Independence in Belief Networks
&
Causality Introduction

CMPUT 366: Intelligent Systems

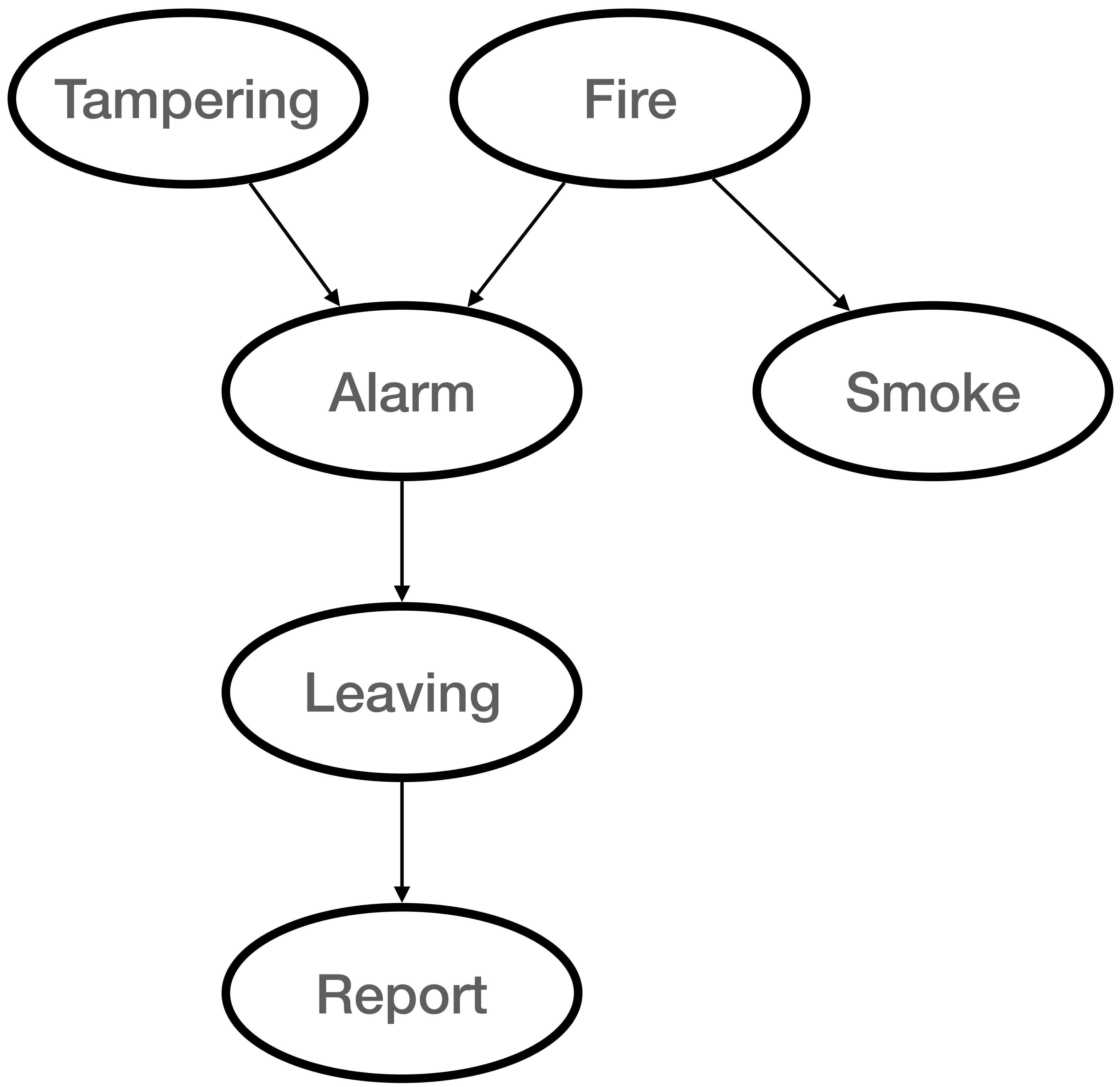
P&M §8.4, Bar §3.4

Lecture Outline

1. Recap
2. Reasoning About Independence
3. Causality Introduction

Recap: Belief Network Semantics

- Graph representation represents a specific **factorization** of the full **joint distribution**
    - Distribution on each node **conditional on its parents**
    - **Marginal distributions** on nodes with no parents
    - **Product** of these distributions is the joint distribution
    - Not every possible factorization is a **correct** factorization
- **Semantics:**
  Every node is **independent** of its **non-descendants**, **conditional** on its **parents**



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**:

