Causality

CMPUT 261: Introduction to Artificial Intelligence

Bar §3.4

Lecture Outline

1. Recap
2. Causality Introduction
3. Causal Queries

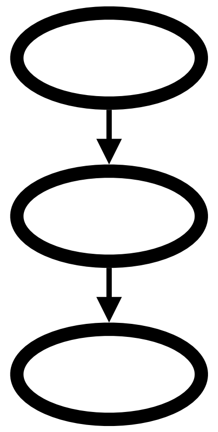
After this lecture, you should be able to:

- distinguish between an observation and an intervention
- construct the post-intervention distribution for a causal query
- evaluate a causal query on a given causal network
- justify whether a causal model is valid
- define selection effect

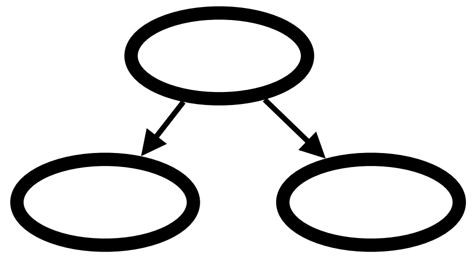
Recap: Independence in a Belief Network

Belief Network Independence Property:
Every node is **independent** of its **non-descendants**, **conditional only** on its **parents**

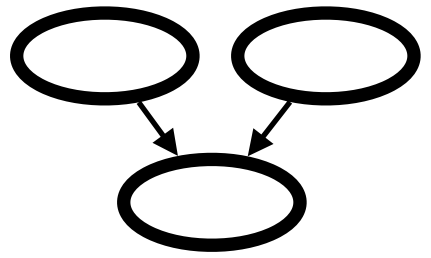
Patterns of dependence:



1. **Chain:** **Ends** are **not marginally** independent, but **conditionally** independent given middle



2. **Common ancestor:** **Descendants** are **not marginally** independent, but **conditionally** independent given ancestor



3. **Common descendant:** **Ancestors** are **marginally** independent, but **not conditionally** independent given descendant

Recap: Variable Elimination

1. Condition on observations by **conditioning**
2. Construct joint distribution factor by **multiplication**
3. Remove non-query, non-observed variables by **summing out**
4. **Normalize** at the end

Interleaving order of sums and products can improve **efficiency**:

$$\sum_A \sum_E f_1(Q, A, B, C) \times f_2(C, D, E)$$ **112** computations

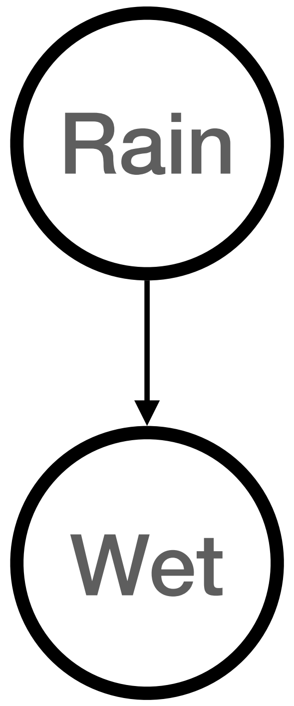
$$= \left(\sum_A f_1(Q, A, B, C) \right) \times \left(\sum_E f_2(C, D, E) \right)$$ **28** computations

Causality Introduction: A Tale of Two Belief Networks

$P(\text{Rain} = T) = 0.5$ (rows 1–2)

$P(\text{Rain} = F) = 0.5$ (rows 3–4)

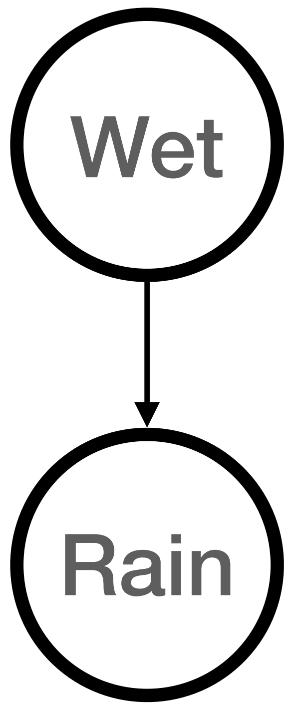
Rain	Wet	P(Rain, Wet)
F	T	0.125
F	F	0.375
T	T	0.45
T	F	0.05



Natural network

- Two different ways to **factor** the joint distribution between whether the sidewalk is **Wet** and whether it is **Raining:**

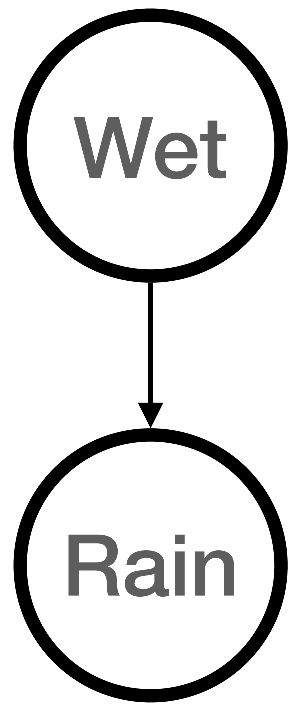
$$P(\text{Rain}, \text{Wet}) = P(\text{Wet} \mid \text{Rain})P(\text{Rain})$$
$$= P(\text{Rain} \mid \text{Wet})P(\text{Wet})$$



