

CS 480

Introduction to Artificial Intelligence

October 27, 2022

Announcements / Reminders

- Follow Week 10 TO DO List

- Grading TA assignment:

https://docs.google.com/spreadsheets/d/1ExS0bKnGt_fdf4LHa3YS1qRA7-lq4xqXVjfSAPMaGVk/edit?usp=sharing

- UPDATED Final Exam date:

- December 1st, 2022 (last week of classes!)
 - Ignore the date provided by the Registrar

Plan for Today

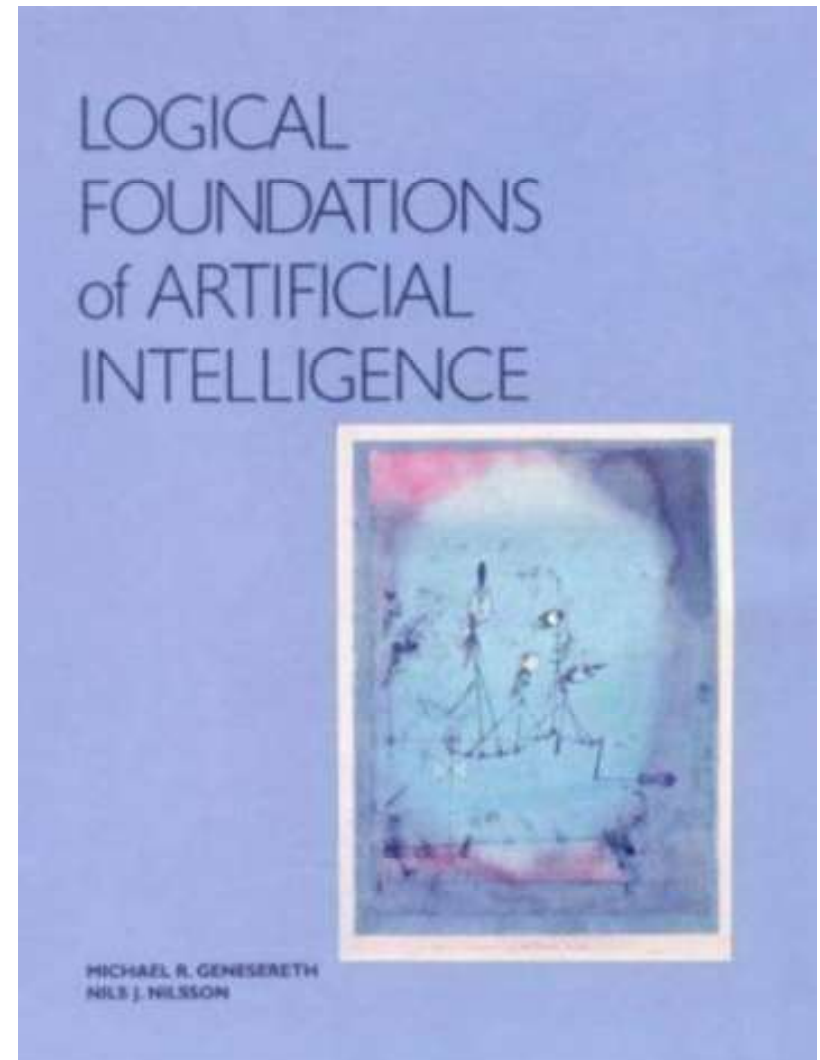
- Quantifying and dealing with uncertainty

If you want more on Logic...

Michael Genesereth, Nils J. Nilsson

“Logical foundations of artificial intelligence”

Elsevier 1978



Probability Theory: Need to Know

- What is an **event** A ?
- What is the **probability of event** A occurring ($P(A)$)?
- What is a **random variable** X ?
- What is the **probability distribution** for X ?
- What is the **probability density function** for X ?
- What are the **expectation** and **variance** of X ?
- Check out <https://seeing-theory.brown.edu/> for a refresher

Probability Theory: Need to Know

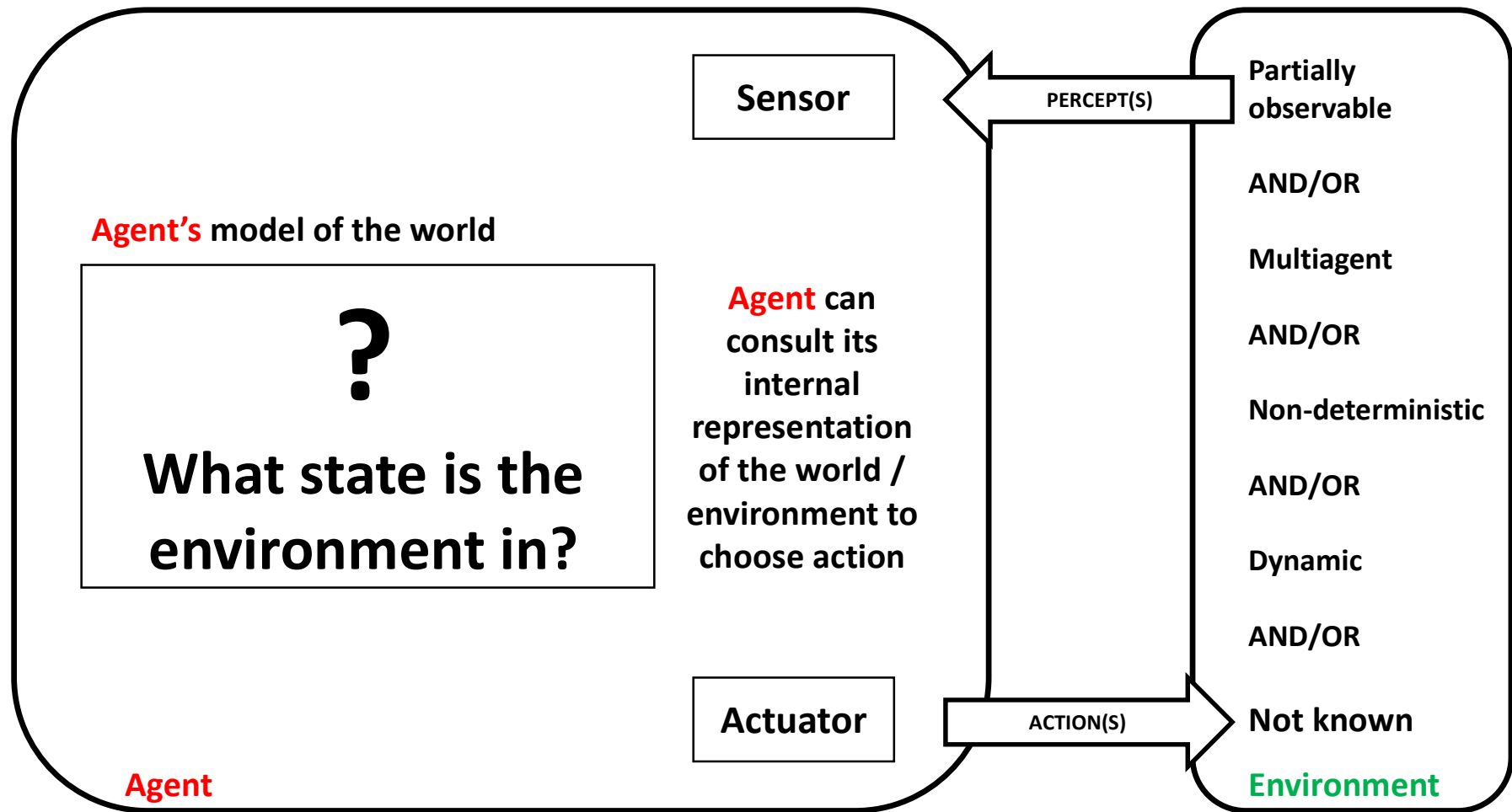
- $P(\text{sure event}) = 1$ and
- $P(\text{impossible event}) = 0$
- **If A, B are exclusive events:** $P(A \vee B) = P(A) + P(B)$
- **If A, B are complementary events:** $P(A) + P(\neg A) = 1$
- **If A, B are arbitrary events:**

