CAUSAL LEARNING IN HEALTHCARE

Irene BALELLI
irene.balelli@inria.fr

OVERVIEW OF THE COURSE AND PLANNING

- Lesson 1 (3h) - Introduction to causality and causal discovery – 12/03

- Lesson 2 (3h) – Causal inference (the *do* operator and the potential outcome framework) – 23/03

- Lesson 3 (1h30) – A healthcare use case – 26/03

➢ Exam (1h30) – Pen and paper – 26/03

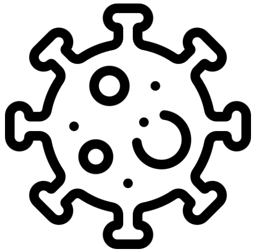
CAUSALITY AND THE CAUSAL HIERARCHY

- Why does causality matter?
- The ladder of causation

WHY DOES CAUSALITY MATTER?

Simpson's Paradox

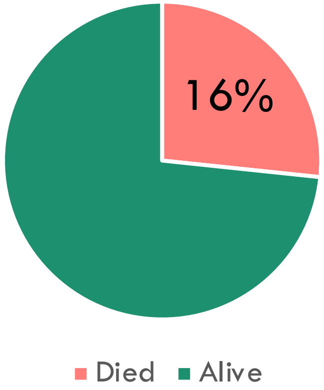
New virus



2050 subjects observed

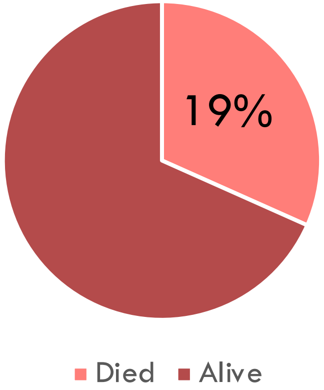
Treatment A

1500 subjects



Treatment B

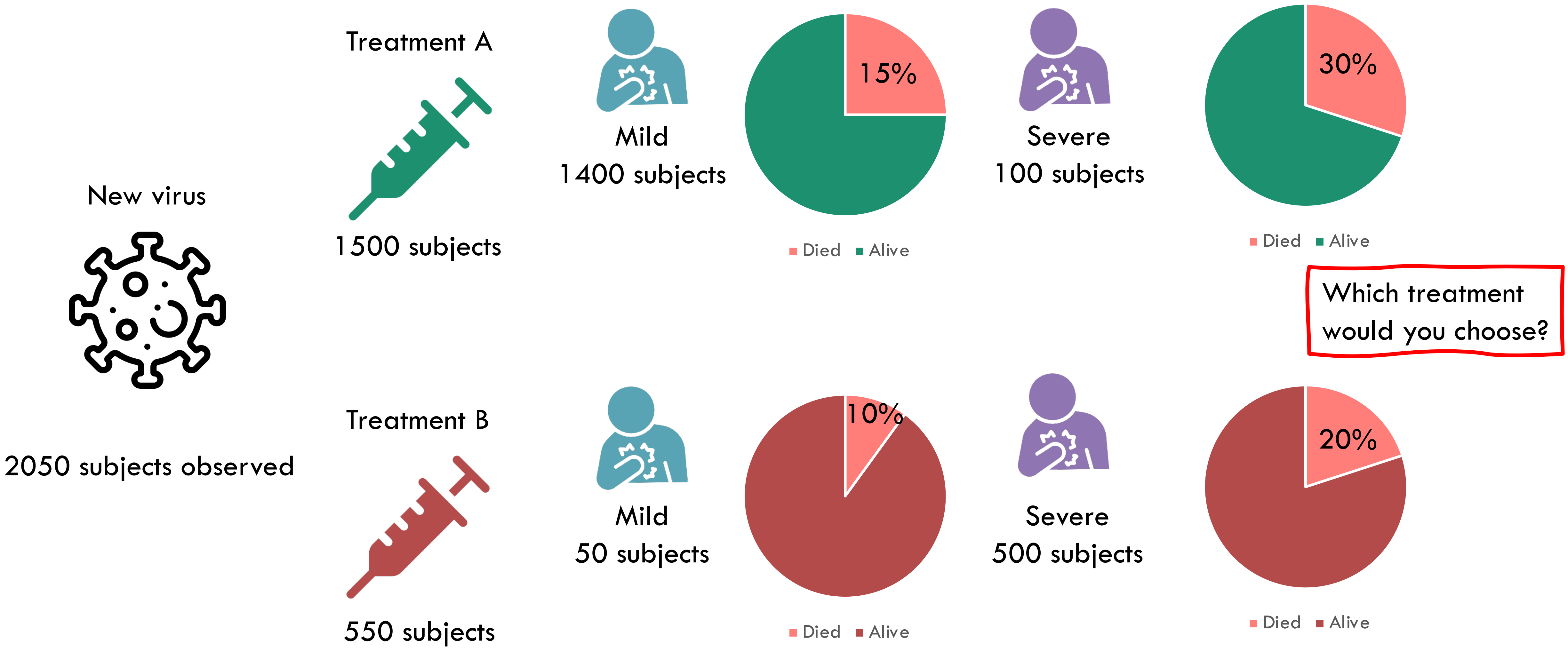
550 subjects



Which treatment would you choose?

WHY DOES CAUSALITY MATTER?

Simpson's Paradox



WHY DOES CAUSALITY MATTER?

Simpson's Paradox

	Output (Y)		
	Condition (C)		
Treatment (T)	Mild	Severe	Total
A (73%)	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B (27%)	**10%** (5/50)	**20%** (100/500)	19% (105/550)

$$\mathbb{E}[Y|T, C = \text{Mild}]$$

$$\mathbb{E}[Y|T]$$

$$\mathbb{E}[Y|T, C = \text{Severe}]$$

WHY DOES CAUSALITY MATTER?

Simpson's Paradox

	Output (Y)		
	Condition (C)		
Treatment (T)	**Mild**	**Severe**	**Total**
A (73%)	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B (27%)	**10%** (5/50)	**20%** (100/500)	19% (105/550)
	$\mathbb{E}[Y \mid T, C = \text{Mild}]$	$\mathbb{E}[Y \mid T, C = \text{Severe}]$	$\mathbb{E}[Y \mid T]$

$$\frac{1400}{1500} 0.15 + \frac{100}{1500} 0.30 = \mathbf{0.16}$$

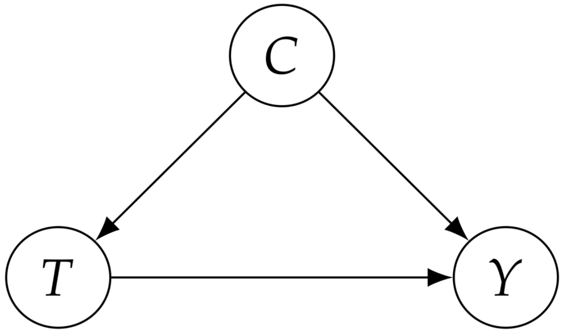
$$\frac{50}{550} \mathbf{0.10} + \frac{500}{550} \mathbf{0.20} = 0.19$$

WHY DOES CAUSALITY MATTER?

Simpson's Paradox

Scenario 1: Condition C is a cause of treatment assignment T (e.g., doctors decide to give treatment A, less expensive/easily accessible to patients with mild symptoms, and keep treatment B, more expensive/more limited to patients with severe symptoms).

Graphically:



Conclusion: Treatment B is more effective at reducing mortality Y!

Can you see why?

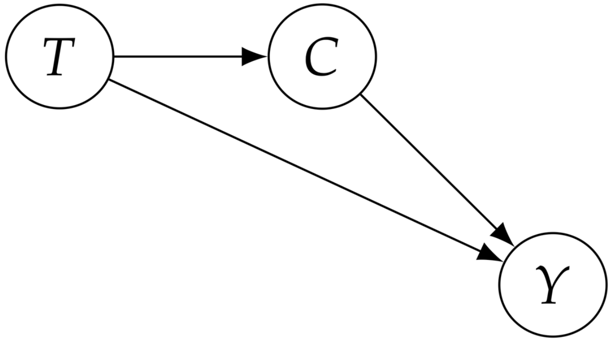
	Output (Y)		
	Condition (C)		
Treatment (T)	Mild	Severe	Total
A (73%)	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B (27%)	**10%** (5/50)	**20%** (100/500)	19% (105/550)

WHY DOES CAUSALITY MATTER?

Simpson's Paradox

Scenario 2: Treatment prescription T is a cause of condition C (e.g., treatment B is so scarce that patients with prescribed B have to wait longer before effective administration, and their condition worsens from mild to severe).

Graphically:



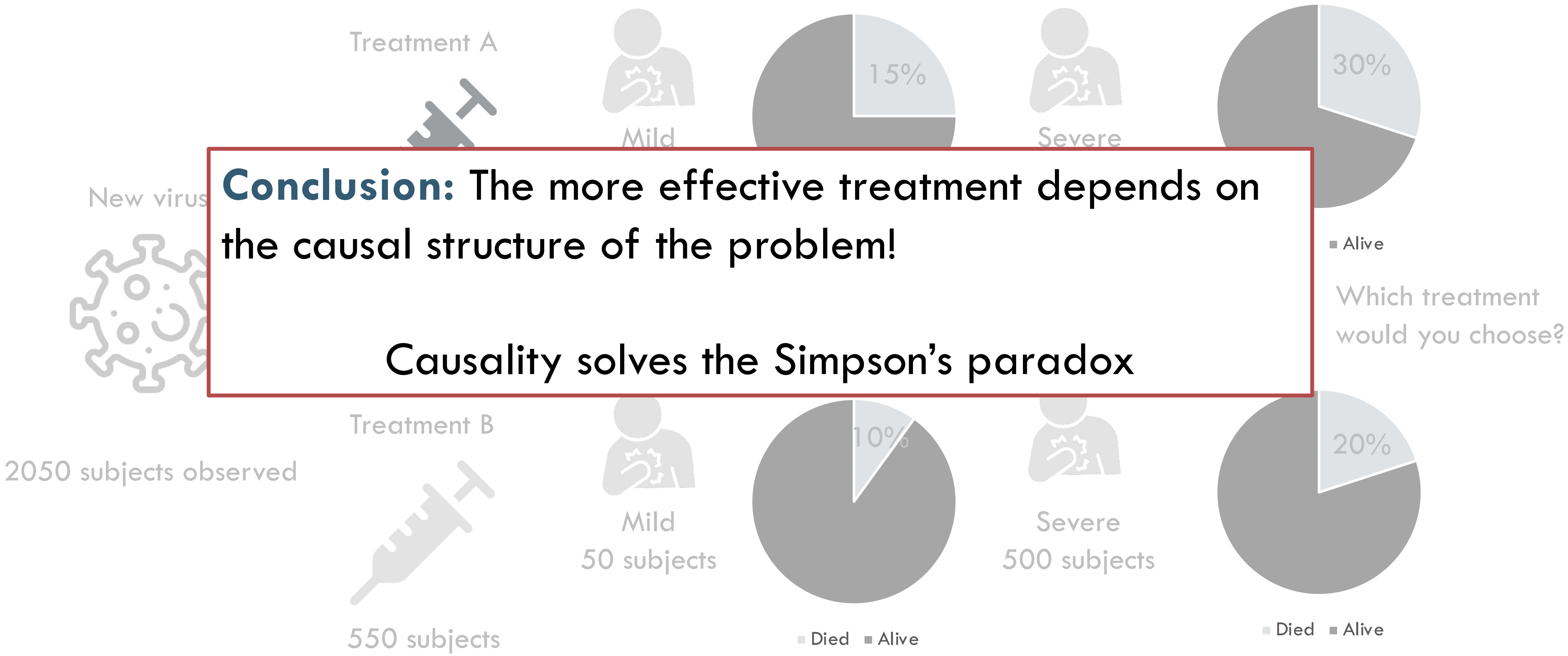
Conclusion: Treatment A is more effective!