Inverted network

- Each factorization corresponds to a different **Belief Network**

The Inverted Network Isn't Crazy

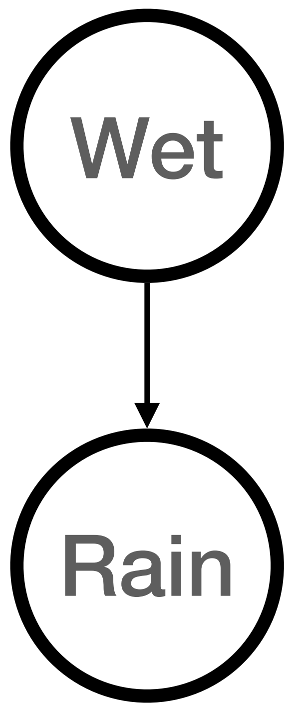


Inverted network

Corresponds to the factoring $P(Rain \mid Wet)P(Wet)$

- Sometimes you want to answer the question
Given that I observe that the sidewalk is Wet, what is the probability that it is currently Raining?
  - This is just updating our confidence in a hypothesis (it is **Raining**) given our observations (**Wet** sidewalk)
- Could **preprocess** the natural network into this form to avoid having to do a lot of computations with **Bayes' Rule**

The Inverted Network Is Crazy



Inverted network

Corresponds to the factoring $P(Rain \mid Wet)P(Wet)$

- **If I cause my sidewalk to be Wet (by throwing water on it), what is the probability that it is Raining?**
    - So, condition on **Wet**=true
    - This network seems to imply that it will be $P(Rain \mid Wet = True) = .78 > P(Rain) = .5$
    - .... wait, what?
- **Question:** What is going wrong in this example?

Wet	P(Wet)
T	0.575
F	0.425

Rain	Wet	P(Rain | Wet)
F	F	0.88
T	F	0.12
F	T	0.22
T	T	0.78

Observations vs. Interventions

- The semantics of Belief Networks are defined for **observational** questions
  - They don't directly model **causal** questions
  - In fact, in our Rainy Sidewalk example, we would get **exactly the same** (crazy) answer to our causal question from querying the **natural network**
- The joint distribution represented by the networks **doesn't model** the situation in which I **intervene**
  - Adding a variable **James_Throws_Water** to the distribution

Observations vs. Interventions: Examples

- *Observation:* If I **observe** soil pH is high, will this plant grow more?
- *Intervention:* If I **make the pH** of the soil high, will this plant grow more?
- *Observation:* If I **observe** that the clock says 11pm, will it be dark out when I look out the window?
- *Intervention:* If I **set the clock** to 11pm, will it be dark out when I look out the window?
- *Observation:* In the past, did students who **volunteered** for extra tutoring get higher grades than those who didn't?
- *Intervention:* If I **assign** extra tutoring to a student, will they get a higher grade than if I don't?

Simpson's Paradox

Suppose we have information from two trials of a new drug:
One on young test subjects, and one on old test subjects.

A - age
D - received drug
R - recovered

A	D	R	count	P(A,D,R)
Y	T	T	18	0.225
Y	T	F	12	0.15
Y	F	T	7	0.0875
Y	F	F	3	0.0375
O	T	T	2	0.025
O	T	F	8	0.1
O	F	T	9	0.1125
O	F	F	21	0.2625

- Is the drug **effective for young patients**?

$$P(R = true \mid D = true, A = young) = 0.60$$
$$P(R = true \mid D = false, A = young) = 0.70$$

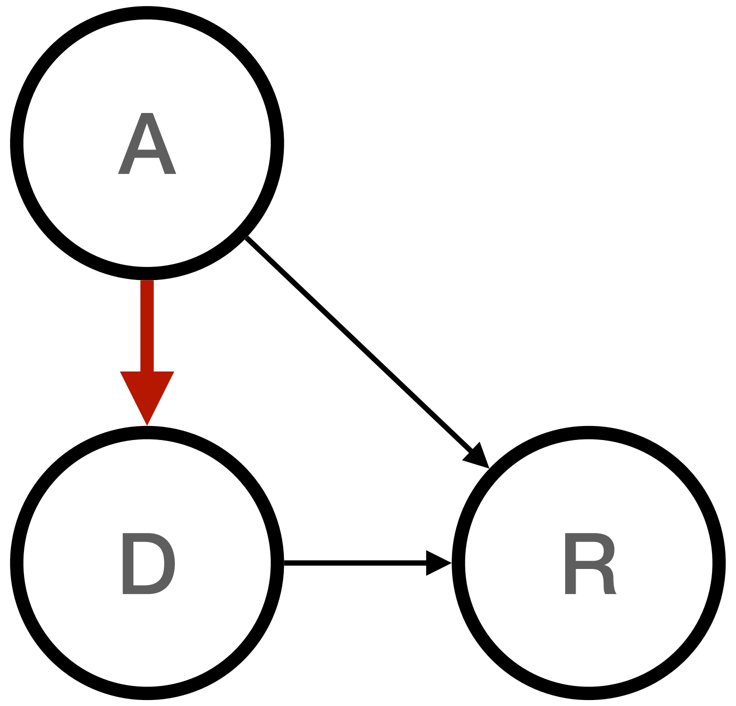
- Is the drug **effective for old patients**?