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

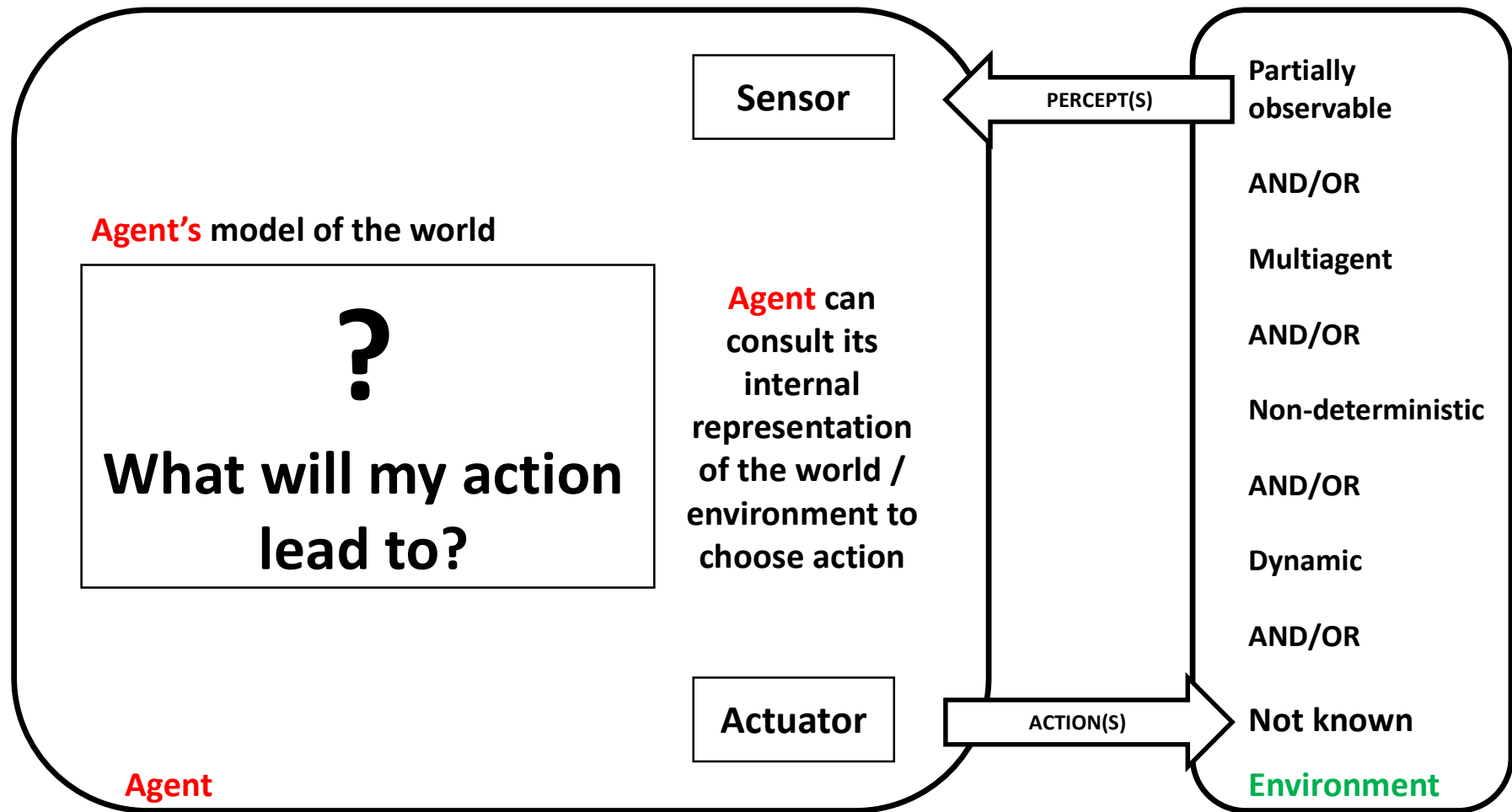
- **If $A \subseteq B$, it is true that $P(A) \leq P(B)$**
- **If A_1, A_2, \dots, A_n are elementary events, then:**