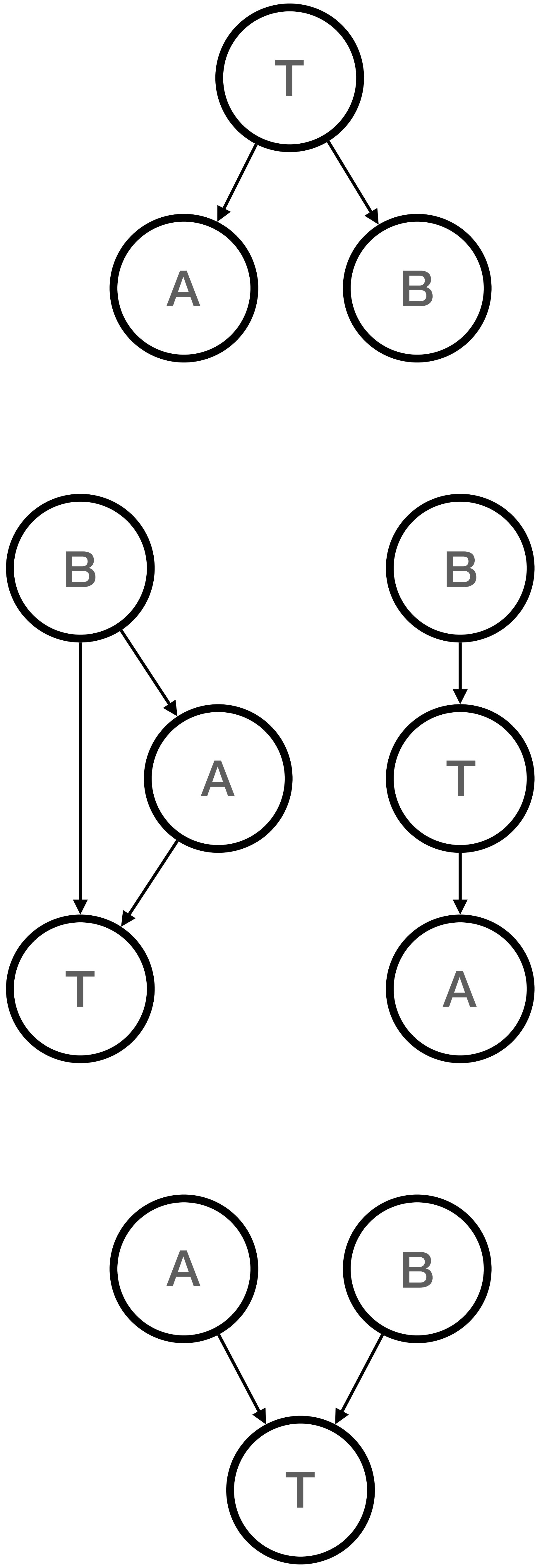
$$\sum_A \sum_E f_1(Q,A,B,C) \times f_2(C,D,E)$$

about 72 computations

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

about 28 computations

Reasoning About Independence



- A joint distribution can be factored in **multiple** different ways
- A belief network represents a **single** such factoring
- Some factorings are correct, some are incorrect

Independence in a Joint Distribution

Question: How can we answer questions about independence using the **joint distribution**?

Examples using P(*A*,*B*,*T*):

1. Is *A* independent of *B*?
   - P(*a* | *b*) = P(*a*) **for all** *a* ∈ dom(*A*), b ∈ dom(*B*)?
2. Is *T* independent of *A*?
   - P(*t* | *a*) = P(*t*) **for all** *a* ∈ dom(*A*), *t* ∈ dom(*T*)?
3. Is *A* independent of *B* given *T*?
   - P(*a* | *b*, *t*) = P(*a* | *t*) **for all** *a* ∈ dom(*A*), *b* ∈ dom(*B*), *t* ∈ dom(*T*)?