$$P(R = true \mid D = true, A = old) = 0.20$$
$$P(R = true \mid D = false, A = old) = 0.30$$

- Is the drug **effective**?

$$P(R = true \mid D = true) = 0.50$$
$$P(R = true \mid D = false) = 0.40$$

Simpson's Paradox, explained



- The joint distribution factors as
  $P(A, D, R) = P(A) \times P(D \mid A) \times P(R \mid D, A)$

- **Per-age** queries are answered **directly** by $P(R \mid D, A)$

- For the **overall query**, we want $P(R \mid D) = \dfrac{\sum_A P(R \mid A, D) P(A)}{\sum_{A,R} P(R \mid A, D) P(A)}$

- But that's not how the distribution factors. If we follow the factoring above, we will instead compute

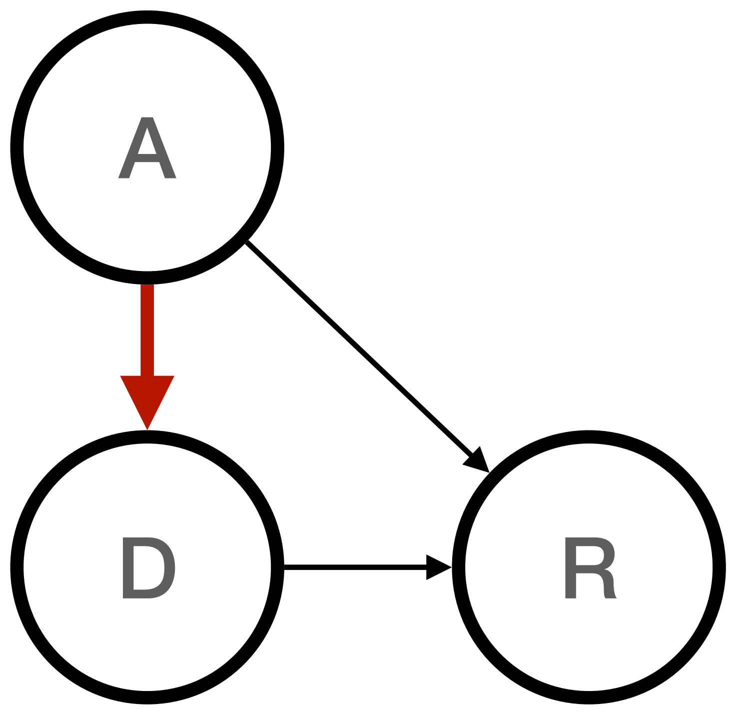
$$P(R \mid D) = \frac{P(R, D)}{P(D)} = \frac{\sum_A P(A, D, R)}{\sum_{A,R} P(A, D, R)} = \frac{\sum_A P(R \mid A, D) \boxed{P(D \mid A)} P(A)}{\sum_{A,R} P(R \mid A, D) \boxed{P(D \mid A)} P(A)}$$

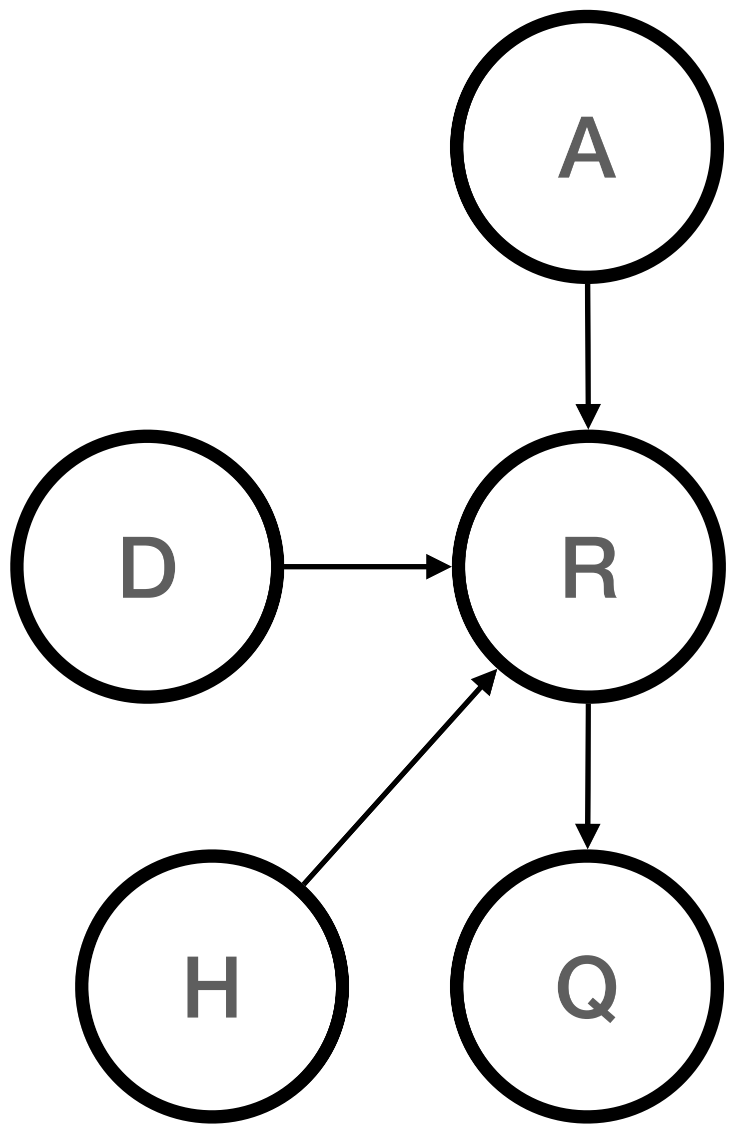
  - In our dataset, knowing whether a subject **got the drug** tells you something about their **age**, and younger patients have a **higher overall recovery** rate than older patients

