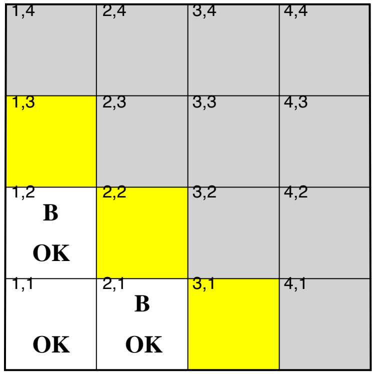
$to\ get\ \mathbf{P}(Effect|Cause)$

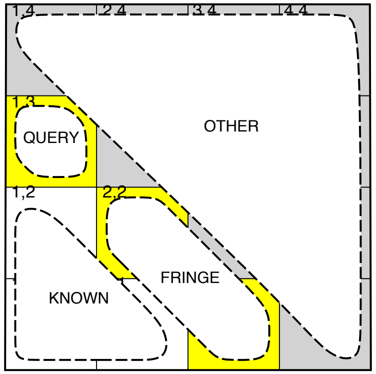
First term: value is 1 if pits are adjacent to breezes, 0 otherwise
Second term: priors, pits are placed randomly, probability 0.2 per square

$$\mathbf{P}(P_{1,1}, ..., P_{4,4}) = \prod_{ij=1,1}^{4,4} \mathbf{P}(P_{ij}) = 0.2^{n} \times 0.8^{16-n}$$

$for\ n\ pits.$

Observations and Query





We know the following facts:

$$b = \neg b_{1,1} \wedge b_{1,2} \wedge b_{2,1}$$

$$known = \neg p_{1,1} \wedge \neg p_{1,2} \wedge \neg p_{2,1}$$

Query:

$$\mathbf{P}(P_{1,3}|known, b)$$

Defining $Unkown = P_{i,j}$s other than $P_{1,3}$ and $Known$

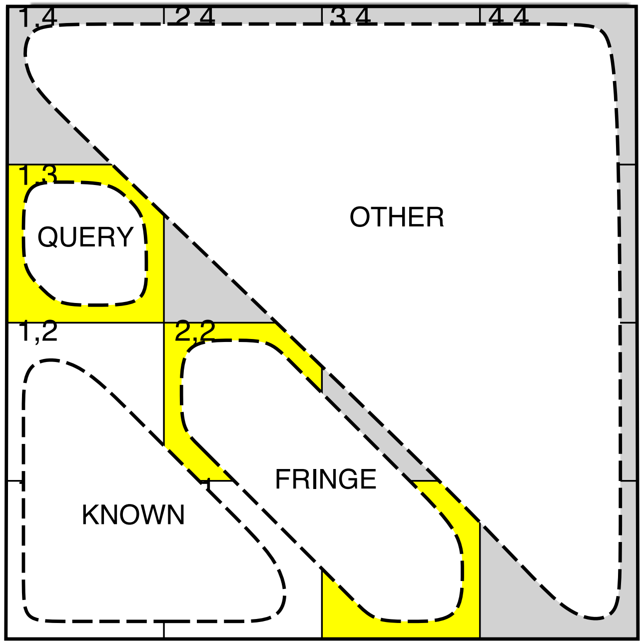
For inference by enumeration we have

$$\mathbf{P}(P_{1,3}|known, b) = \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b)$$

Grows exponentially with the number of squares!

Using conditional independence

Basic insight: observations are conditionally independent of other hidden squares given neighbouring hidden squares



Defining $Unknown = Fringe \cup Other$

$\mathbf{P}(b|P_{1,3}Known, Unknown) = \mathbf{P}(b|P_{1,3}, Known, Fringe)$

Manipulate query into a form where we can use this!