$$P(A,B) = \sum_{t \in T} P(A,B,T=t)$$

$$P(A,T) = \sum_{b \in B} P(A,B=b,T)$$

$$P(B,T) = \sum_{a \in A} P(A=a,B,T)$$

$$P(A) = \sum_{b \in B} P(A,B=b)$$

$$P(B) = \sum_{a \in A} P(A=a,B)$$

$$P(T) = \sum_{a \in A} P(A=a,T)$$

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

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

$$P(A\,|\,T) = \frac{P(A,T)}{P(T)}$$

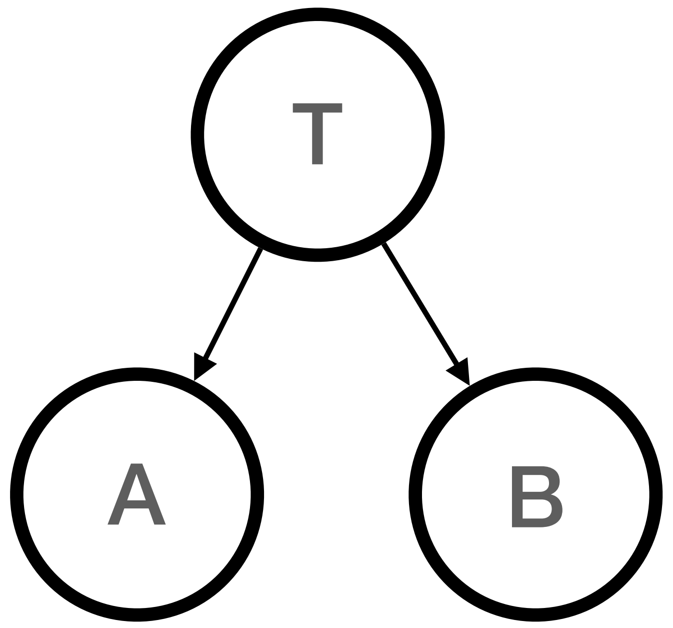
$$P(T\,|\,A) = \frac{P(A,T)}{P(A)}$$

Independence in a Belief Network

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

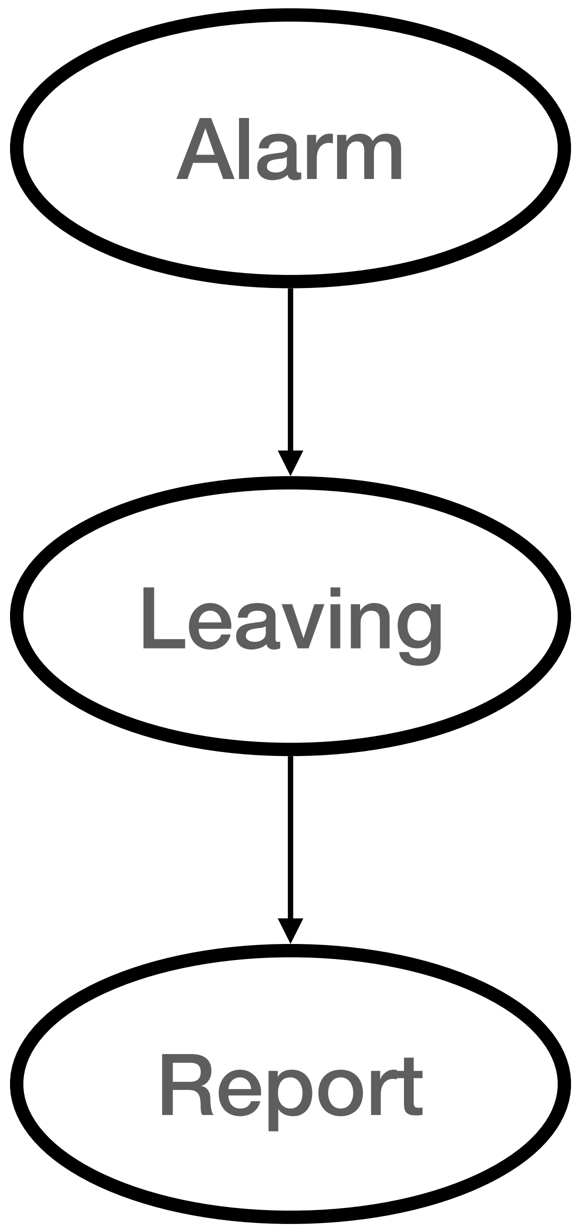
- We can use the semantics of a correct belief network to answer questions about independence
- Examples using the belief network at right:

1. Is **T** independent of **A**?
2. Is **A** independent of **B** given **T**?
3. Is **A** independent of **B**?



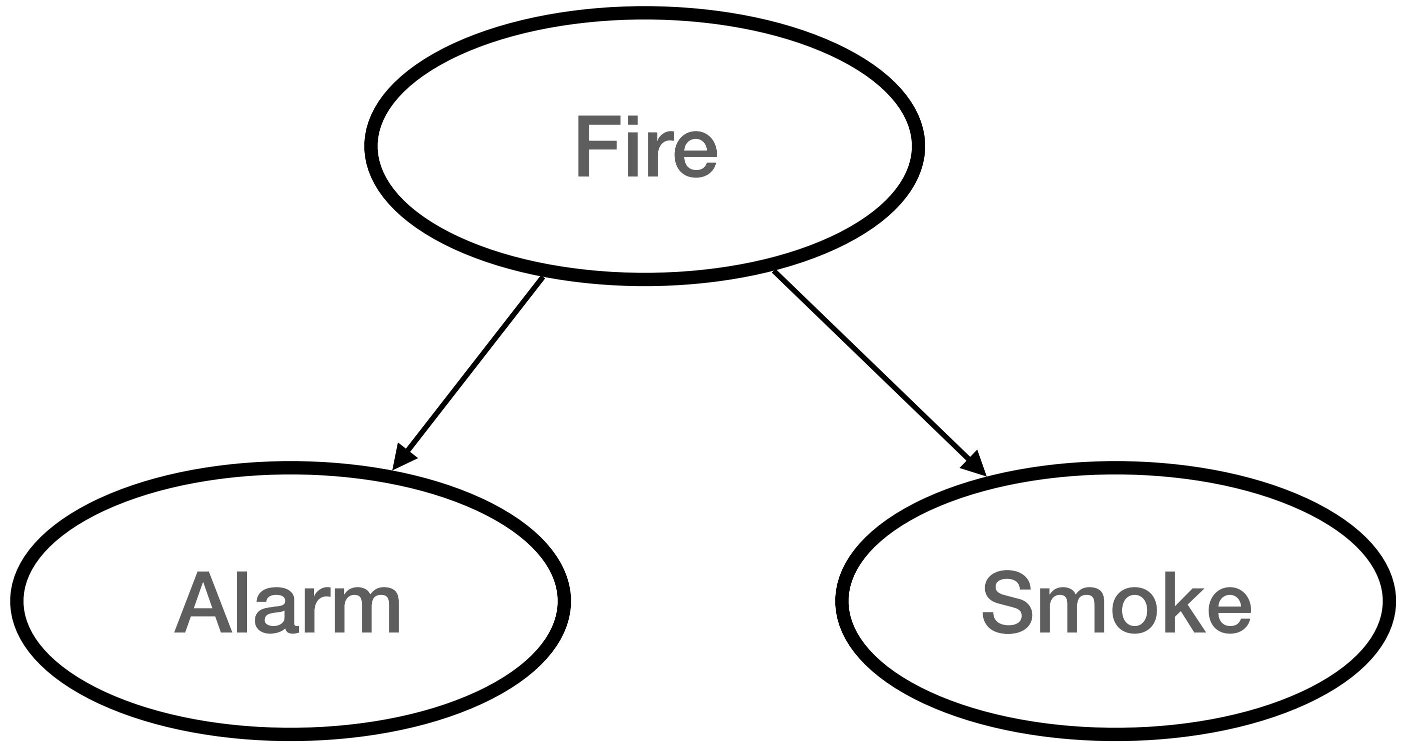
Chain

- **Question:** Is **Report** independent of **Alarm** given **Leaving**?
  - *Intuitively:* The only way learning **Report** tells us about **Alarm** is because it tells us about **Leaving**; but **Leaving** has already been observed
  - *Formally:* **Report** is independent of its non-descendants given only its parents
    - **Leaving** is **Report's** parent
    - **Alarm** is a non-descendant of **Report**
- **Question:** Is **Report** independent of **Alarm**?
  - *Intuitively:* Learning **Report** gives us information about **Leaving**, which gives us information about **Alarm**
  - *Formally:* **Report** is independent of **Alarm** given **Report's** parents; but the question is about **marginal** independence