  - $P(R \mid A = young) = 0.625$ vs $P(R \mid A = old) = 0.275$

Selection Bias

- This problem is an example of **selection bias**
- Whether subjects received treatment is **systematically related** to their **response** to the treatment
- This is why **randomized trials** are the gold standard for causal questions:
  - The only thing that determines whether or not a subject is treated is a **random number**
  - Random number is definitely independent of **anything else** (including **response** to treatment)





Why do we need this?

The problem we're trying to solve here:

- Given an **observational** dataset (which may include selection bias), answer **causal** questions (i.e., questions about the results of **interventions**)

Question: If randomized trials never have selection bias, why would we even want to do this? (i.e., why not always just use randomized trials?)

1. *Practicality:* Sometimes it is **impossible**. (There's only one national economy, so you cannot raise taxes and also not raise taxes)

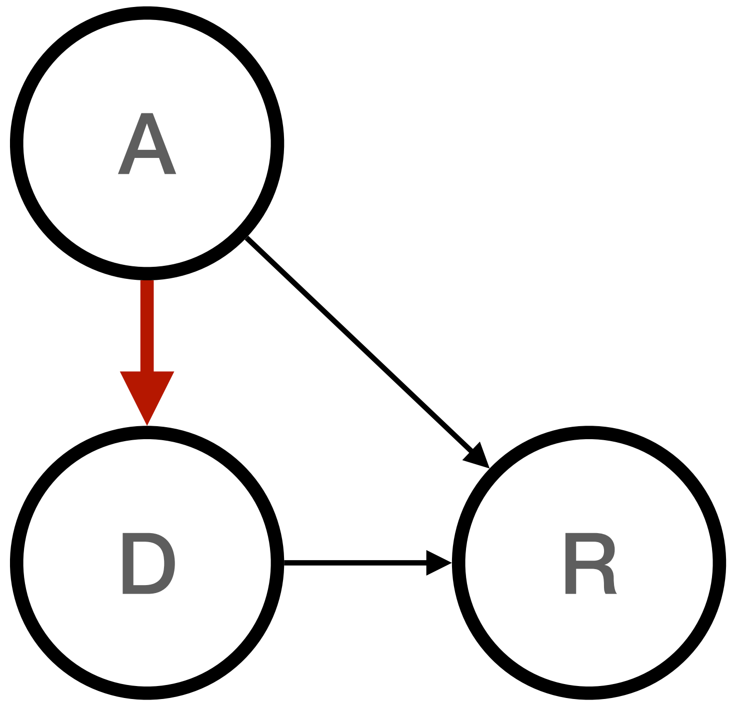
2. *Ethics:* Sometimes it is **immoral**. (It would be unfair to randomly assign different prison sentences to different convicts)

Post-Intervention Distribution

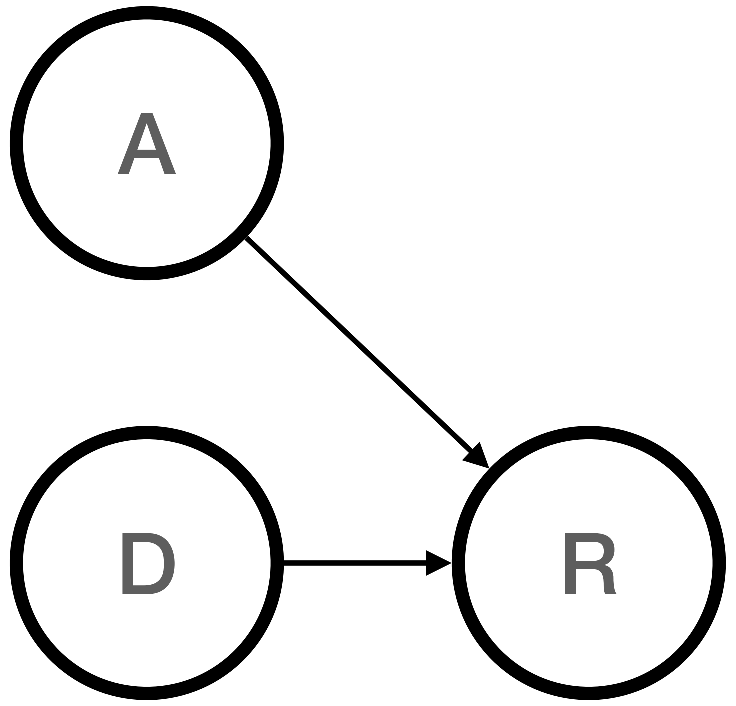
- The causal query is really a query on a **different distribution** in which we have **forced** $D = true$
    - Different from the original joint distribution **conditioned** on **observing** that $D = true$
    - We will refer to the two distributions as the **observational** distribution and the **post-intervention** distribution
- With a post-intervention distribution, we can compute the answers to causal queries using **existing techniques** (e.g., variable elimination)