→ We do not have to control for C to assess effectiveness

	Output (Y)		
	Condition (C)		
Treatment (T)	Mild	Severe	Total
A (73%)	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B (27%)	**10%** (5/50)	**20%** (100/500)	19% (105/550)

WHY DOES CAUSALITY MATTER?

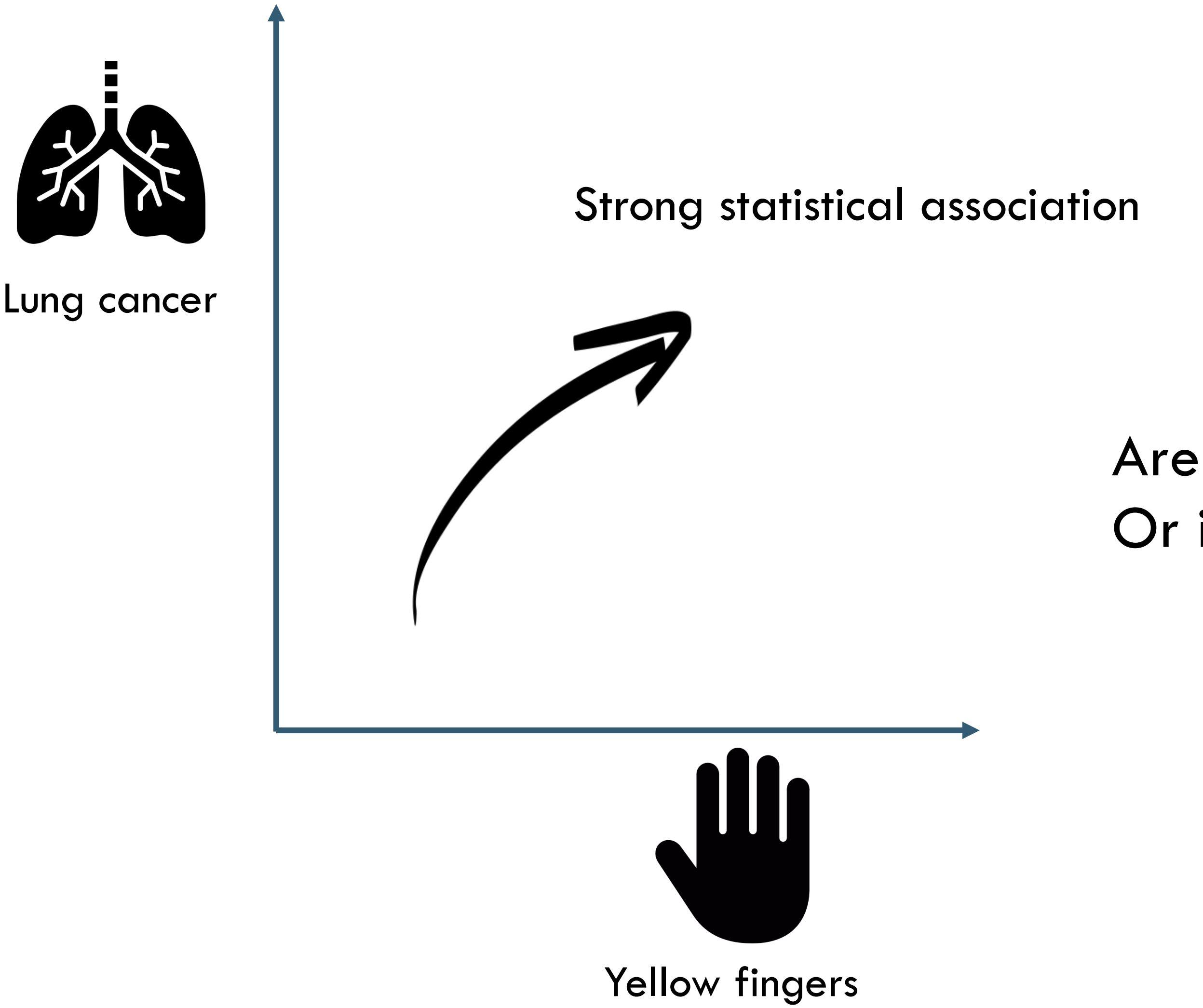
Simpson’s Paradox



Conclusion: The more effective treatment depends on the causal structure of the problem!

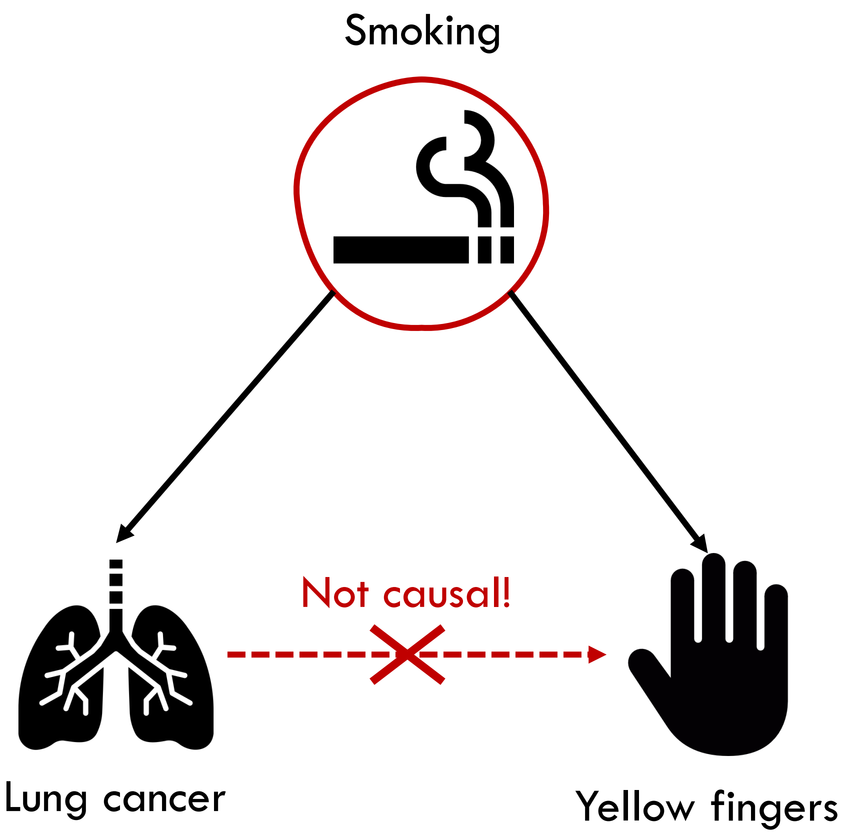
Causality solves the Simpson’s paradox

STATISTICAL ASSOCIATION IS NOT CAUSATION



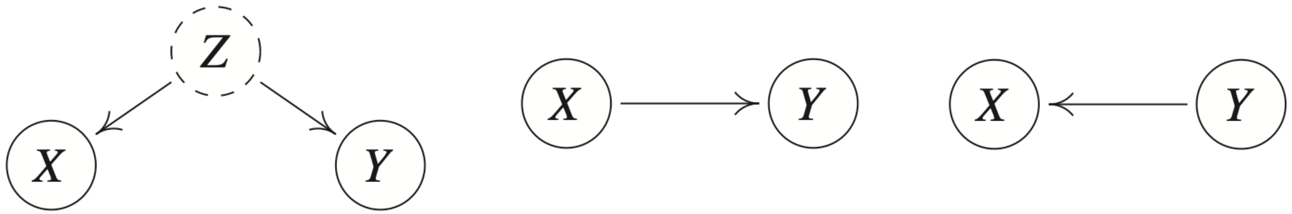
Are yellow fingers causing lung cancer?
Or is lung cancer responsible for yellow fingers?

STATISTICAL ASSOCIATION IS NOT CAUSATION

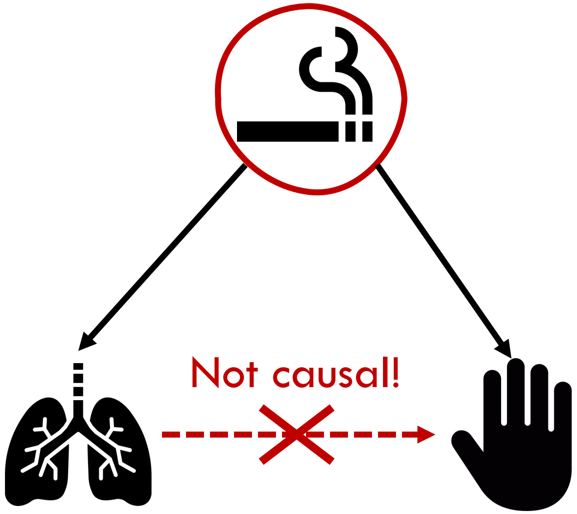


Reichenbach's common cause principle. If two random variables X and Y are statistically dependent ($X \not\perp\!\!\!\perp Y$), then there exists a third variable Z that causally influences both. (As a special case, Z may coincide with either X or Y)

Furthermore, this variable Z screens X and Y from each other in the sense that given Z, they become independent, $X \perp\!\!\!\perp Y \mid Z$.



STATISTICAL ASSOCIATION IS NOT CAUSATION



Reichenbach's common cause principle. If two random variables X and Y are statistically dependent ($X \not\perp\!\!\!\perp Y$), then there exists a third variable Z that causally influences both. (As a special case, Z may coincide with either X or Y)

Furthermore, this variable Z screens X and Y from each other in the sense that given Z, they become independent, $X \perp\!\!\!\perp Y | Z$.