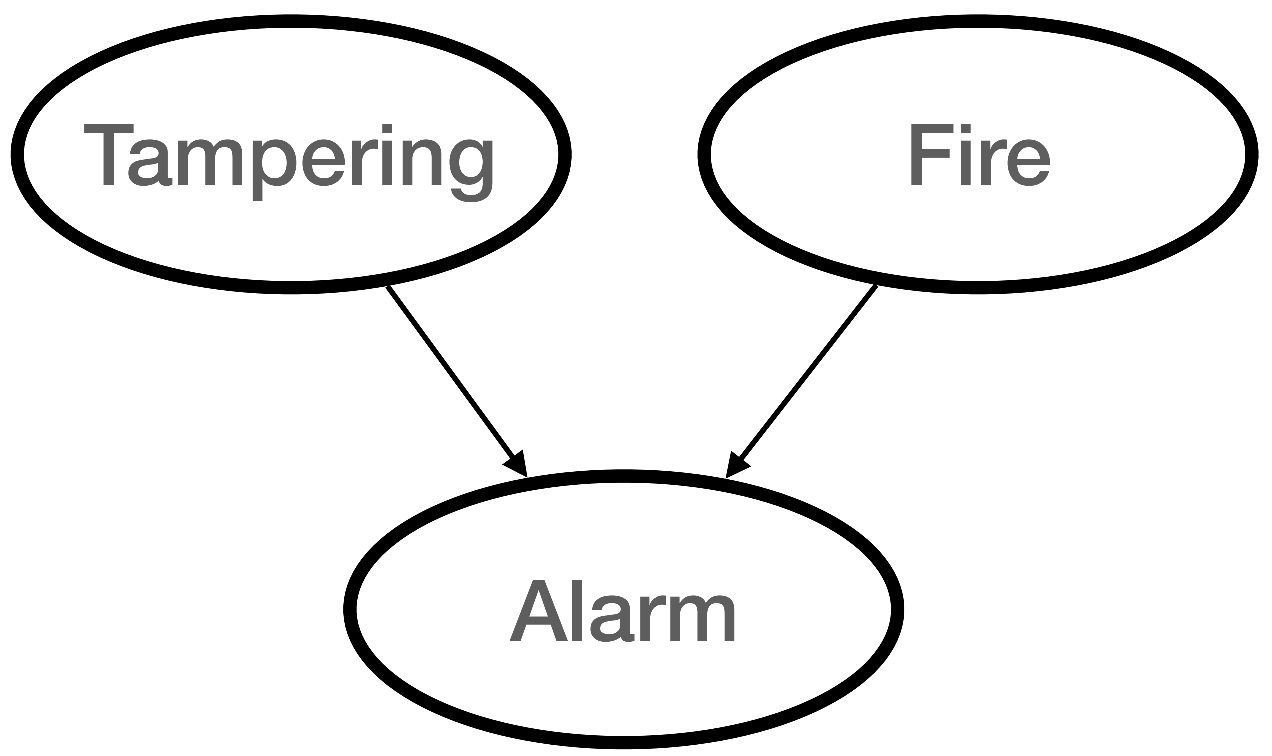
Common Ancestor

- **Question:** Is Alarm independent of Smoke given Fire?
  - *Intuitively:* The only way learning **Smoke** tells us about **Alarm** is because it tells us about **Fire**; but **Fire** has already been observed
  - *Formally:* **Alarm** is independent of its non-descendants given only its parents
    - **Fire** is **Alarm's** parent
    - **Smoke** is a non-descendant of **Fire**
- **Question:** Is Alarm independent of Smoke?
  - *Intuitively:* Learning **Smoke** gives us information about **Fire**, which gives us information about **Alarm**
  - *Formally:* **Alarm** is independent of **Smoke** given only **Alarm's** parents; but the question is about **marginal independence**