Post-Intervention Distribution for Simpson's Paradox

- **Observational distribution:**
  $P(A, D, R) = P(A) \times P(D \mid A) \times P(R \mid D, A)$



- **Question:** What is the **post-intervention distribution** for Simpson's Paradox?
  - We're **forcing** $D = true$, so $P(D = true \mid A = a) = 1$ for all $a \in \mathrm{dom}(A)$
  - That's the same as just omitting the $P(D \mid A)$ factor



- **Post-intervention distribution:**
  $P(G, D, R) = P(A) \times P(D) \times P(R \mid D, A)$

The Do-Calculus

- How should we **express** causal queries?
- One approach: The **do-calculus**
- Condition on **observations**:
  $P(Y \mid X = x)$
- Express **interventions** with special **do** operator:
  $P(Y \mid do(X = x))$
- Allows us to **mix** observational and interventional information:
  $P(Y \mid Z = z, do(X = x))$

Evaluating Causal Queries With the Do-Calculus

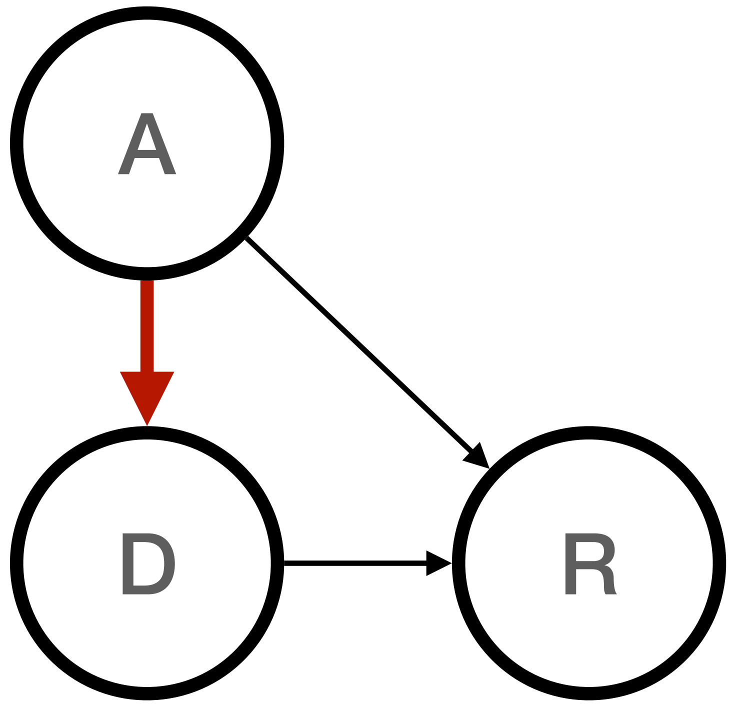
Given a query $P(Y \mid do(X = x), Z = z)$:

1. Construct **post-intervention** distribution $\hat{P}$ by **removing** all links from X's direct parents to X
2. Evaluate the **observational** query $\hat{P}(Y \mid X = x, Z = z)$ in the **post-intervention distribution**

Example: Simpson's Paradox

- **Observational distribution:** $P(A, D, R) = \times P(A) \times P(D \mid A) \times P(R \mid A, D)$
- **Observational query:**

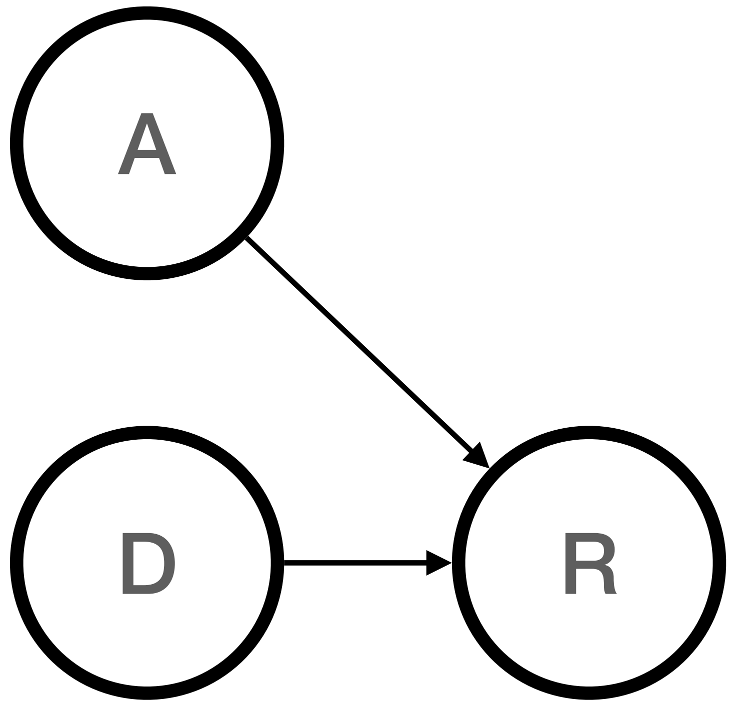
$$P(R|D) = \frac{P(R, D)}{P(D)} = \frac{\sum_A P(A, D, R)}{\sum_{A,R} P(A, D, R)} = \frac{\sum_A P(R|D, A)P(D|A)P(A)}{\sum_{A,R} P(R|D, A)P(D|A)P(A)}$$