$$\sum_{i=1}^n P(A_i) = 1$$

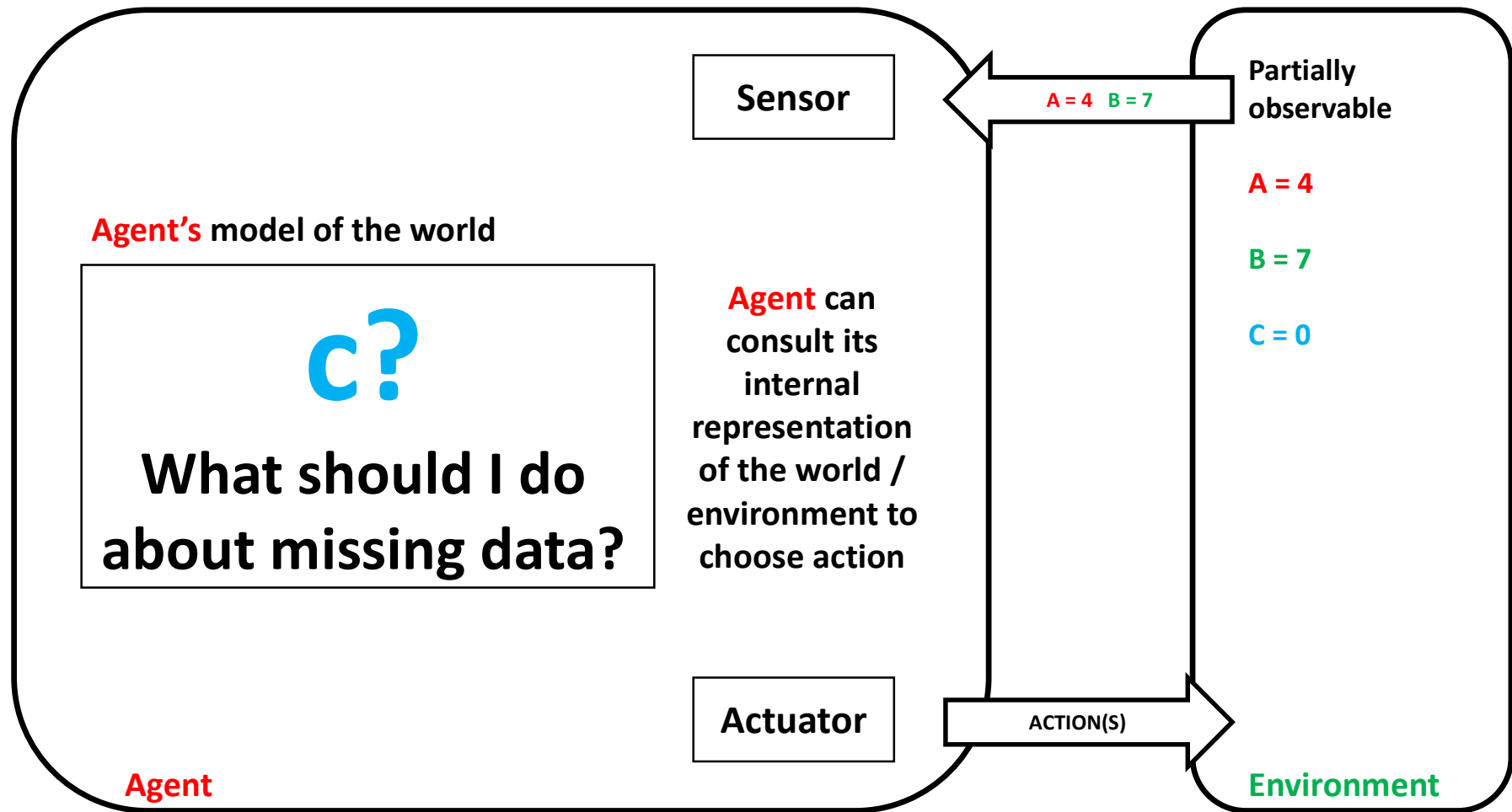
Agents and Uncertainty



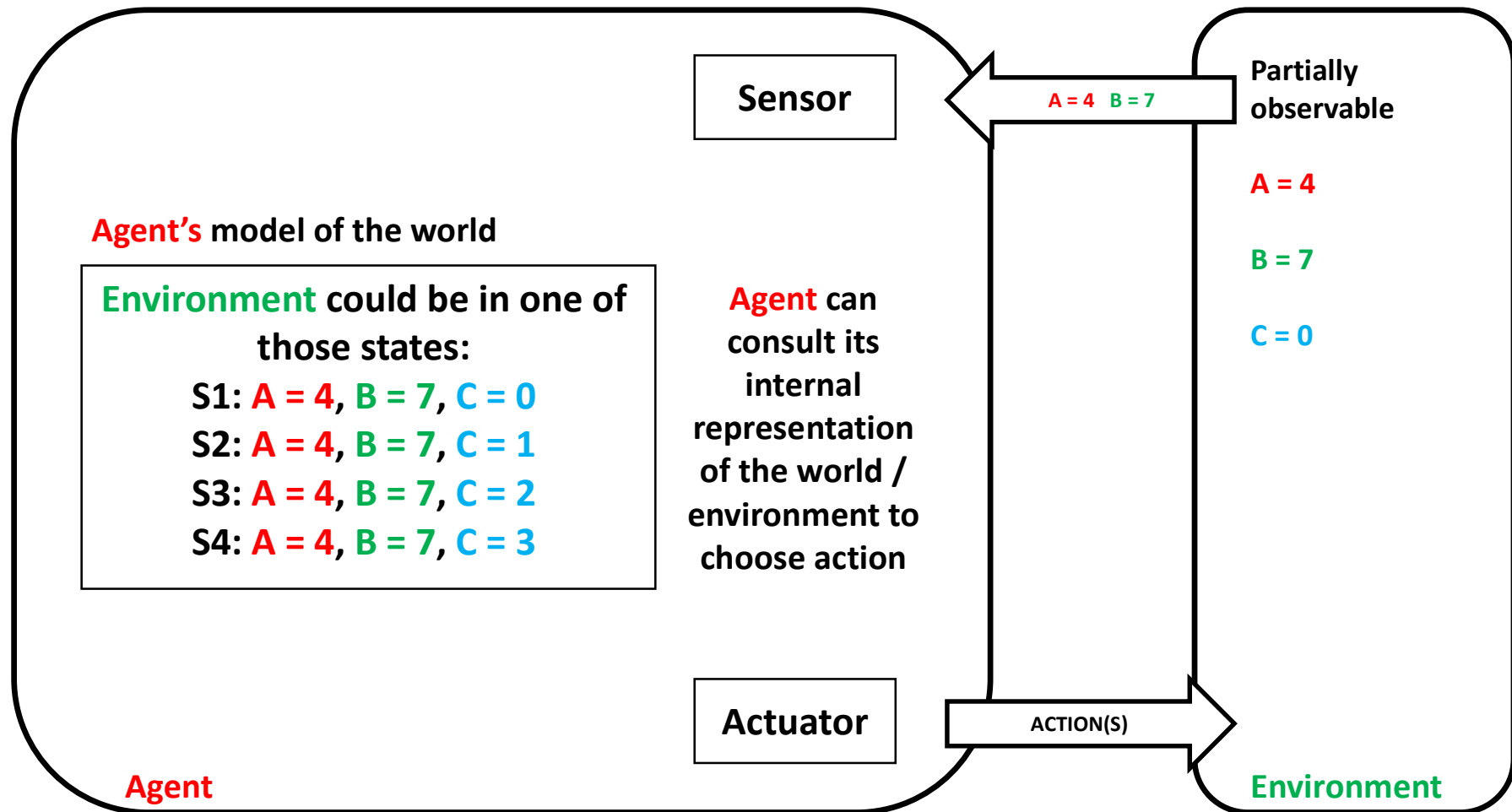
Agents and Uncertainty



Agents and Uncertainty



Agents and Belief State



Assume: $D_c = \{0,1,2,3\}$

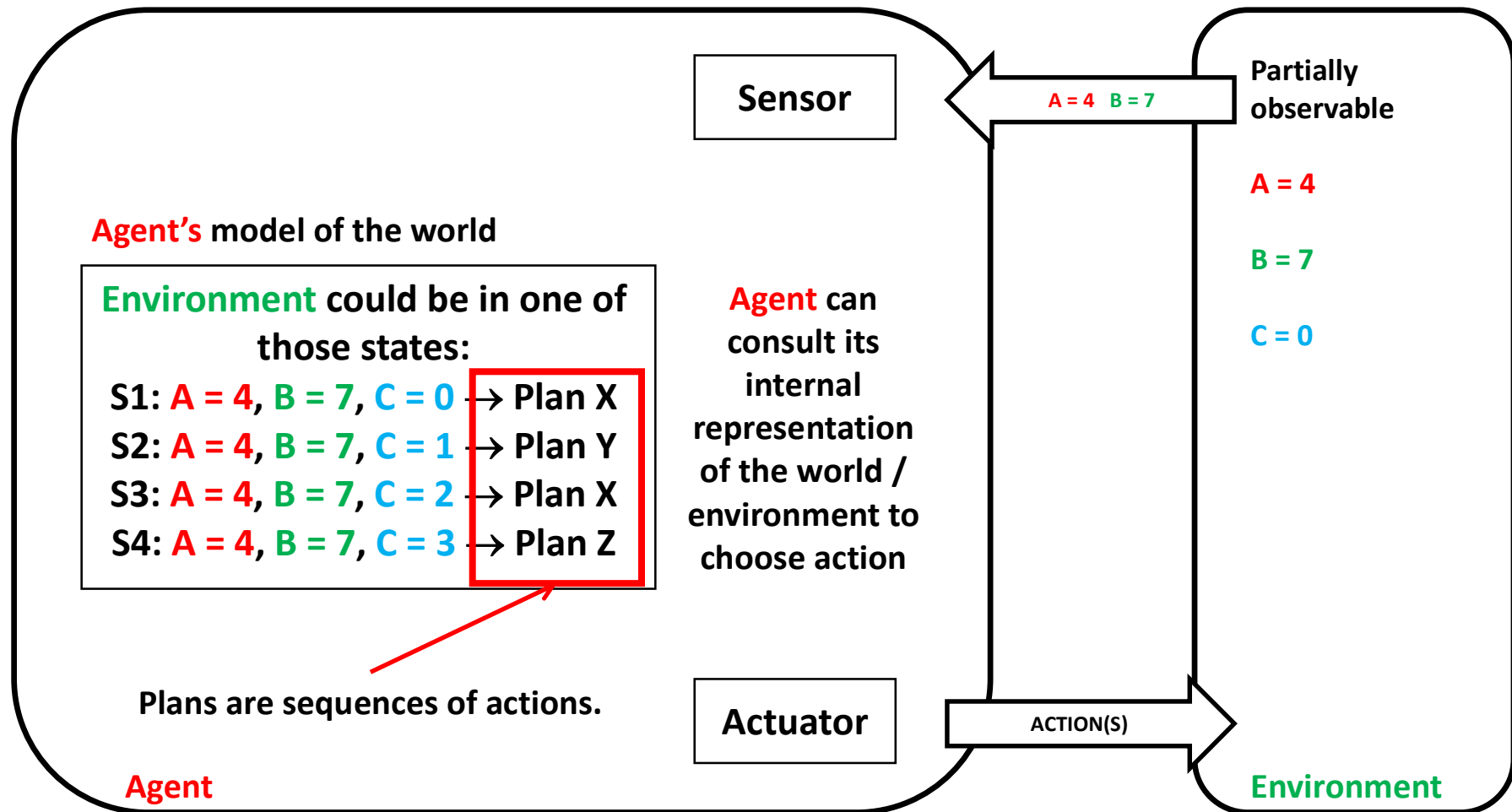
Agent Belief State

Belief state: a set of all possible environment states that the agent can be in and needs to keep track of to handle uncertainty.

Problems:

- agent needs to consider every possible state some are going to be unlikely
- agent needs plans for every eventuality