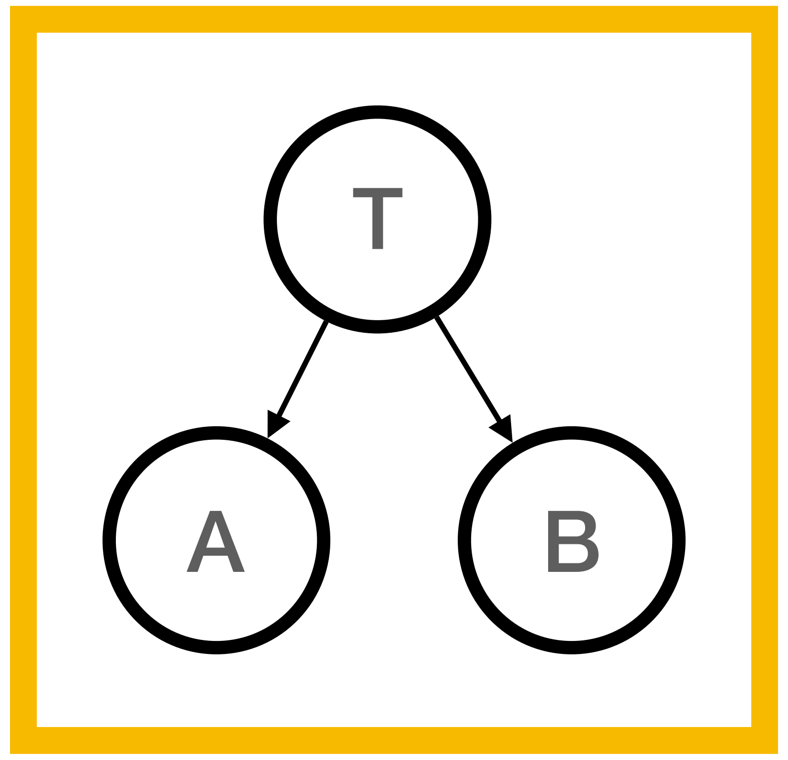
Common Descendant

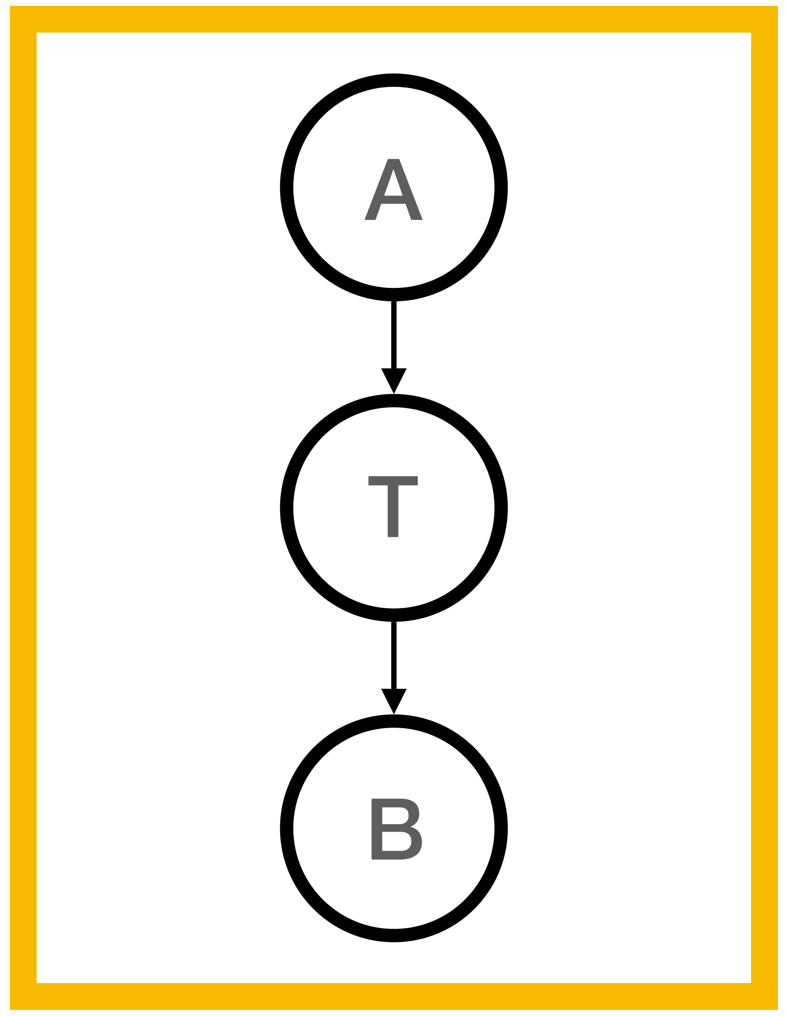
- **Question:** Is Tampering independent of Fire given Alarm?
  - *Intuitively:* If we know Alarm is ringing, then both Tampering and Fire are more likely. If we then learn that Tampering is false, that makes it less likely that the Alarm is ringing because of a Fire.
  - *Formally:* Tampering is independent of Fire given **only** Tampering's parents; but we are conditioning on one of Tampering's **descendants**
    - Conditioning on a **common descendant** can make independent variables dependent through the **explaining away** effect
- **Question:** Is Tampering independent of Fire?
  - *Intuitively:* Learning Tampering doesn't tell us anything about whether a Fire is happening
  - *Formally:* Tampering is independent of Fire given Tampering's parents
    - Tampering has no parents, so we are always conditioning on them
    - Fire is a non-descendant of Tampering