- Observational query values: $P(R \mid D = true) = 0.50 \quad P(R \mid D = false) = 0.40$
- **Post-intervention distribution** for causal query $P\left(R \mid do(D = true)\right)$:
  $\hat{P}(A, D, R) = P(R \mid D, A) \times P(A)$
- **Causal query:**

$$P(R|do(D = true)) = \hat{P}(R|D = true) = \frac{\sum_A P(R|D, A)P(A)}{\sum_{A,R} P(R|D, A)P(A)}$$

- Causal query values:
  $P\left(R \mid do(D = true)\right) = 0.40 \quad P\left(R \mid do(D = false)\right) = 0.50$

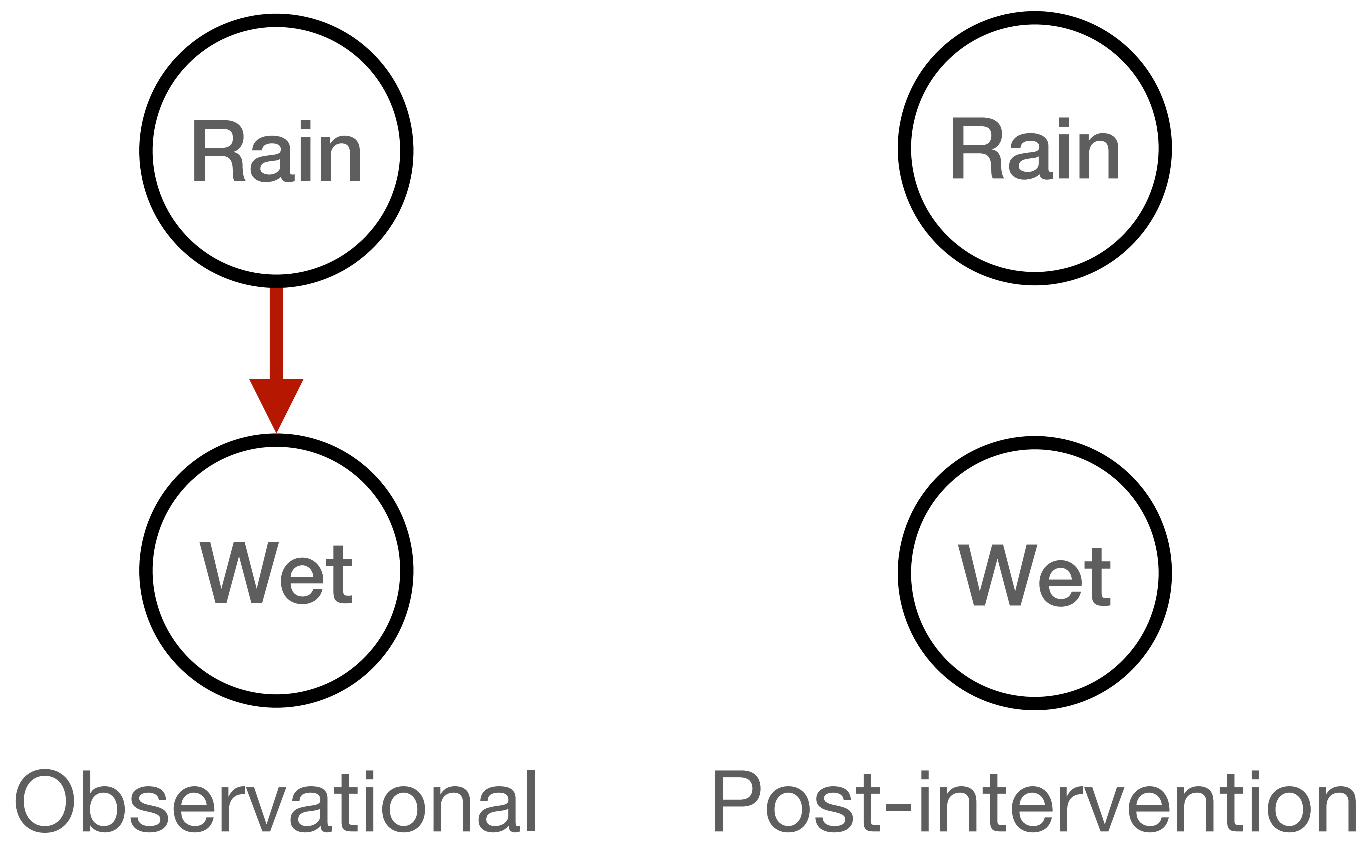


Example: Rainy Sidewalk

Query: $P(Rain \mid do(Wet = true))$

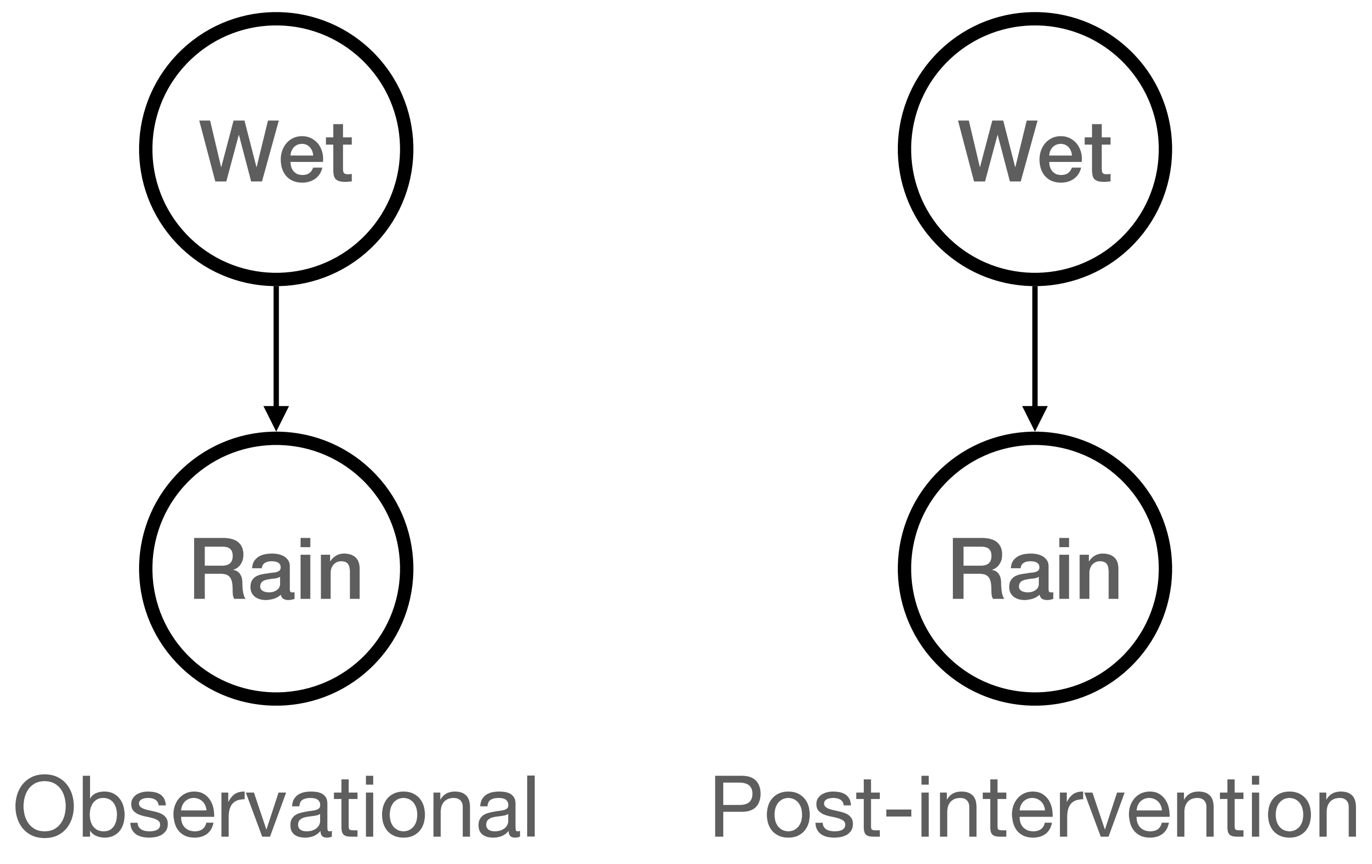
Natural network:

- Observational distribution: $P(Wet, Rain) = P(Wet \mid Rain)P(Rain)$
- Post intervention distribution: $\hat{P}(Wet = true, Rain) = P(Rain)P(Wet)$
- $P(Rain \mid do(Wet = true)) = .50$



Inverted network:

- Observational distribution: $P(Wet, Rain) = P(Rain \mid Wet)P(Wet)$
- Post intervention distribution:
  $\hat{P}(Wet = true, Rain) = P(Rain \mid Wet)P(Wet)$
- $P(Rain \mid do(Wet = true)) = .78$



Causal Models

- The **natural network** gives the correct answer to our causal query, but the **inverted network** does not (**Why**?)
- Not every factoring of a joint distribution is a **valid** causal model

Definition:
A **causal model** is a directed acyclic graph of random variables such that for every edge $X \to Y$, the value of random variable X is **realized before** the value of random variable Y.