- there may be no known plan, agent needs to act

Agents and Belief State



Assume: $D_c = \{0,1,2,3\}$

Decision Theory

- **Decisions**: every plan (**actions**) leads to an outcome (state)
- Agents have preferences (**preferred outcomes**)
- Preferences → outcome **utilities**
- Agents have **degrees of belief** (**probabilities**) for actions

Decision theory = **probability theory** + **utility theory**

Frequentist versus Causal Perspective

- **Frequentist view:**

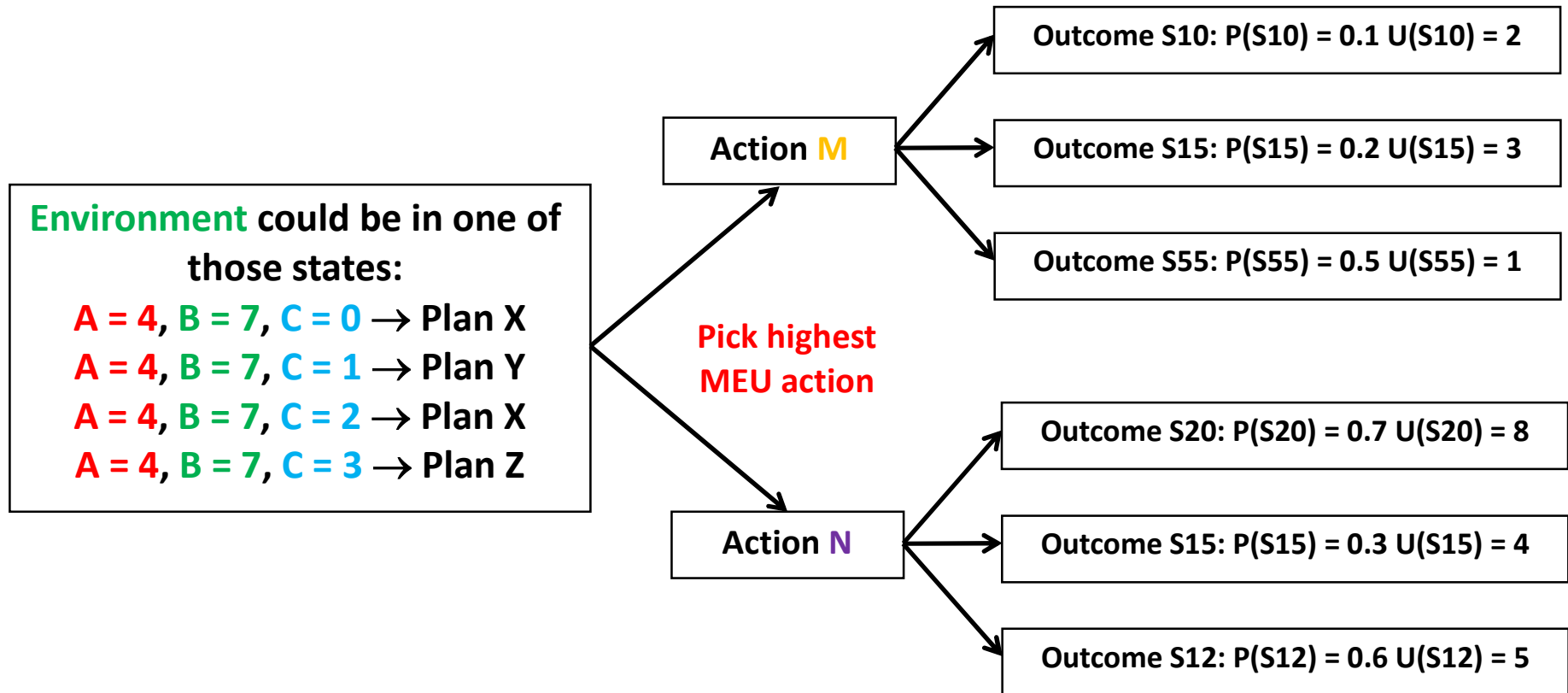
Probability represents long-run frequencies of repeatable events.