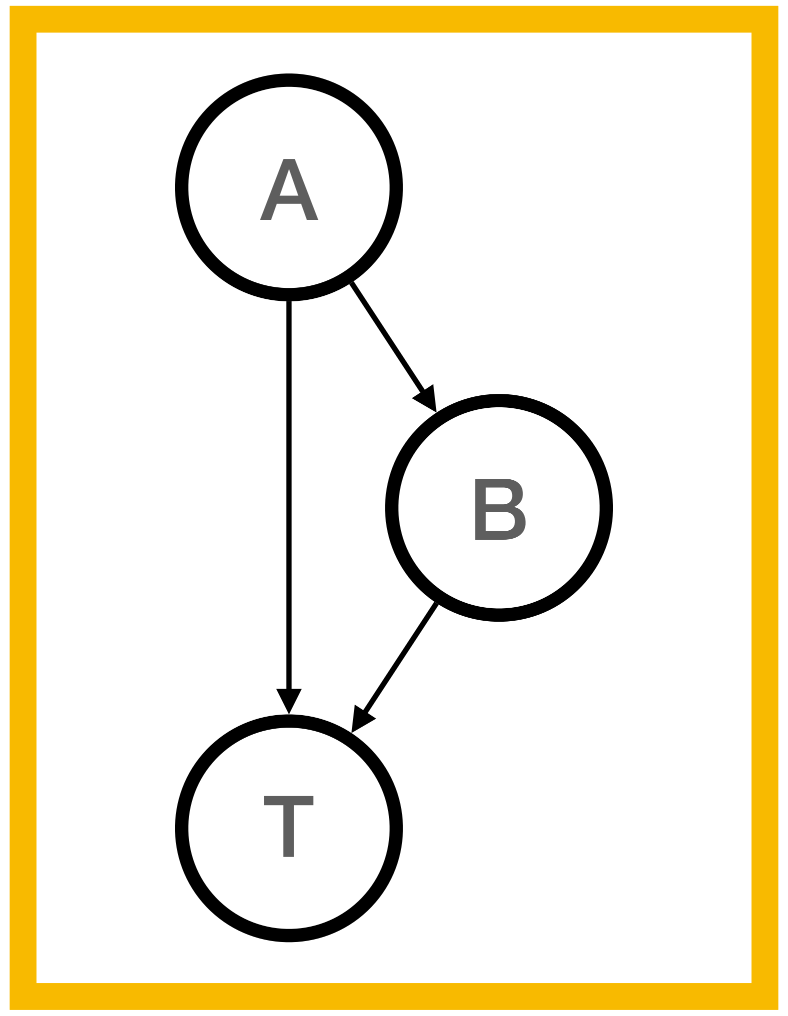


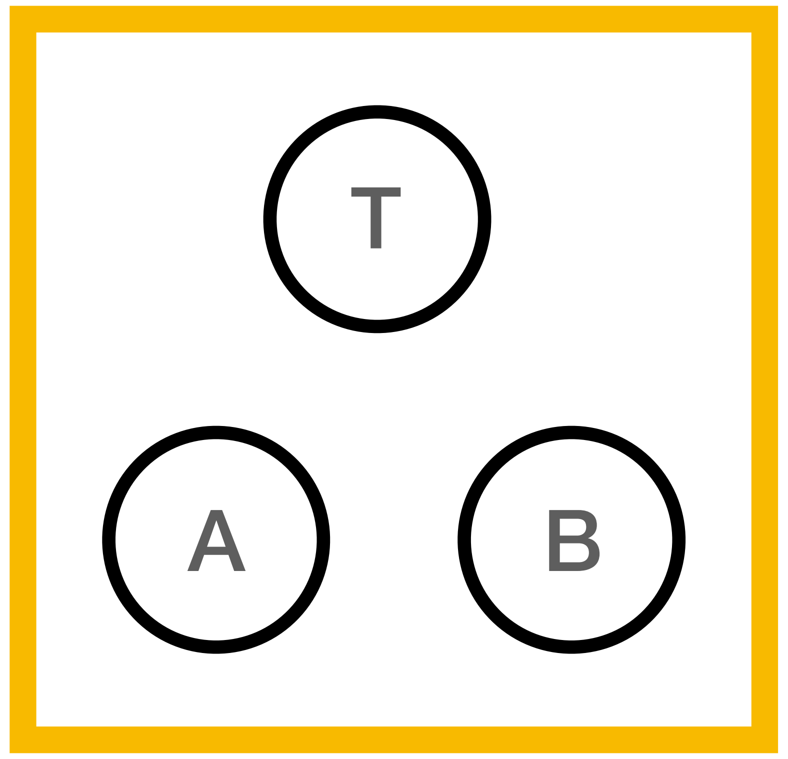
Correctness of a Belief Network

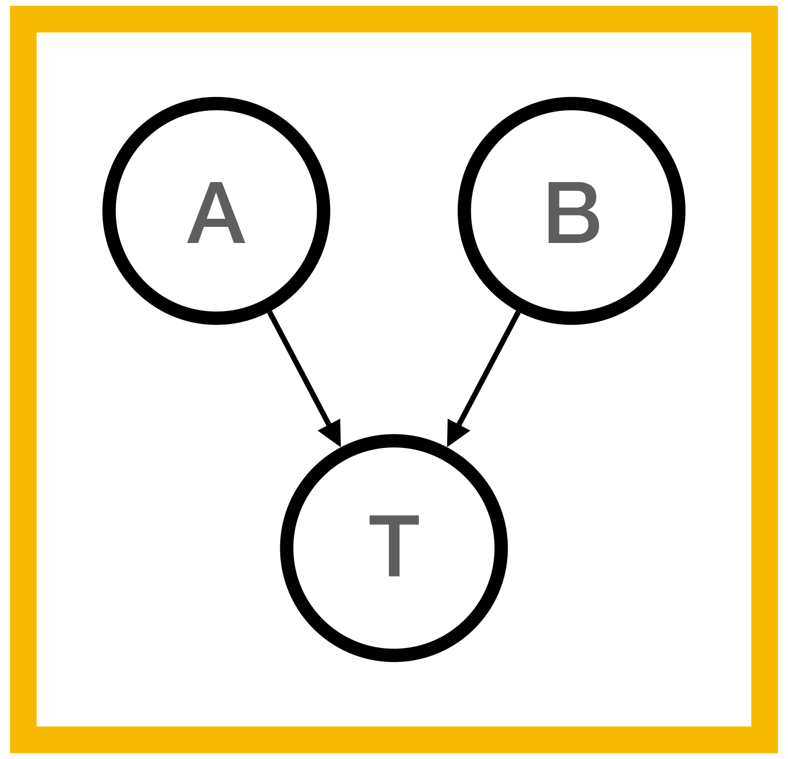
A belief network is a **correct** representation of a joint distribution when the belief network answers "yes" to an independence question **only if** the **joint distribution** answers "yes" to the same question.











Questions:

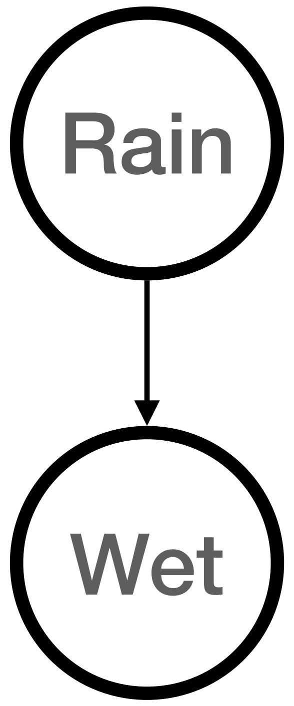
1. Is A independent of B in the above belief networks?
2. Is A independent of B given T in the above belief networks?