Using conditional independence (2)

$$\mathbf{P}(P_{1,3}|known, b) = \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b)$$

$$= \alpha \sum_{unknown} \mathbf{P}(b|P_{1,3}, known, unknown)\,\mathbf{P}(P_{1,3}, known, unknown)$$

$$= \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe, other)\,\mathbf{P}(P_{1,3}, known, fringe, other)$$

$$= \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe)\,\mathbf{P}(P_{1,3}, known, fringe, other)$$

$$= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}, known, fringe, other)$$

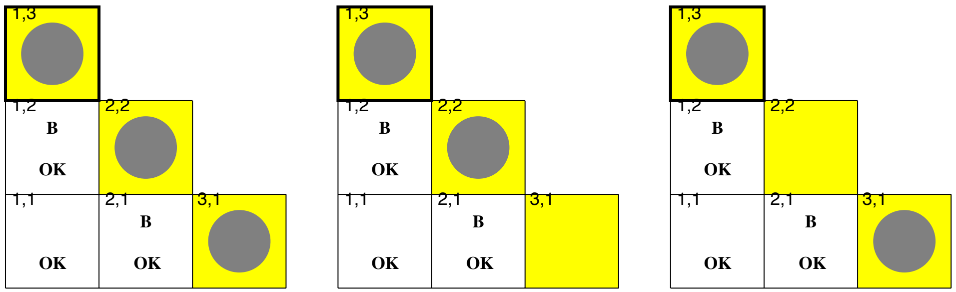
$$= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3})\,P(known)\,P(fringe)\,P(other)$$

$$= \alpha P(known)\,\mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)\,P(fringe) \sum_{other} P(other)$$

$$= \alpha^{'}\mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)\,P(fringe)$$

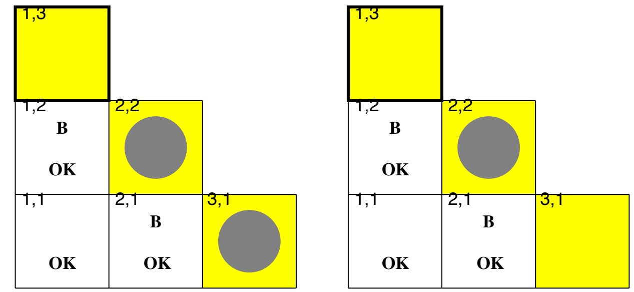
Using conditional independence (3)

a)



0.2 x 0.2 = 0.04 0.2 x 0.8 = 0.16 0.8 x 0.2 = 0.16

b)



0.2 x 0.2 = 0.04 0.2 x 0.8 = 0.16

$$\mathbf{P}(P_{1,3}|known, b) = \alpha' \langle 0.2(0.04 + 0.16 + 0.16),\ 0.8(0.04 + 0.16) \rangle$$
$$\approx \langle 0.31, 0.69 \rangle$$

$$\mathbf{P}(P_{2,2}|known, b) \approx \langle 0.86, 0.14 \rangle$$

Summary

- Probability is a rigorous formalism for uncertain knowledge.
  - Uncertainty arises because of both laziness and ignorance. It is inescapable in complex, dynamic, or inaccessible worlds.
  - Uncertainty means that many of the simplifications that are possible with deductive inference are no longer valid.
  - Probabilities express the agent's inability to reach a definite decision regarding the truth of a sentence, and summarize the agent's beliefs.
  - Basic probability statements include **prior probabilities** and **conditional probabilities** over simple and complex propositions.
  - The **full joint distribution** specifies the probability of each complete assignment of values to random variables. It is usually too large to create or use in its explicit form.

Summary (2)

- The axioms of probability constrain the possible assignments of probabilities onto propositions. An agent that violates the axioms will behave irrationally in some circumstances.
- When the full joint distribution is available, it can be used to answer queries simply by adding up entries for the atomic events corresponding to the query propositions.
- **Absolute independence** between subsets of random variables may allow the full joint to be factored into smaller joint distributions. This may greatly reduce complexity but seldom occurs in practice.
- **Bayes' rule** allows unknown probabilities to be computed from known conditional probabilities, usually in the causal direction. Applying Bayes' rule with many pieces of evidence may in general run into the same scaling problems as the full joint distribution.

Summary (3)

- **Conditional independence** brought about by direct causal relationships in the domain may allow the full joint to be factored into smaller, conditional distributions. The **naive Bayes** model assumes conditional independence of all effect variables given a single cause variable, and grows linearly with the number of effects.
- A wumpus-world agent can calculate probabilities for unobserved aspects of the world and use them to make better decisions than a purely logical agent.