- **Causal perspective:**

Probability is a measure of belief.

Maximum Expected (Average) Utility

$$MEU(M) = P(S10) * U(S10) + P(S15) * U(S15) + P(S55) * U(S55)$$



$$MEU(N) = P(S20) * U(S20) + P(S15) * U(S15) + P(S12) * U(S12)$$

Decision-theoretic Agent

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,

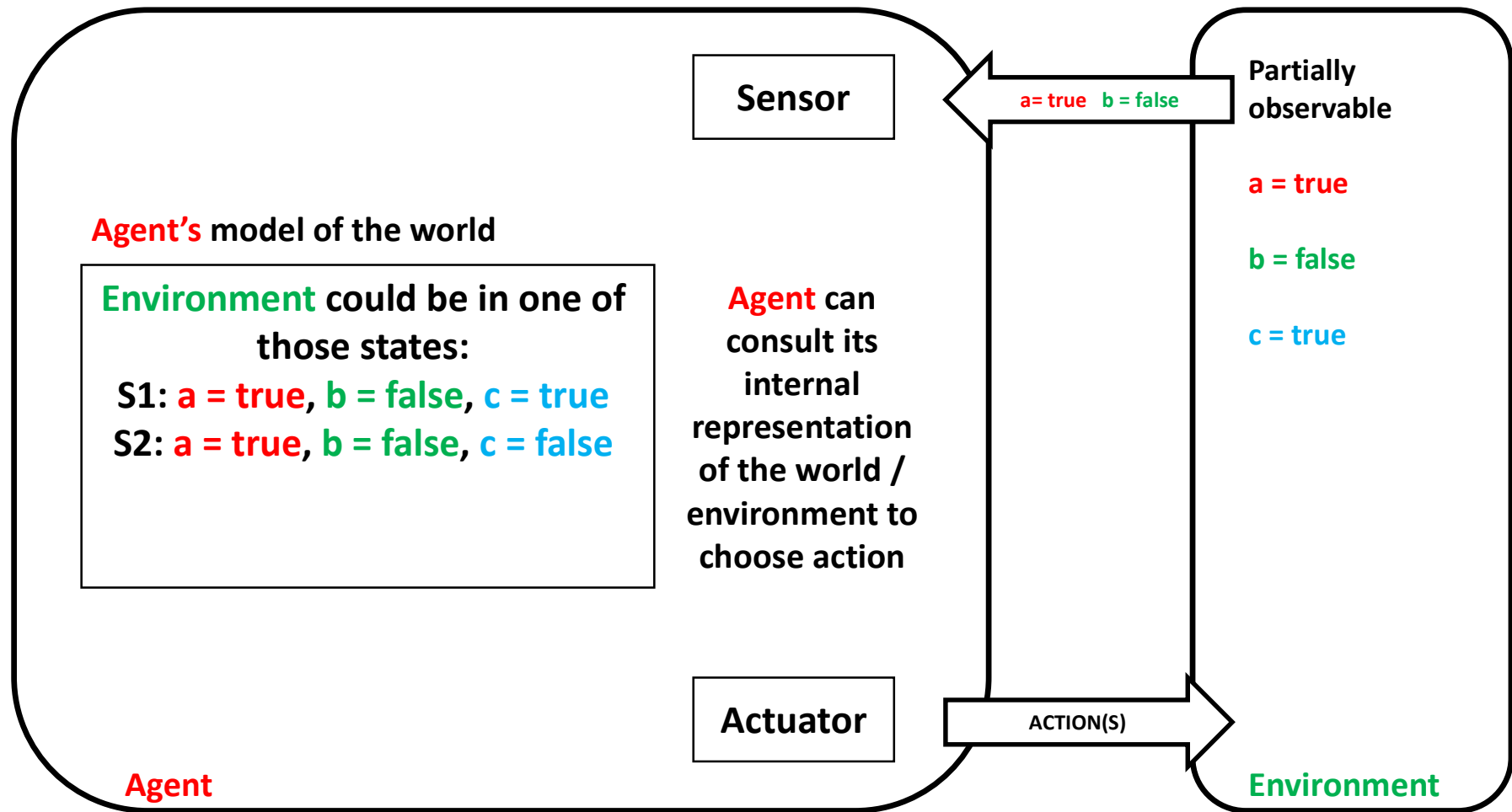
 given action descriptions and current *belief_state*