A: yes in graphs 4 and 5 (rightmost 2)

A: yes in graphs 1, 2, 4

Causality Introduction:
A Tale of Two Belief Networks

Raining	Wet	P(Raining, Sidewalk)
F	T	0.125
F	F	0.375
T	T	0.45
T	F	0.05

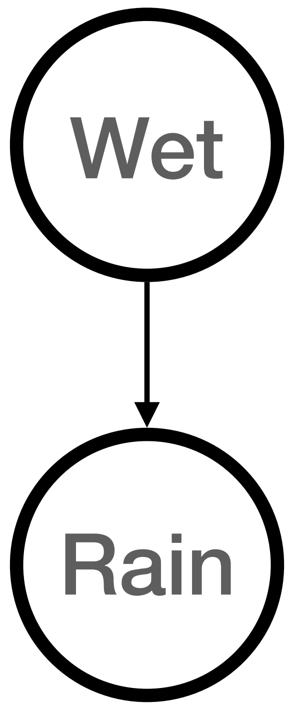


Natural network

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

$$\begin{aligned} P(\text{Rain}, \text{Wet}) &= P(\text{Wet} \mid \text{Rain})P(\text{Rain}) \\ &= P(\text{Rain} \mid \text{Wet})P(\text{Wet}) \end{aligned}$$

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

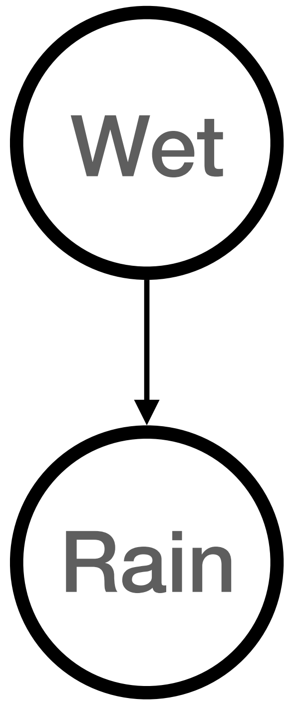


Inverted network

The Inverted Network Isn't Crazy

Corresponds to the factoring P(Rain | 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 causal network into this form to avoid having to do a lot of computations with **Bayes' Rule**



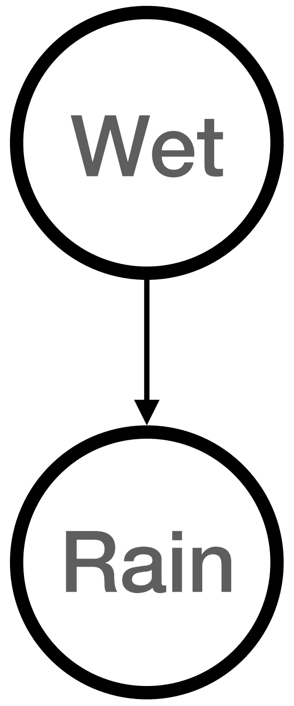
Inverted network

The Inverted Network Is Crazy

Corresponds to the factoring P(Rain | Wet)P(Wet)

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

A: this is a causal query, but we're doing it in the observational graph



Inverted network

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

Simpson's Paradox

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

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

- Is the drug effective for males?
  $P(R \mid D=\text{true}, G=\text{male}) = 0.60$
  $P(R \mid D=\text{false}, G=\text{male}) = 0.70$

- Is the drug effective for females?
  $P(R \mid D=\text{true}, G=\text{female}) = 0.20$
  $P(R \mid D=\text{false}, G=\text{female}) = 0.30$

- Is the drug effective?
  $P(R \mid D=\text{true}) = 0.50$
  $P(R \mid D=\text{false}) = 0.40$

Simpson's Paradox, explained

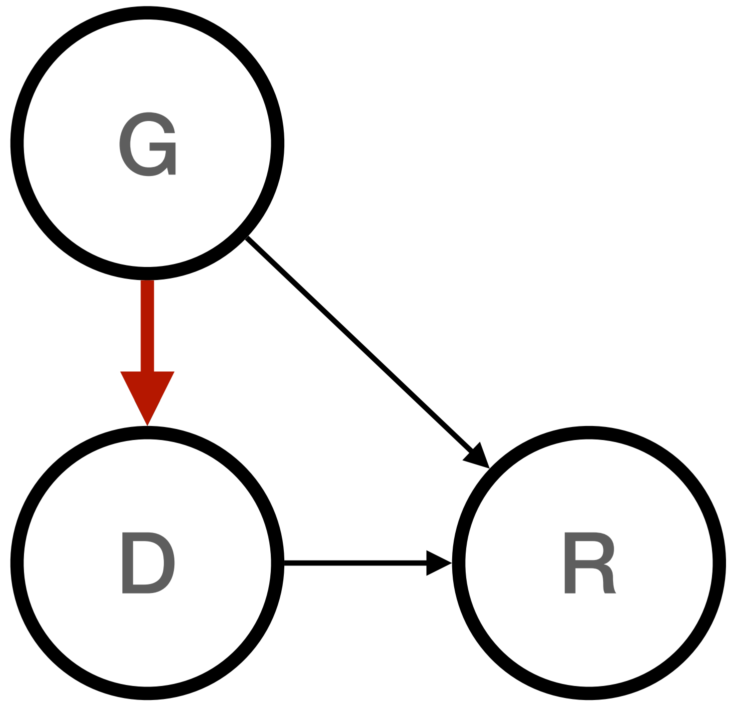
- The joint distribution factors as
  P(G,D,R) = P(R | D, G) x P(D | G) x P(G)