X and Y are **statistically associated** iff:

$$\exists x_i \neq x_2, \mathbb{P}(Y|X = x_1) \neq \mathbb{P}(Y|X = x_2)$$

X is a **cause** of Y iff:

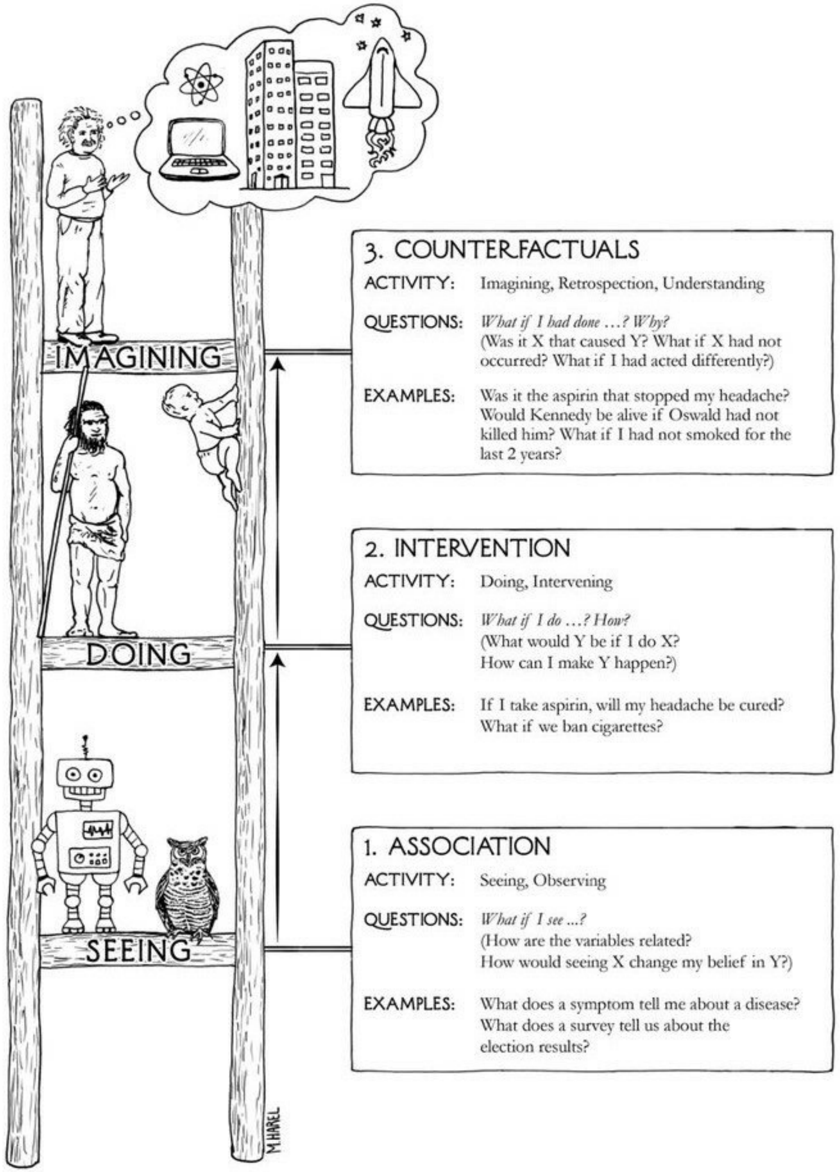
$$\exists x_i \neq x_2, \mathbb{P}(Y|\text{set } X = x_1) \neq \mathbb{P}(Y|\text{set } X = x_2)$$

THE LADDER OF CAUSATION

Causal simulation: causal inference at the **individual** level

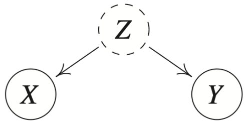
Causality-in-mean: causal inference at the **cohort** level

Statistical **association**
- basic statistics
- classical ML



I'm interested in: a DAG + SCM $\mathbb{P}(Y|\text{set } X = x)$

I'm interested in: a DAG. $\mathcal{G}$ (+ BN)



$X \perp\!\!\!\perp Y | Z$

I'm interested in: $\mathbb{P}(Y|X = x)$

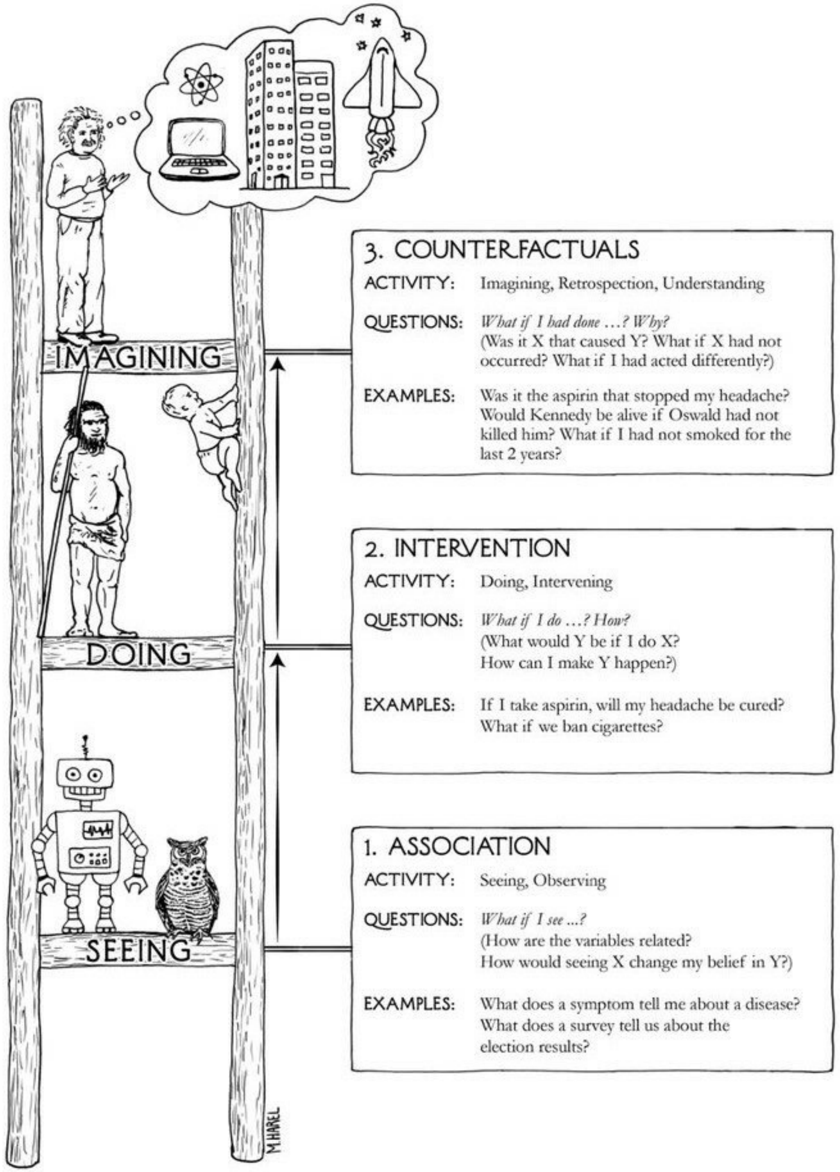
The ladder of causation - Pearl, J. (2000). *Causality: Models, Reasoning, and Inference.*

THE LADDER OF CAUSATION

Causal simulation: causal inference at the **individual** level

Causality-in-mean: causal inference at the **cohort** level

Statistical **association**
- basic statistics
- classical ML



Causal inference

Causal discovery

Not causal reasoning

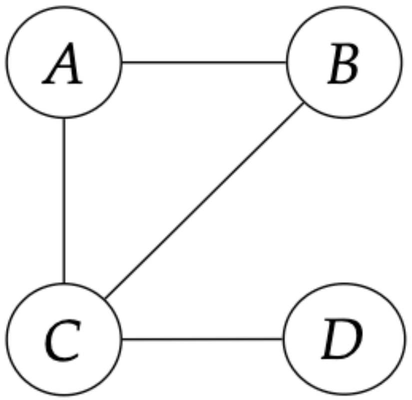
The ladder of causation - Pearl, J. (2000). *Causality: Models, Reasoning, and Inference.*

CAUSAL GRAPHICAL MODELS

- The Markov Property and Faithfulness
- Chain, Fork, and Collider
- d-separation and conditional independence

CAUSALITY THROUGH GRAPHS

- **Undirected Graph:** a set of nodes ($V = \{X_1, \dots, X_N\}$) and edges ($E = \{(X_i, X_j)\}$) establishing pairwise links between nodes in V.

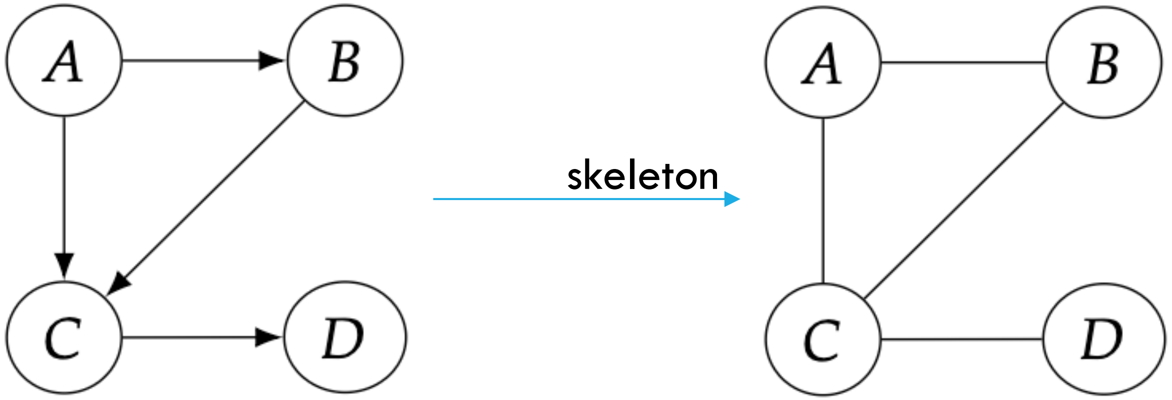


$$V = \{A, B, C, D\}$$
$$E = \{(A, B), (A, C), (B, C), (C, D)\} = \{(B, A), (C, A), (C, B), (D, C)\}$$

CAUSALITY THROUGH GRAPHS

- **Undirected Graph:** a set of nodes ($V = \{X_1, \dots, X_N\}$) and edges ($E = \{(X_i, X_j)\}$) establishing pairwise links between nodes in V.

- **Directed graph:** a graph whose set of edges E only contains ordered tuples, i.e. $(X_i, X_j) \neq (X_j, X_i)$. Its undirected counterpart is often referred to as the **skeleton**.