select *action* with highest expected utility

 given probabilities of outcomes and utility information

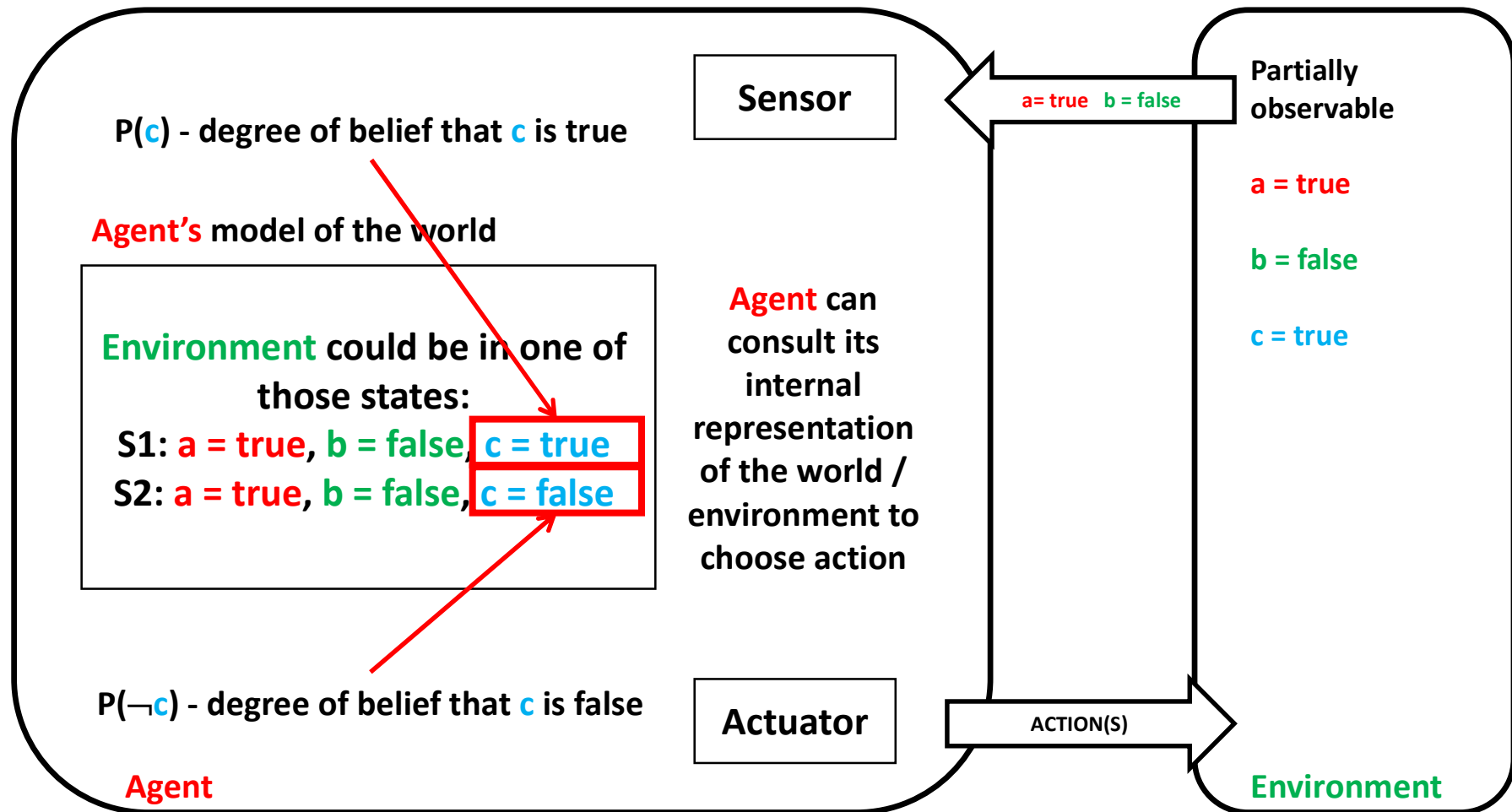
return *action*

Agents and Belief State



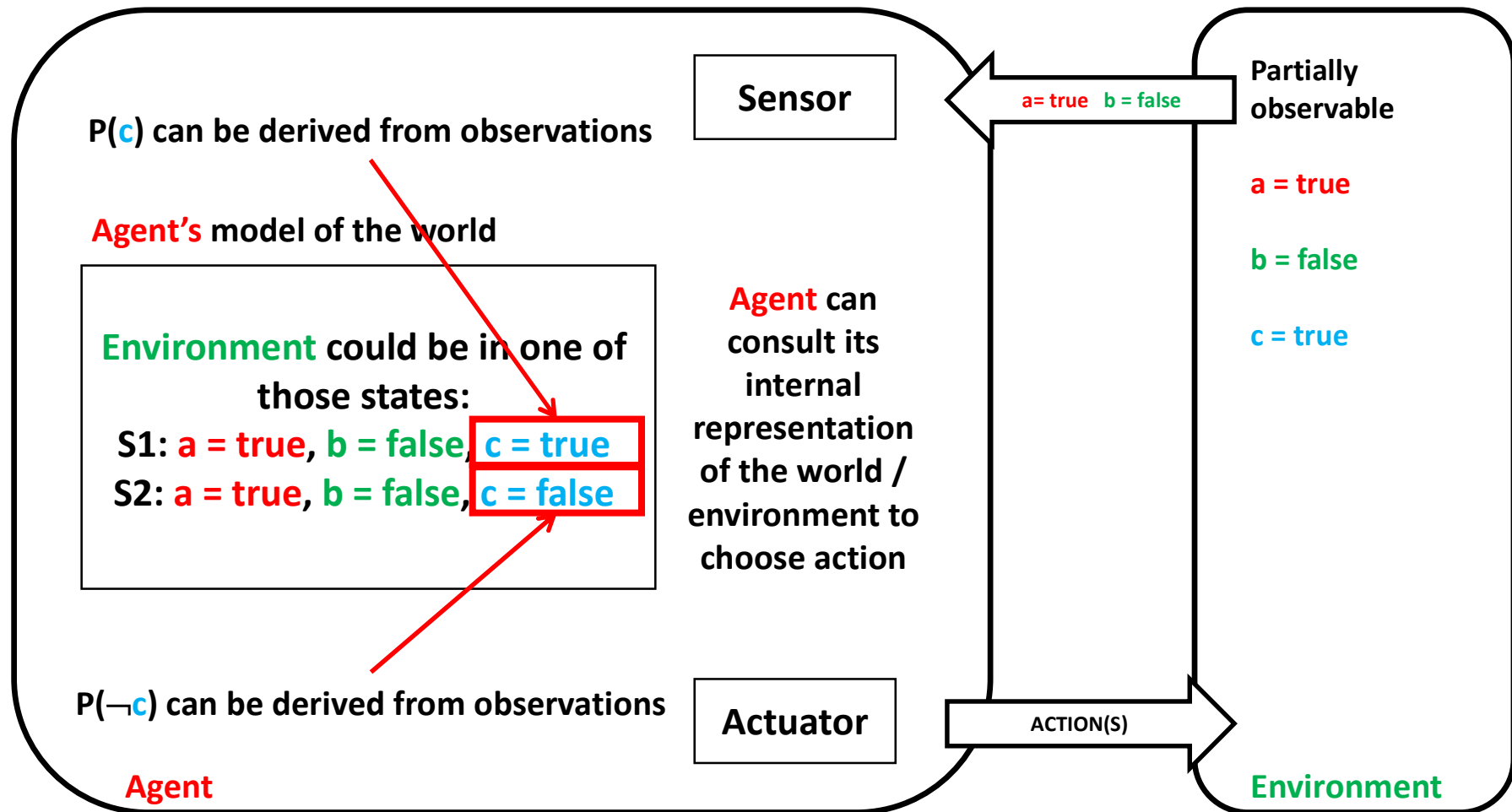
Assume: $D_c = \{\text{true}, \text{false}\}$

Agents and Belief State



Assume: $D_c = \{\text{true}, \text{false}\}$

Agents and Belief State



Assume: $D_c = \{\text{true}, \text{false}\}$

Relationships in Probability Language

- Likelihood:

“Tim is *more likely* to fly than to walk.”

- Conditioning:

“*If* Tim is sick, he can’t fly.”

- Relevance:

“Whether Tim flies *depends on whether* he is sick.”

- Causation:

“Being sick *caused* Tim’s inability to fly.”

Probability Theory and Propositions

Assume that A and B are sentences in propositional logic.

- $P(T) = 1$
- $P(\perp) = 0$
- $P(A \vee B) = P(A) + P(B)$ if $\neg(A \wedge B)$ is a tautology
- $P(A) + P(\neg A) = 1$
- $P(A) = P(B)$ if $(A \Leftrightarrow B)$ is a tautology (logical equivalence)
- $0 \leq P(A)$ for any sentence A

Probability Model

A fully specified probability model associates a numerical probability $P(\omega)$ with each possible world (assume there is a finite number of such worlds):

$$0 \leq P(\omega) \leq 1 \text{ for every } \sum_{\omega \in \Omega} P(\omega) = 1$$

Propositions and Probabilities

The probability associated with a proposition A is defined to be the sum of the probabilities of all the worlds in which it holds:

$$\textit{For any proposition } A, P(A) = \sum_{\omega \in A} P(\omega)$$

Prior (Unconditional) Probabilities

Degree of belief that some proposition A is true *in the absence of any other related information* is called **unconditional** or **prior probability** (or “prior” for short) $P(A)$.