- **Per-gender** queries are answered **directly** by P(R | D, G)

- For the **overall query**, we want $P(R|D) = \dfrac{\sum_G P(R|G,D)P(G)}{\sum_{G,R} P(R|G,D)P(G)}$

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

$$P(R|D) = \frac{P(R,D)}{P(D)} = \frac{\sum_G P(G,D,R)}{\sum_{G,R} P(G,D,R)} = \frac{\sum_G P(R|D,G)\boxed{P(D|G)}P(G)}{\sum_{G,R} P(R|D,G)\boxed{P(D|G)}P(G)}$$

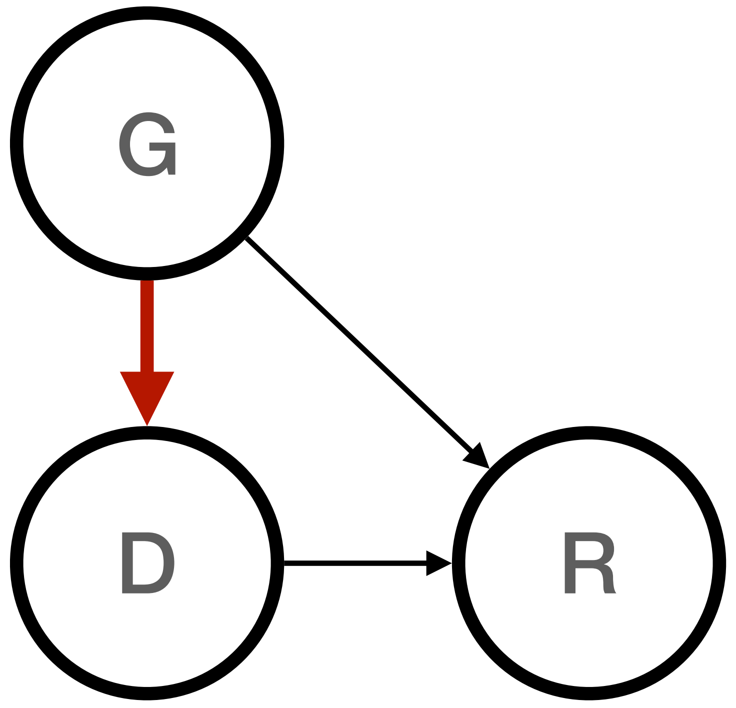


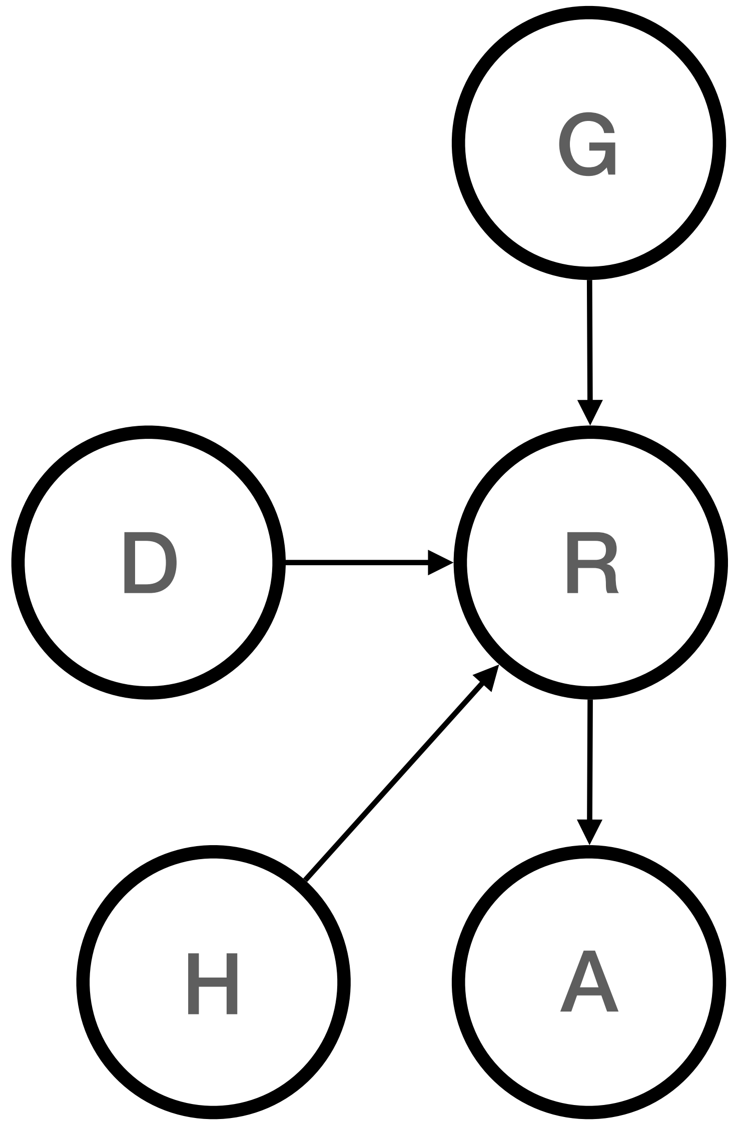
  - In our dataset, knowing whether a subject **got the drug** tells you something about their **gender**, and males have a **higher overall recovery** rate than females

  - P(R | G=male) = 0.625   vs P(R | G=female) = 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)





Causal Inference Summary

In the next lecture, we will learn how to:

- Systematically **express** causal queries
- Mechanically **compute** their answers
- Evaluate when a joint distribution is **informative** about causal queries
    - I.e., which causal queries are **identifiable** in a given dataset