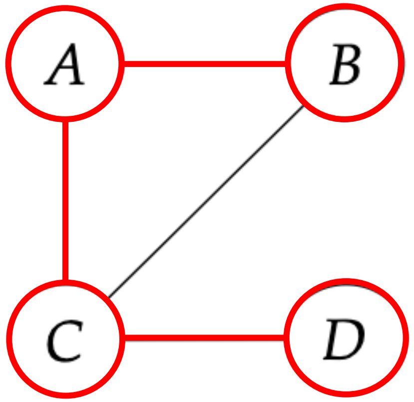


CAUSALITY THROUGH GRAPHS

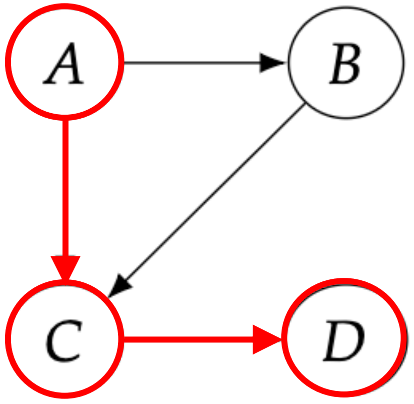
- **Undirected Graph:** a set of nodes ($V = \{X_1, \dots, X_N\}$) and edges ($E = \{(X_i, X_j)\}$) establishing pairwise links between nodes in V.

- **Directed graph:** a graph whose set of edges E only contains ordered tuples, i.e. $(X_i, X_j) \neq (X_j, X_i)$. Its undirected counterpart is often referred to as the **skeleton**.

- **(directed) path:** consecutive (directed) edges.



Undirected path: $(A, B), (A, C), (C, D)$



Directed path: $(A, C), (C, D)$

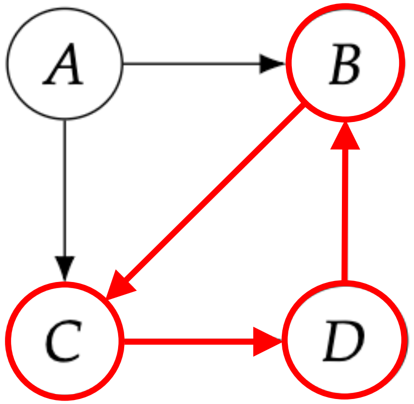
CAUSALITY THROUGH GRAPHS

- **Undirected Graph:** a set of nodes ($V = \{X_1, \dots, X_N\}$) and edges ($E = \{(X_i, X_j)\}$) establishing pairwise links between nodes in V.

- **Directed graph:** a graph whose set of edges E only contains ordered tuples, i.e. $(X_i, X_j) \neq (X_j, X_i)$. Its undirected counterpart is often referred to as the **skeleton**.

- **(directed) path:** consecutive (directed) edges.

- **Directed Acyclic graph (DAG):** a directed graph without cycles, i.e. with no directed path starting and ending in the same node $(X_i, X_{i+1})(X_{i+1}, X_{i+2}) \dots (X_{i+k}, X_i)$.



A cycle: $(B, C), (C, D), (D, B)$

CAUSALITY THROUGH GRAPHS

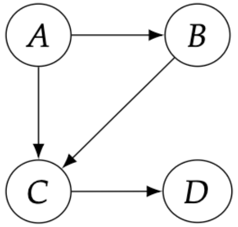
- **Undirected Graph:** a set of nodes ($V = \{X_1, \ldots, X_N\}$) and edges ($E = \{(X_i, X_j)\}$) establishing pairwise links between nodes in V.

- **Directed graph:** a graph whose set of edges E only contains ordered tuples, i.e. $(X_i, X_j) \neq (X_j, X_i)$. Its undirected counterpart is often referred to as the **skeleton**.

- **(directed) path:** consecutive (directed) edges.

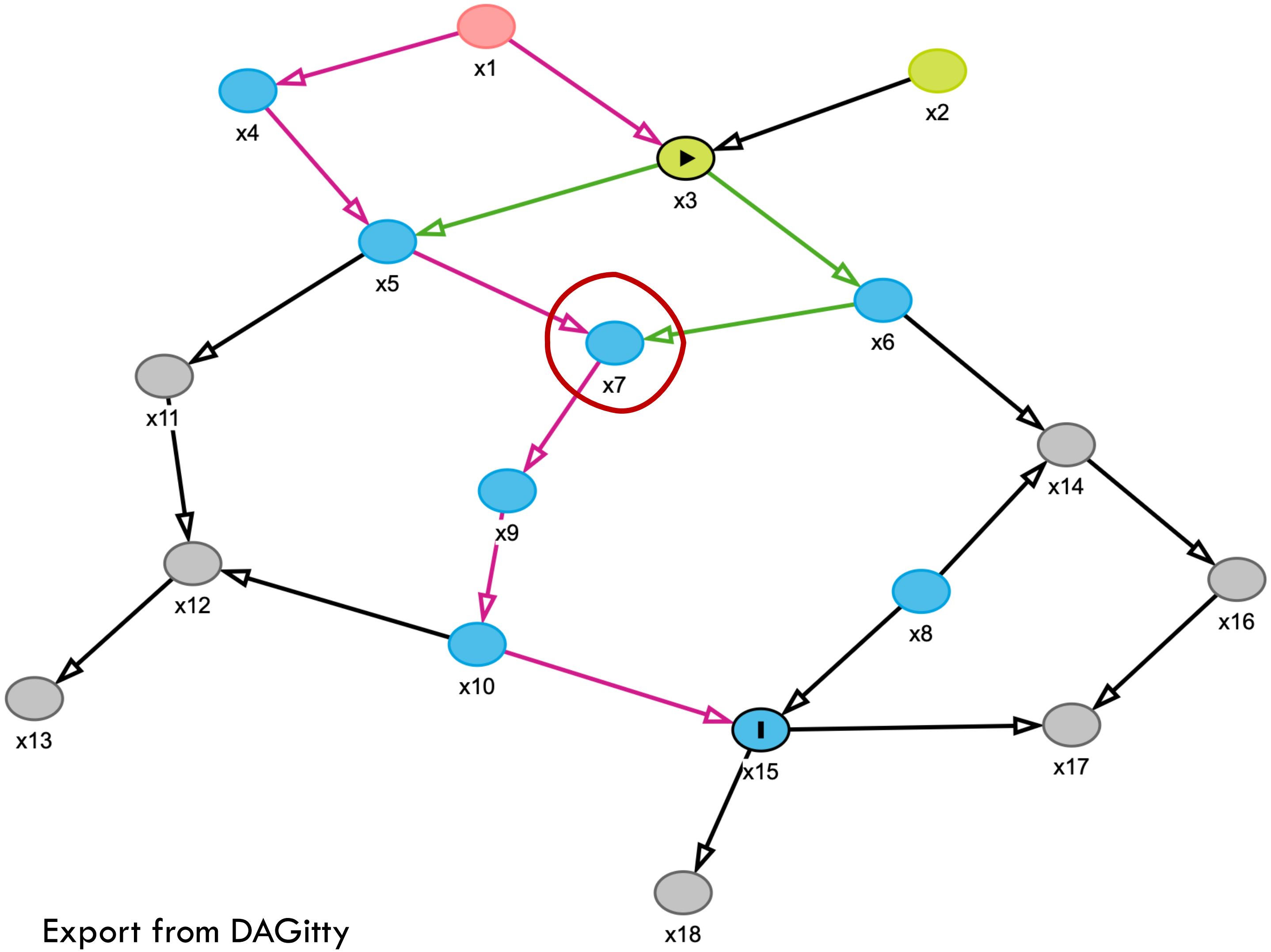
- **Directed Acyclic graph (DAG):** a directed graph without cycles, i.e. with no directed path starting and ending in the same node $(X_i, X_{i+1})(X_{i+1}, X_{i+2}) \ldots (X_{i+k}, X_i)$.

- **Causal DAG:** a DAG where each node represent a random variable, each directed edge a (potential) direct causal relationship (→ a probabilistic interaction) and the absence of an edge denotes no direct causal relationship.



$$\mathbb{P}(D) = \mathbb{P}(D|C)\mathbb{P}(C|A,B)\mathbb{P}(B|A)\mathbb{P}(A)$$

CAUSALITY THROUGH GRAPHS – SOME TERMINOLOGY

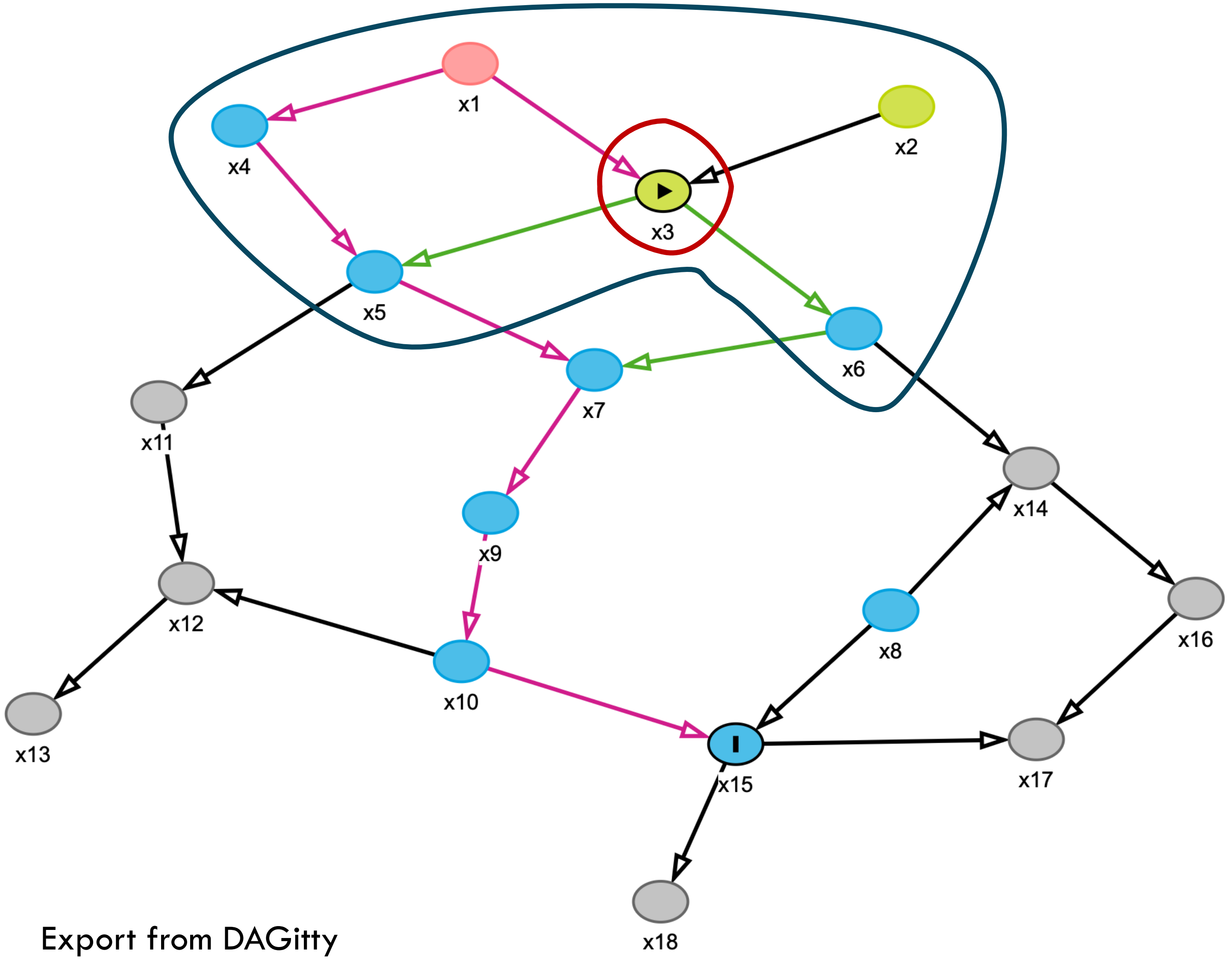


Export from DAGitty

Let's focus on node X_7:

- **Parents** – direct causes:
  $Pa_{X_7} = \{X_5, X_6\}$
- **Ancestors** – (un)direct causes:
  $Anc_{X_7} = Pa_{X_7} \cup \{X_1, X_2, X_3, X_4\}$
- **Children** – direct effects:
  $Ch_{X_7} = \{X_9\}$
- **Descendant** – (un)direct effects:
  $Des_{X_7} = Ch_{X_7} \cup \{X_{10}, X_{12}, X_{13}, X_{15}, X_{17}, X_{18}\}$

CAUSALITY THROUGH GRAPHS – SOME TERMINOLOGY



Export from DAGitty

Markov blanket

The Markov blanket of a node X, MB_X, is the set consisting of:

- Parents of X
- Children of X
- Parents of children of X (spouses)

Could you see the MB_{X_3}?

CAUSALITY THROUGH GRAPHS – SOME TERMINOLOGY

Local Markov Property:

In a DAG, a node X is independent from all its non-descendants given its parents.

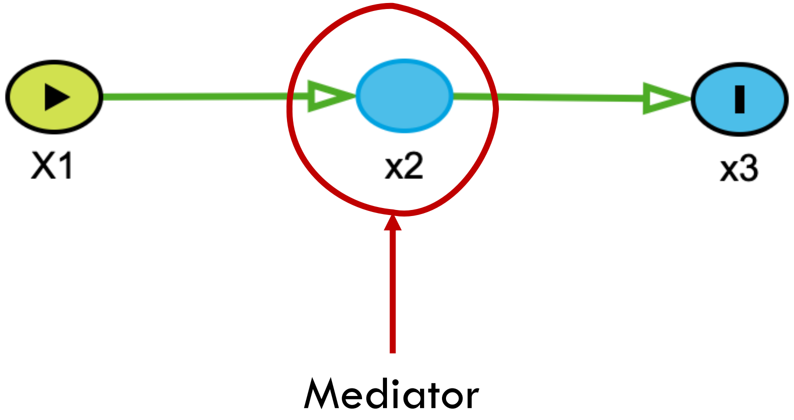
Local Markov Condition:

The joint distribution of the variables in a DAG can be factorized as follows:

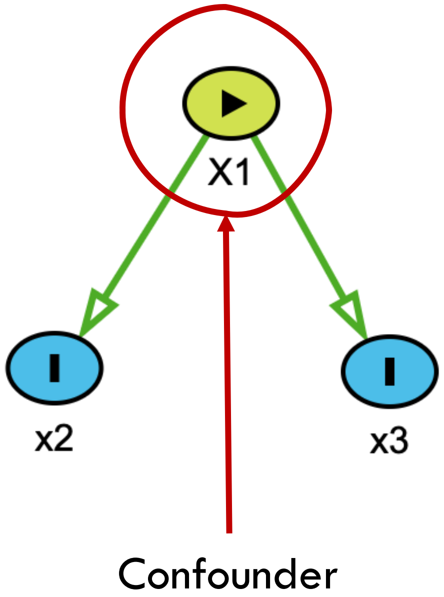
$$\mathbb{P}(X_1, \ldots, X_N) = \prod_{i=1}^{N} \mathbb{P}(X_i | \mathrm{Pa}_{X_i})$$

CAUSALITY THROUGH GRAPHS – SOME TERMINOLOGY

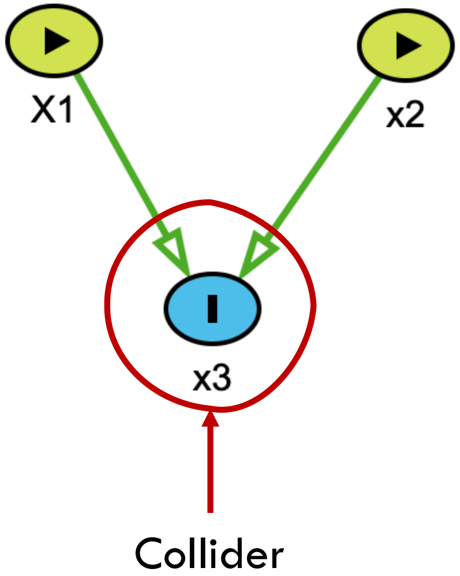
A chain

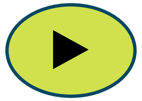


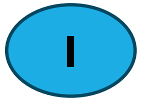
A fork



A v-structure



 Root (no parents)

 Sink (no children)

CAUSAL STRUCTURE AND STATISTICAL INDEPENDENCE

Recall:

$$X \perp\!\!\!\perp Y \Leftrightarrow \mathbb{P}(X, Y) = \mathbb{P}(X)\mathbb{P}(Y)$$
$$X \perp\!\!\!\perp Y \Leftrightarrow \mathbb{P}(X|Y) = \mathbb{P}(X)$$

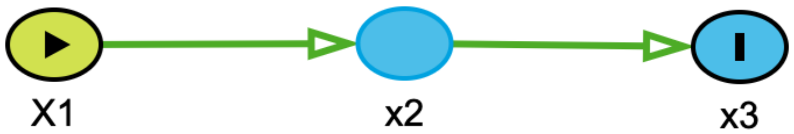
Independence

$$X \perp\!\!\!\perp Y|Z \Leftrightarrow \mathbb{P}(X, Y|Z) = \mathbb{P}(X|Z)\mathbb{P}(Y|Z)$$
$$X \perp\!\!\!\perp Y|Z \Leftrightarrow \mathbb{P}(X|Y, Z) = \mathbb{P}(X|Z)$$

Conditional independence

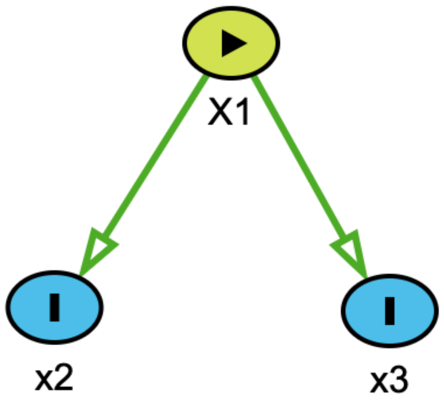
CAUSAL STRUCTURE AND STATISTICAL INDEPENDENCE

1. All **adjacent nodes** (nodes linked through a single edge) are **dependent.**



$$X_1 \not\perp\!\!\!\perp X_2$$

$$X_2 \not\perp\!\!\!\perp X_3$$



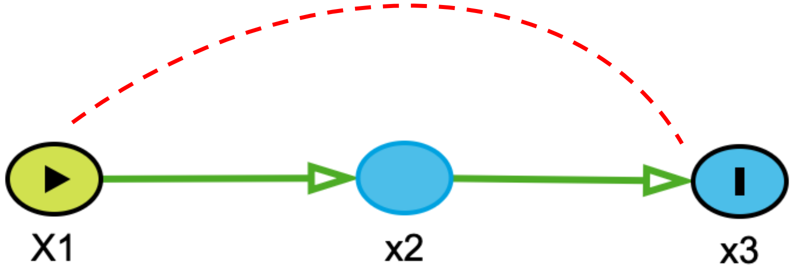
$$X_1 \not\perp\!\!\!\perp X_2$$

$$X_1 \not\perp\!\!\!\perp X_3$$

CAUSAL STRUCTURE AND STATISTICAL INDEPENDENCE

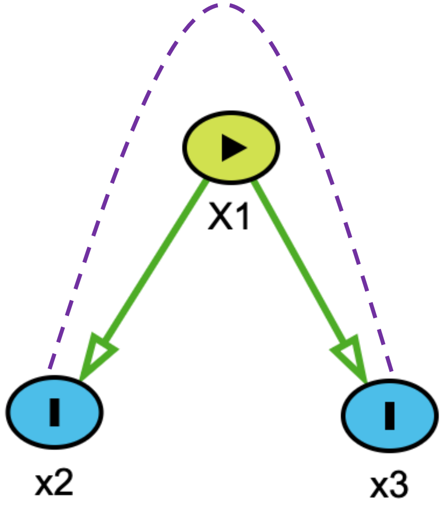
1. All **adjacent nodes** (nodes linked through a single edge) are **dependent.**
2. Statistical **association** can arise between **non-adjacent** nodes through **paths.**

(Causal) statistical association



$$X_1 \not\perp\!\!\!\perp X_3$$

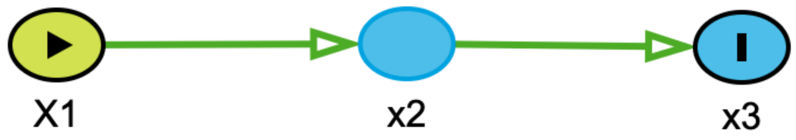
(Non-causal) statistical association



$$X_2 \not\perp\!\!\!\perp X_3$$

CAUSAL STRUCTURE AND STATISTICAL INDEPENDENCE

1. All **adjacent nodes** (nodes linked through a single edge) are **dependent**.
2. Statistical **association** can arise between **non-adjacent** nodes through **paths.**
3. A **DAG + Markov condition** define a **set of (conditional) independence facts.**



Show that: $X_3 \perp\!\!\!\perp X_1 | X_2$

From the Markov factorisation: $\mathbb{P}(X_1, X_2, X_3) = \mathbb{P}(X_3|X_2)\mathbb{P}(X_2|X_1)\mathbb{P}(X_1)$

Conditional probability formula:

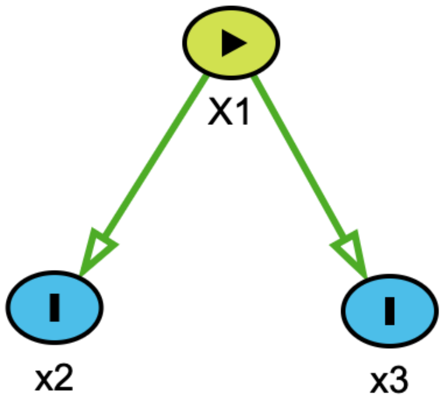
$$\mathbb{P}(X_1, X_3|X_2) = \frac{\mathbb{P}(X_1, X_2, X_3)}{\mathbb{P}(X_2)} \Rightarrow \mathbb{P}(X_1, X_3|X_2) = P(X_3|X_2)\frac{\mathbb{P}(X_2|X_1)\mathbb{P}(X_1)}{\mathbb{P}(X_2)}$$

Bayes rule:

$$\mathbb{P}(X_1|X_2) = \frac{\mathbb{P}(X_2|X_1)\mathbb{P}(X_1)}{\mathbb{P}(X_2)} \Rightarrow \mathbb{P}(X_1, X_3|X_2) = \mathbb{P}(X_1|X_2)\mathbb{P}(X_3|X_2)$$

CAUSAL STRUCTURE AND STATISTICAL INDEPENDENCE

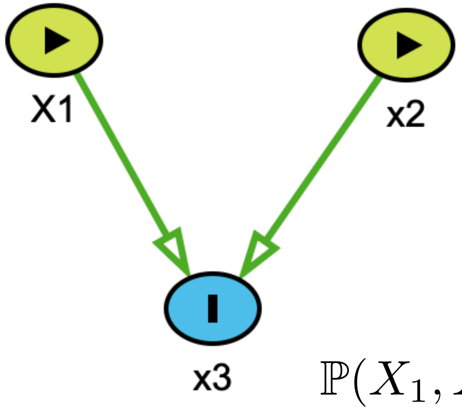
1. All **adjacent nodes** (nodes linked through a single edge) are **dependent.**
2. Statistical **association** can arise between **non-adjacent** nodes through **paths.**
3. A **DAG + Markov condition** define a **set of (conditional) independence facts.**



The same works for the fork structure (X_2 and X_3 are dependent, but becomes independent when conditioning on X_1) → try at home!

CAUSAL STRUCTURE AND STATISTICAL INDEPENDENCE

1. All **adjacent nodes** (nodes linked through a single edge) are **dependent.**
2. Statistical **association** can arise between **non-adjacent** nodes through **paths.**
3. A **DAG + Markov condition** define a **set of (conditional) independence facts.**



For the v-structure, things change: X_1 and X_2 are independent, but they turn out to be dependent when conditioning on X_3, their collider.

$$\mathbb{P}(X_1, X_2) = \sum_{X_3 = x_3} \mathbb{P}(X_1, X_2, X_3 = x_3)$$

Total probabilities low

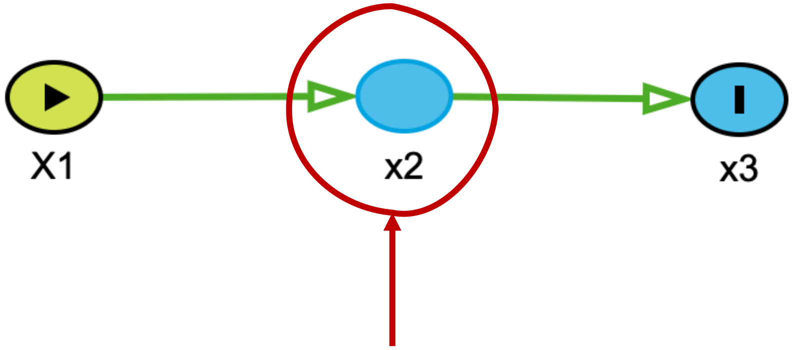
$$= \sum_{X_3 = x_3} \mathbb{P}(X_3 = x_3 | X_1, X_2) \mathbb{P}(X_1) \mathbb{P}(X_2) \quad = P(X_1)\mathbb{P}(X_2) \underbrace{\sum_{X_3 = x_3} \mathbb{P}(X_3 = x_3 | X_1, X_2)}_{=1}$$

$$= P(X_1)\mathbb{P}(X_2)$$

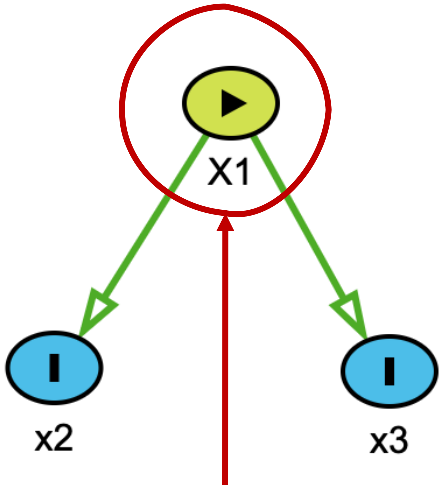
D-SEPARATION

Blocking paths

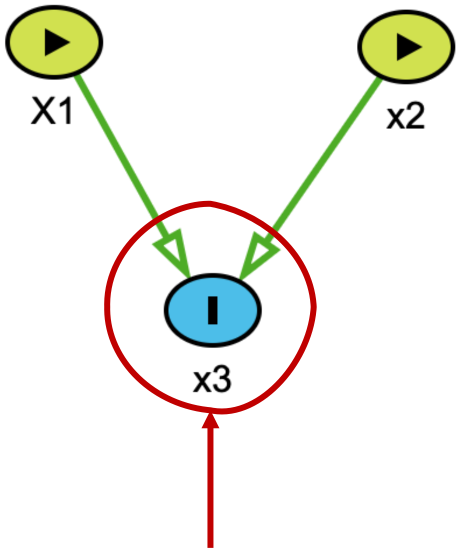
A path **(not talking about direction here!)** between two nodes, X and Y, is blocked conditioning on a node $Z \neq X, Y$ if:
- Z is a mediator along the path between X and Y ($X \rightarrow \cdots \rightarrow Z \rightarrow \cdots \rightarrow Y$) or a confounder ($X \leftarrow \cdots \leftarrow Z \rightarrow \cdots \rightarrow Y$)
- Z is not a collider along the path between X and Y, nor a descendant of a collider.



Mediator:
X_2 is blocking the path $X_1 \rightarrow X_2 \rightarrow X_3$



Confounder:
X_1 is blocking the path $X_2 \leftarrow X_1 \rightarrow X_3$

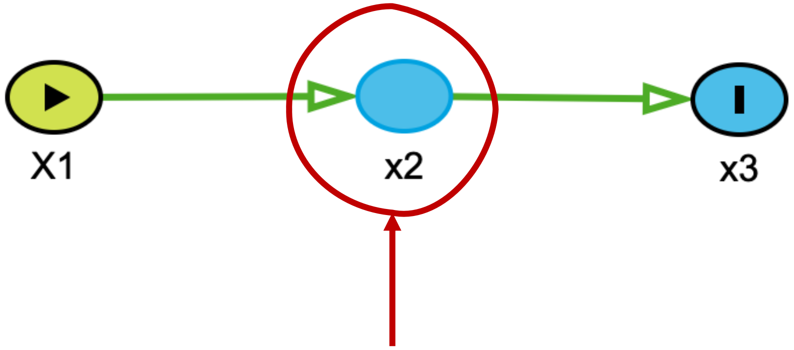


Collider
X_3 is NOT blocking the path $X_1 \rightarrow X_3 \leftarrow X_2$

D-SEPARATION

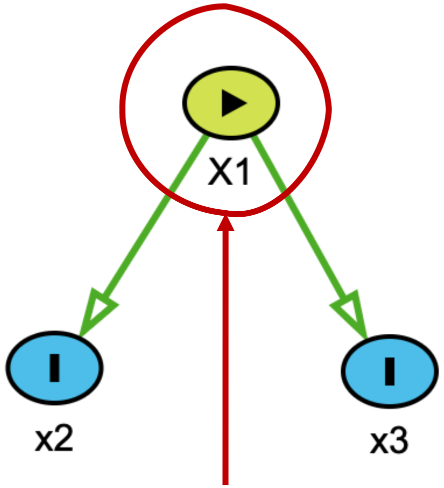
Intuition 1

A path between two nodes, X and Y, is blocked by a node Z if conditioning on Z makes X and Y independent.



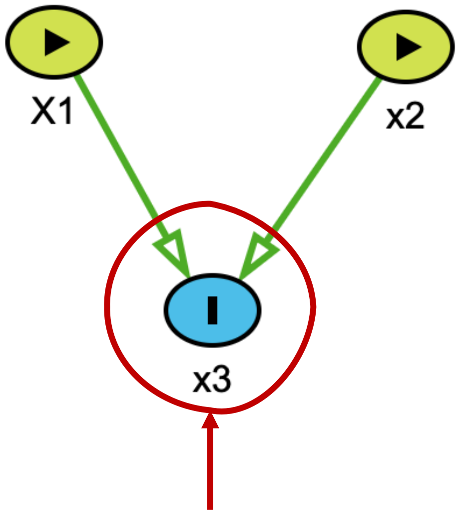
Mediator:

X_2 is blocking the path $X_1 \rightarrow X_2 \rightarrow X_3$



Confounder:

X_1 is blocking the path $X_2 \leftarrow X_1 \rightarrow X_3$



Collider

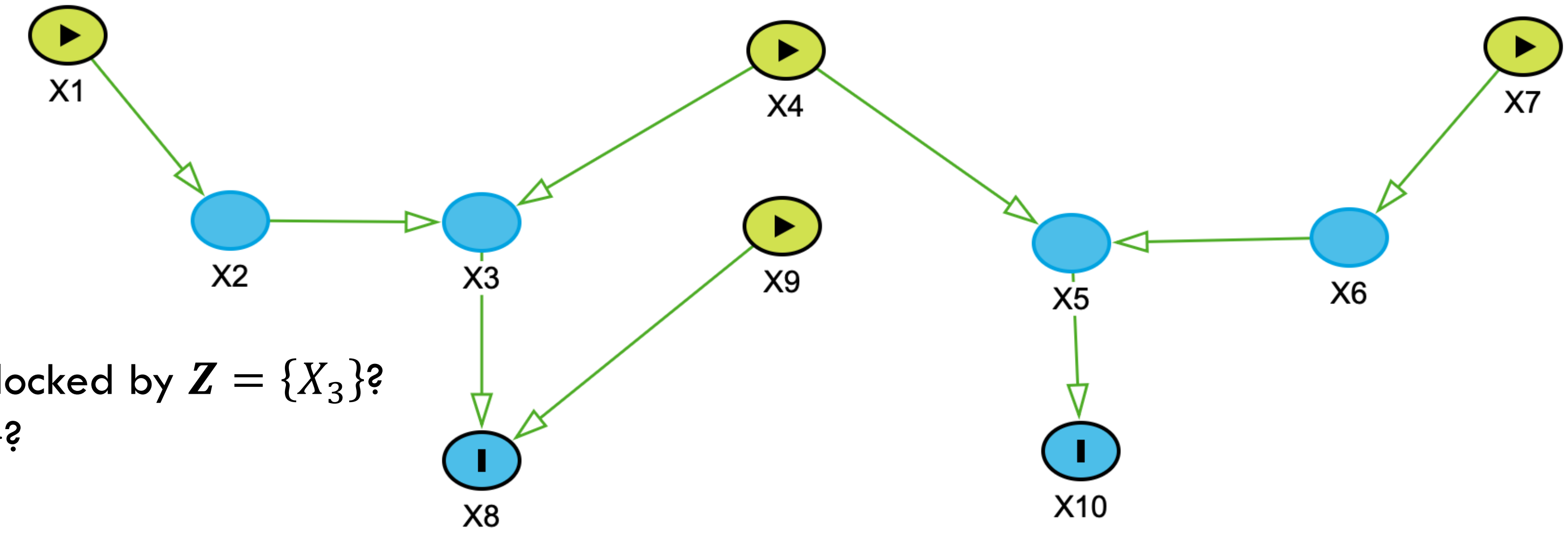
X_1 and X_2 are independent without any conditionaing