Examples:

$P(\text{isRaining})$

$P(\text{dieRoll} = 5)$

$P(\text{CS480FinalGrade} = \text{'A'})$

$P(\text{toothache})$

Conditioning

Conditioning is a process of revising beliefs based on new evidence e :

- start by taking all background information (**prior probabilities**) into account
- if new evidence e is acquired, a conditional probability of some proposition A given evidence e can be calculated (**posterior probability**): $P(A | e)$

Posterior (Conditional) Probabilities

Typically, there is going to be some information, called **evidence** e , that affects our degree of belief about some proposition A being true. This allows us to also consider **conditional** or **posterior probability** (or “posterior” for short) $P(A \mid e)$.

Examples ($P(A \text{ given } e)$):

$$P(\text{isRaining} \mid \text{cloudy})$$

$$P(\text{CS480FinalGrade} = \text{'A'} \mid \text{CS480PA1Score} > 80)$$

$$P(\text{cavity} \mid \text{toothache})$$

Evidence e

Evidence e rules out possible worlds incompatible with e .

Prior vs. Posterior Probabilities

Prior Probability



$$P(A)$$

BTW: it is also $P(A | T)$

Posterior Probability



$$P(A | e)$$

Conditional Probability

$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)}$$

where $P(B) > 0$

Conditional Probability

$$P(A \mid \textit{evidence}) = \frac{P(A \wedge \textit{evidence})}{P(\textit{evidence})}$$

where $P(\textit{evidence}) > 0$

Conditional Probability (Product Rule)

$$P(A \wedge B) = P(A | B) * P(B)$$

Conditional Probability (Product Rule)

$$P(A \wedge \textit{evidence}) = P(A \mid \textit{evidence}) * P(\textit{evidence})$$

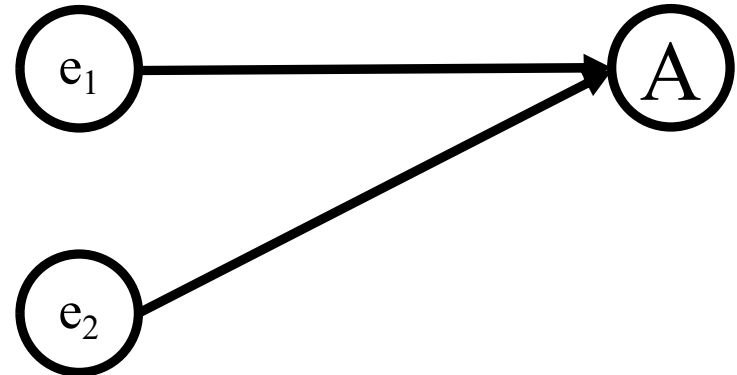
Prior vs. Posterior Probabilities

Prior Probability



$$P(A)$$

Posterior Probability



$$P(A \mid e_1 \wedge e_2)$$

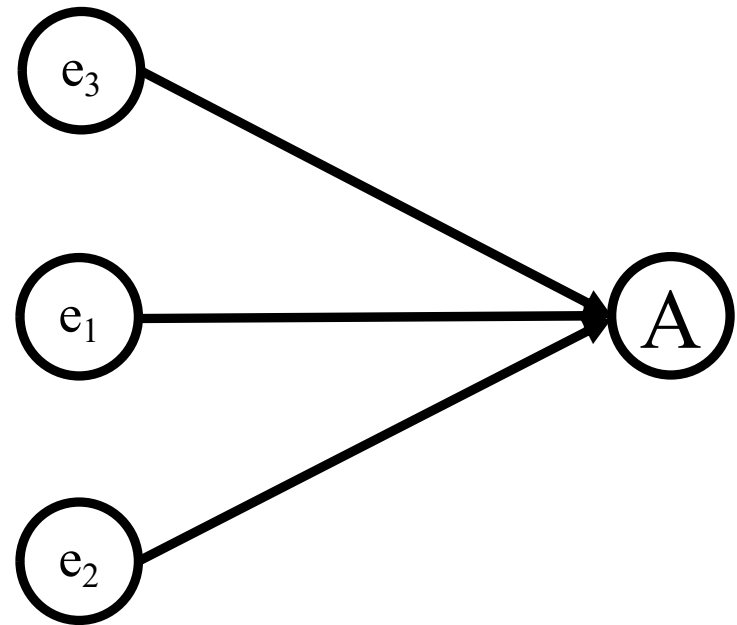
Prior vs. Posterior Probabilities

Prior Probability



$$P(A)$$

Posterior Probability



$$P(A \mid e_1 \wedge e_2 \wedge e_3)$$

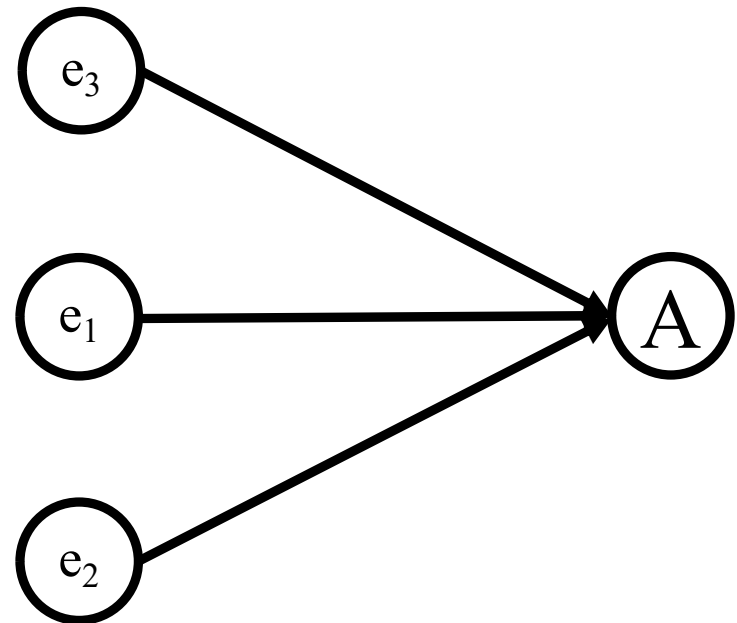
Prior vs. Posterior Probabilities

Prior Probability



$P(A)$

Posterior Probability



$P(A \mid \text{parents}(A))$

Marginal Probability

Marginal probability: the probability of an event occurring $P(A)$.

It may be thought of as an unconditional probability.

It is not conditioned on another event.