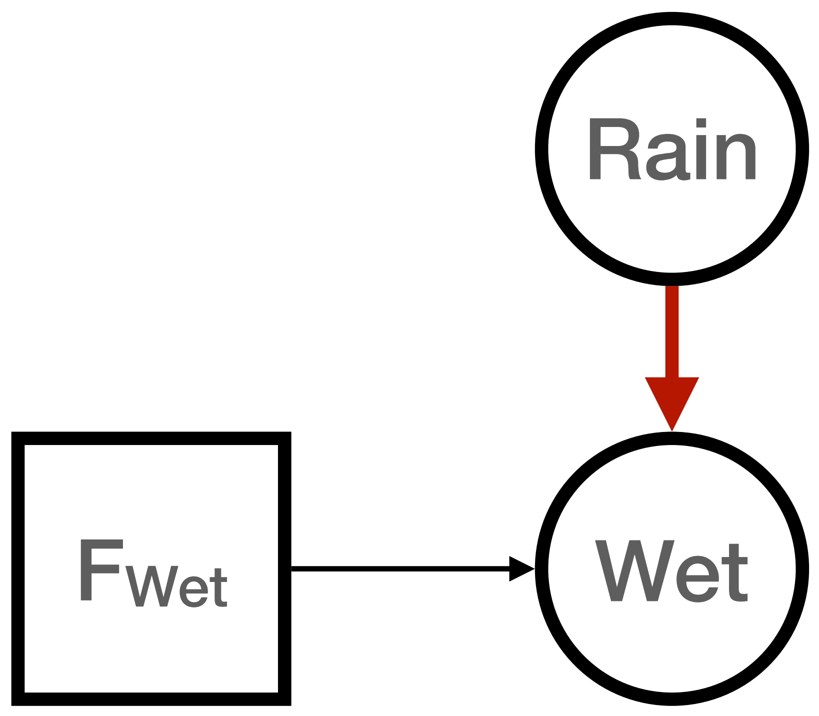
Alternative Representation: Influence Diagrams

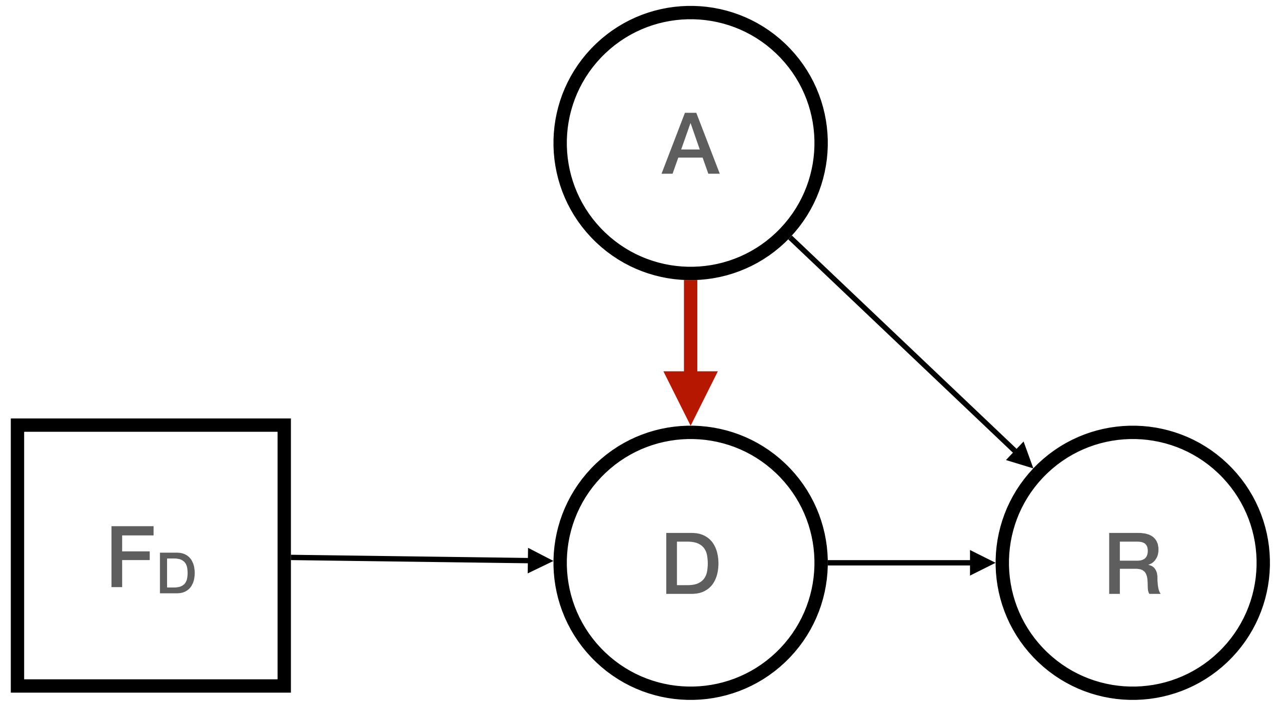
Instead of adding a new operator, we can instead represent causal queries by **augmenting** the causal model with **decision variables** F_D for each potential intervention target D.

$$dom(F_D) = dom(D) \cup \{idle\}$$

$$P(D \mid parents(D), F_D) = \begin{cases} P(D \mid parents(D)) & \text{if } F_D = idle, \\ 1 & \text{if } F_D \neq idle \wedge D = F_D, \\ 0 & \text{otherwise.} \end{cases}$$

Influence Diagrams Examples





Summary

- **Observational** queries $P(Y \mid X = x)$ are different from **causal** queries $P(Y \mid do(X = x))$
- To evaluate **causal query** $P(Y \mid do(X = x))$:
    1. Construct **post-intervention distribution** $\hat{P}$ by **removing** all links from X's direct parents to X
    2. Evaluate the **observational** query $\hat{P}(Y \mid X = x, Z = z)$ in the **post-intervention distribution**
- Not every correct Bayesian network is a valid causal model