D-SEPARATION

This concept can be extended to sets of nodes!

d-separation

Two disjoint sets of nodes, $\boldsymbol{X}$ and $\mathbf{Y}$, are d-separated by a third (potentially empty) disjoint set of nodes $\boldsymbol{Z}$ if **every path** between every node in $\boldsymbol{X}$ and every node in $\mathbf{Y}$ **is blocked given $\boldsymbol{Z}$.**



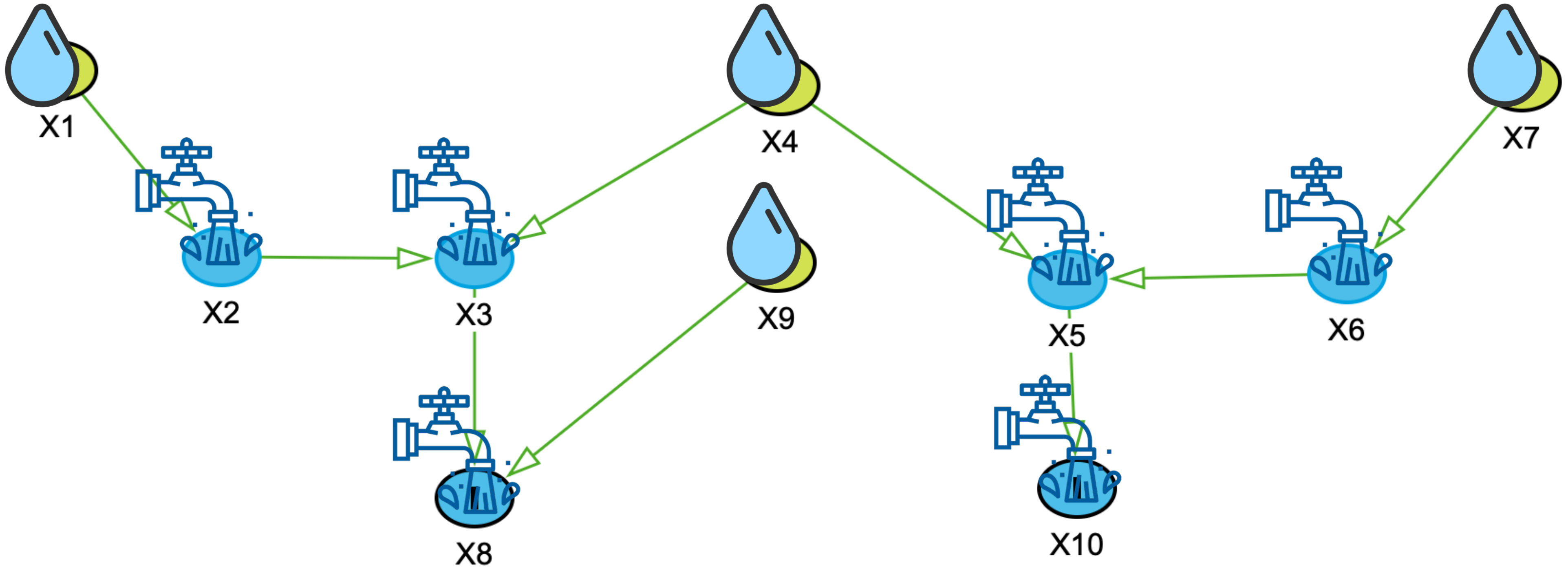
X_1 and X_8 are blocked by $\boldsymbol{Z} = \{X_3\}$?
And by $\boldsymbol{Z} = \{X_2\}$?

X_4 and X_6 are blocked by $\boldsymbol{Z} = \{X_5\}$? Or by $\boldsymbol{Z} = \{X_5, X_{10}\}$?
What set is blocking X_1 and X_9? And X_3 and X_5?

D-SEPARATION

Intuition 2

Think about information flow like water flow through the causal paths



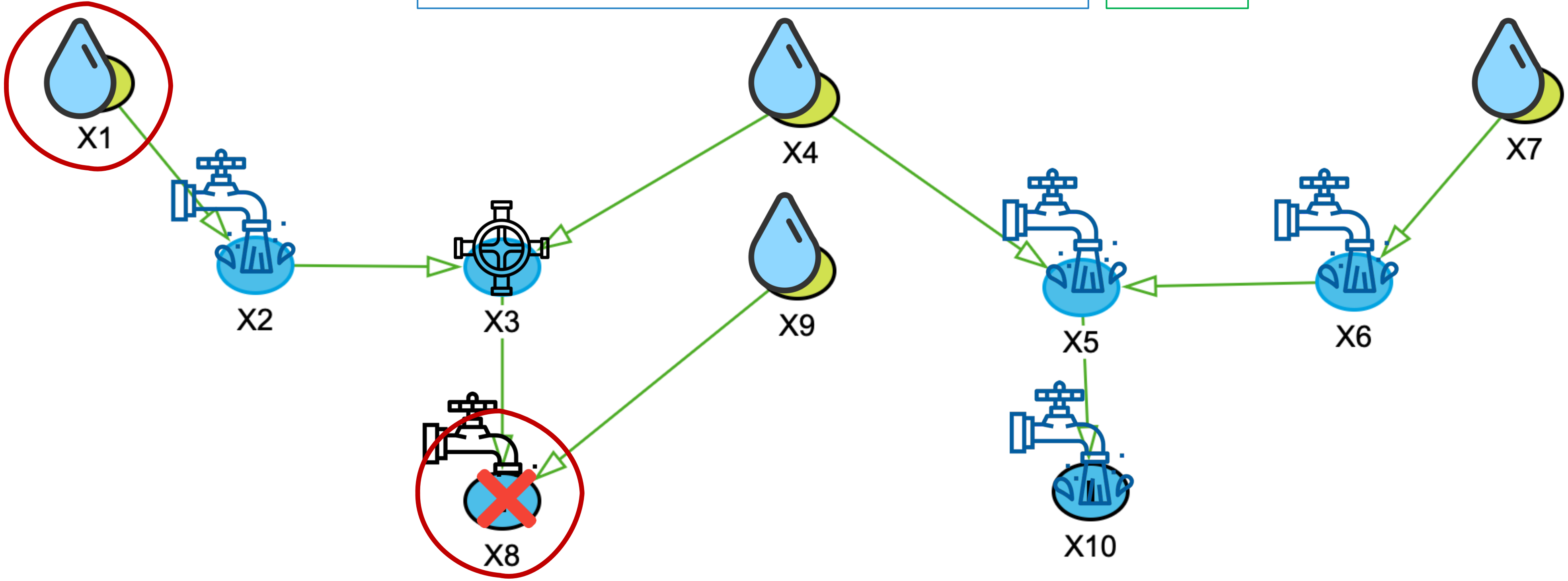
D-SEPARATION

Intuition 2

Think about information flow like water flow through the causal paths

X_1 and X_8 are blocked by $\boldsymbol{Z} = \{X_3\}$?

YES!