Joint Probability

The probability of event A and event B occurring (or more than two events). It is the probability of the intersection of two or more events.

$$P(A \wedge B)$$

For any propositions f_1, f_2, \dots, f_n :

$$P(f_1 \wedge f_2 \wedge \dots \wedge f_n)$$

Chain Rule

Conditional probabilities can be used to decompose conjunctions using the chain rule. For any propositions f_1, f_2, \dots, f_n :

$$\begin{aligned} P(f_1 \wedge f_2 \wedge \dots \wedge f_n) &= \\ P(f_1) &* \\ P(f_2 \mid f_1) &* \\ P(f_3 \mid f_1 \wedge f_2) &* \\ \dots & \\ P(f_n \mid f_1 \wedge \dots \wedge f_{n-1}) &= \\ = \prod_{i=1}^n P(f_i \mid f_1 \wedge \dots \wedge f_{i-1}) \end{aligned}$$

Chain Rule

Conditional probabilities can be used to decompose conjunctions using the chain rule. For any propositions f_1, f_2, \dots, f_n :

$$\begin{aligned} P(f_1 = x_1 \wedge f_2 = x_2 \wedge \dots \wedge f_n = x_n) &= \\ P(f_1 = x_1) &* \\ P(f_2 \mid f_1 = x_1) &* \\ P(f_3 \mid f_1 = x_1 \wedge f_2 = x_2) &* \\ \dots & \\ P(f_n = x_n \mid f_1 = x_1 \wedge \dots \wedge f_{n-1} = x_{n-1}) &= \\ = \prod_{i=1}^n P(f_i = x_i \mid f_1 = x_1 \wedge \dots \wedge f_{i-1} = x_{i-1}) \end{aligned}$$

Bayes' Rule

$$P(A | B) = \frac{P(B | A) * P(A)}{P(B)}$$

Bayes' Rule

$$P(\textit{cause} \mid \textit{effect}) = \frac{P(\textit{effect} \mid \textit{cause}) * P(\textit{cause})}{P(\textit{effect})}$$

Bayes' Rule

$P(\textit{cause} \mid \textit{effect})$ diagnostic direction relation

$$P(\textit{cause} \mid \textit{effect}) = \frac{P(\textit{effect} \mid \textit{cause}) * P(\textit{cause})}{P(\textit{effect})}$$

$P(\textit{effect} \mid \textit{cause})$ causal direction relation

Bayes' Rule

$P(\textit{disease} \mid \textit{symptoms})$ diagnostic direction relation

$$P(\textit{disease} \mid \textit{symptoms}) = \frac{P(\textit{symptoms} \mid \textit{disease}) * P(\textit{disease})}{P(\textit{symptoms})}$$

$P(\textit{symptoms} \mid \textit{disease})$ causal direction relation

Bayes' Rule

$$P(\textit{cause} \mid \textit{effect}) = \frac{P(\textit{effect} \mid \textit{cause}) * P(\textit{cause})}{P(\textit{effect})}$$

Problem: a single card is drawn from a standard deck of cards. What is the probability that we **drew a queen** if we **know that a face card (J, Q, K) was drawn**?

$$P(\textit{queen} \mid \textit{face}) = \frac{P(\textit{face} \mid \textit{queen}) * P(\textit{queen})}{P(\textit{face})}$$

$$P(\textit{queen} \mid \textit{face}) = \frac{1 * 4 / 52}{12 / 52} = \frac{1}{3}$$

Bayes' Rule

$$P(\textit{cause} \mid \textit{effect}) = \frac{P(\textit{effect} \mid \textit{cause}) * P(\textit{cause})}{P(\textit{effect})}$$

Problem: Calculate probability that **a patient has meningitis if a patient has stiff neck**. Meningitis is a cause of neck stiffness in 70% of cases, probability of having meningitis is 1/50000. Stiff neck happens to 1% of patients.

$$P(\textit{m} \mid \textit{s}) = \frac{P(\textit{s} \mid \textit{m}) * P(\textit{m})}{P(\textit{s})}$$

$$P(\textit{m} \mid \textit{s}) = \frac{0.7 * 1/50000}{0.01} = 0.0014$$

Independence

Assume that the knowledge of the truth of one proposition Y , does not affect the agent's belief in another proposition, X , in the context of other propositions Z . We say that X is **independent** of Y given Z .

Conditional Independence

Random variable X is **conditionally independent** of random variable Y given Z if for all $x \in D_X$, for all $y \in D_Y$, and for all $z \in D_Z$, such that

$$P(Y = y \wedge Z = z) > 0 \text{ and } P(Y = y' \wedge Z = z) > 0$$

$$P(X = x \mid Y = y \wedge Z = z) = P(X = x \mid Y = y' \wedge Z = z)$$

In other words, given a value of Z , knowing Y 's value **DOES NOT** affect your belief in the value of X .

Conditional Independence

The following four statements are equivalent as long as conditional probabilities:

1. X is conditionally independent of Y given Z
2. Y is conditionally independent of X given Z
3. $P(X | Y, Z) = P(X | Z)$
4. $P(X, Y | Z) = P(X | Z) * P(Y | Z)$

Bayesian (Belief) Network

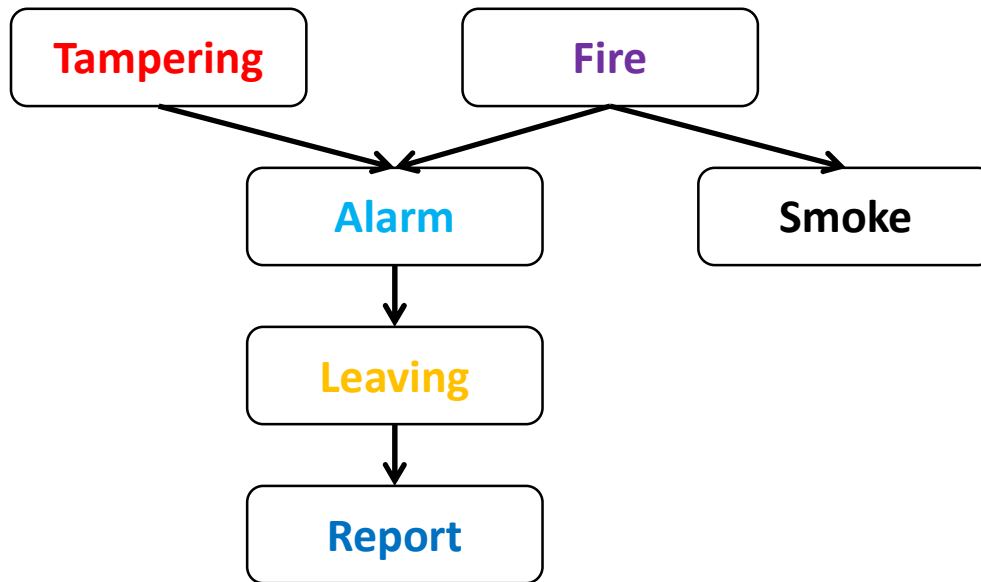
A Bayesian belief network describes the joint probability distribution for a set of variables.

A Bayesian network is **an acyclic, directed graph (DAG)**, where the nodes are random variables (propositions). There is an edge (arc) from each elements of $\text{parents}(X_i)$ into X_i . Associated with the Bayesian network is a set of conditional probability distributions - the conditional probability of each variable given its parents (which includes the prior probabilities of those variables with no parents).

Consists of:

- a graph (DAG) with nodes corresponding to random variables
- a domain for each random variable
- a set of conditional probability distributions $P(X_i \mid \text{parents}(X_i))$

Bayesian (Belief) Network: Example



Random Variables (Propositions):

- **Tampering**: true if the alarm is tampered with
- **Fire**: true if there is a fire
- **Alarm**: true if the alarm sounds
- **Smoke**: true if there is smoke
- **Leaving**: true if people leaving the building at once
- **Report**: true if someone who left the building reports fire

Domain for all variables: {true, false}

NOTE: RVs don't have to be Boolean