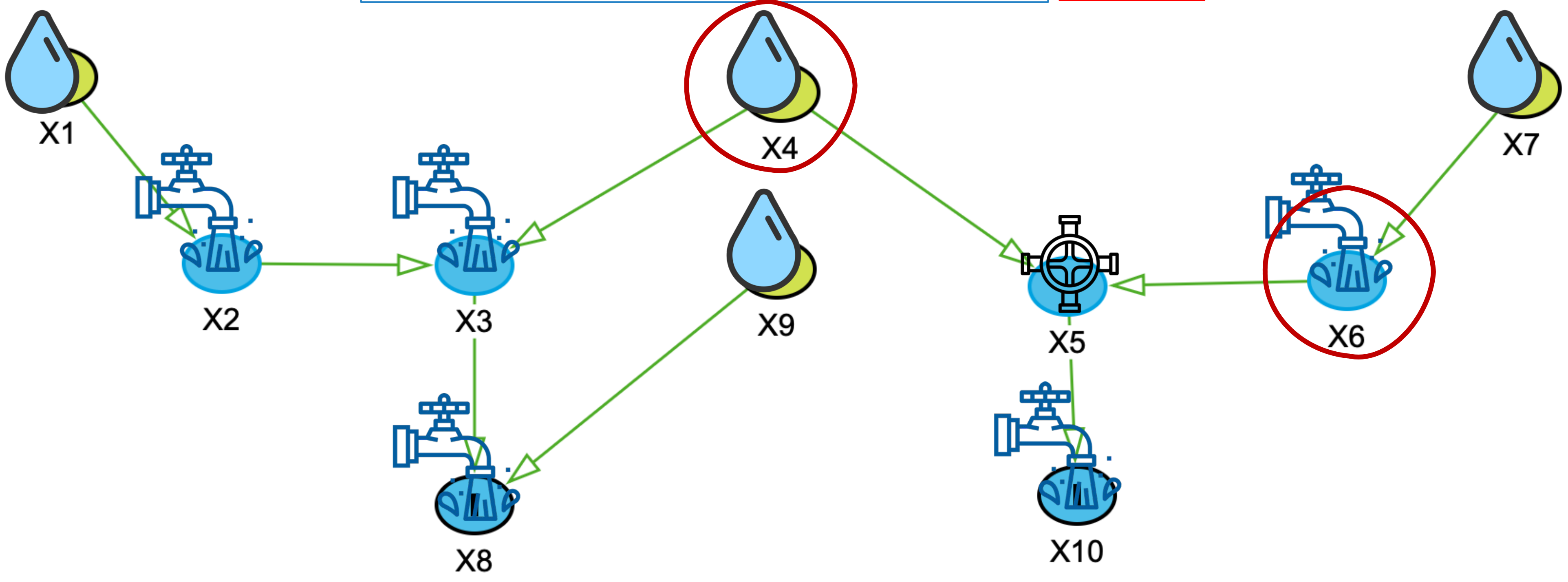


D-SEPARATION

Intuition 2

Think about information flow like water flow through the causal paths

X_4 and X_6 are blocked by $\boldsymbol{Z} = \{X_5\}$? **NO!**



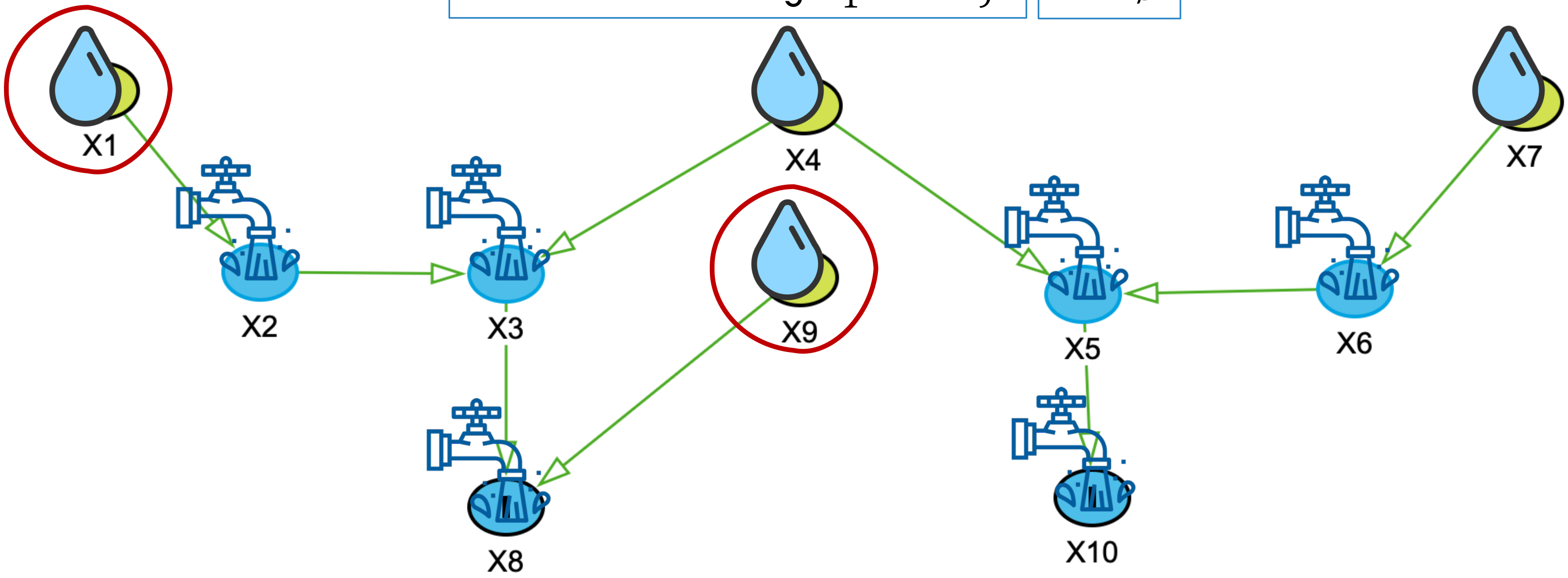
D-SEPARATION

Intuition 2

Think about information flow like water flow through the causal paths

What set is blocking X_1 and X_9?

$\boldsymbol{Z} = \emptyset$



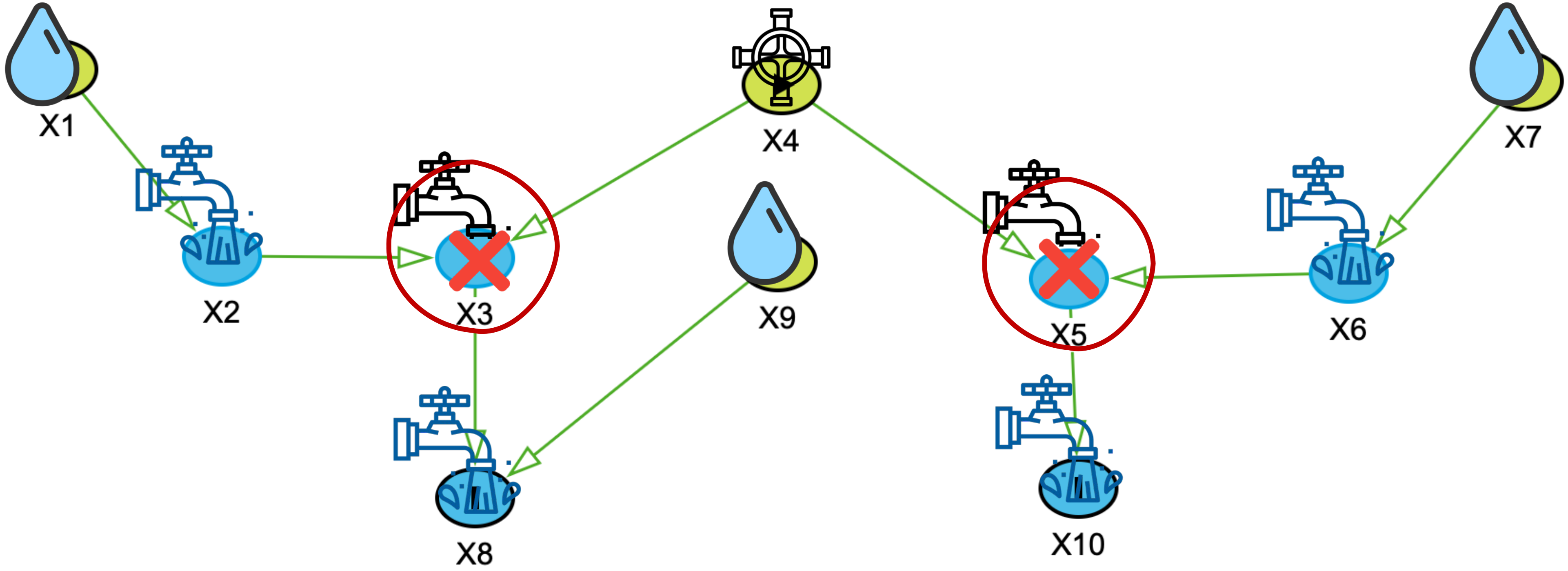
D-SEPARATION

Intuition 2

Think about information flow like water flow through the causal paths

What set is blocking X_3 and X_5?

$\boldsymbol{Z} = \{X_4\}$

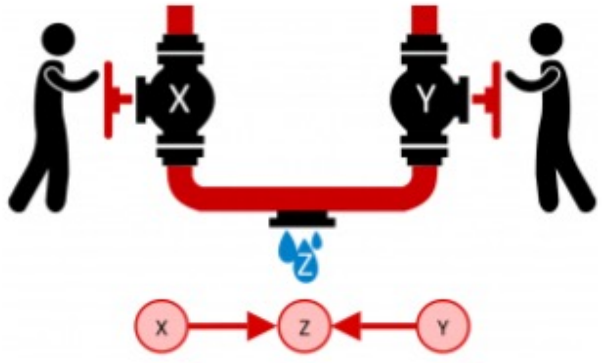


D-SEPARATION

Intuition 3

Conditioning on some variables can either block (→ d-separate) existing flows of information (chain/fork), or activate/open new ones (v-structure).

- **Chain/fork are open by default:** you MUST condition on the intermediate variable (mediator or confounder) to block the path.
- **V-structures are closed by default:** this path becomes open only if you condition on the collider (or its descendant).



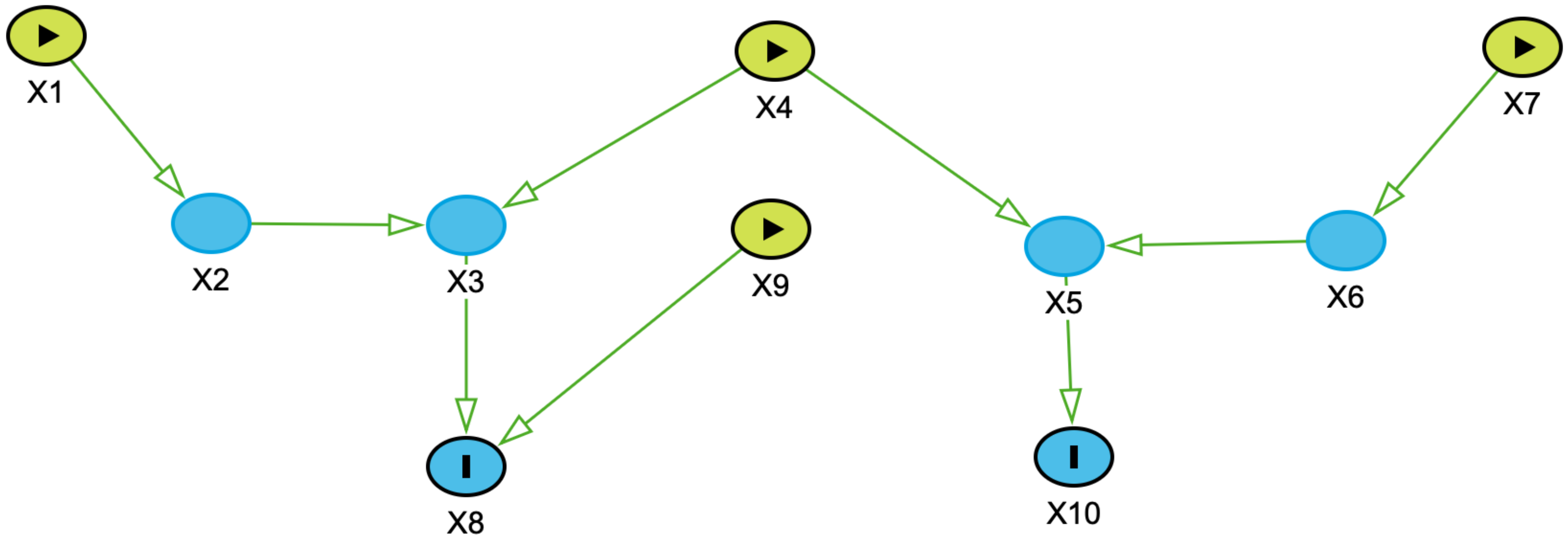
- If X turns rate up, it increases the flow in Z but has no effect on Y.
- If I condition on Z (→ I fix Z to a specific rate of flow), then when X turns up, Y must do the opposite. So someone looking at X and Y under this condition (strata of Z) would see a correlation between them.

D-SEPARATION

Connection with the Markov assumption

If two (sets of) variables are d-separated in a causal DAG, then they are (conditionally) independent.

d-separation $\Longrightarrow$ conditional independence



X_1 and X_8 are d-separated by $X_3 \Rightarrow X_1 \perp\!\!\!\perp X_8 | X_3$