Causal Data Science

Lecture 12.1: Invariant Causal Prediction, Joint Causal Inference

Lecturer: Sara Magliacane

UvA - Spring 2024

Causal discovery overview

Constraint-based causal discovery

- Conditional independence tests
- Observational data
- Output: MEC
- SGS, PC

Score-based causal discovery

- Penalised likelihood
- Observational data
- Output: MEC
- GES

Restricted models

- Nonlinear additive noise, Linear Non-Gaussianity
- Observational data
- Output: DAG
- RESIT, LINGAM

Interventional causal discovery / causal invariance

- Observational and Interventional data
- Output: parents of Y or I-MEC
- ICP, JCI

Learning from interventional data - intuition

Until now we have only used **observational data.**

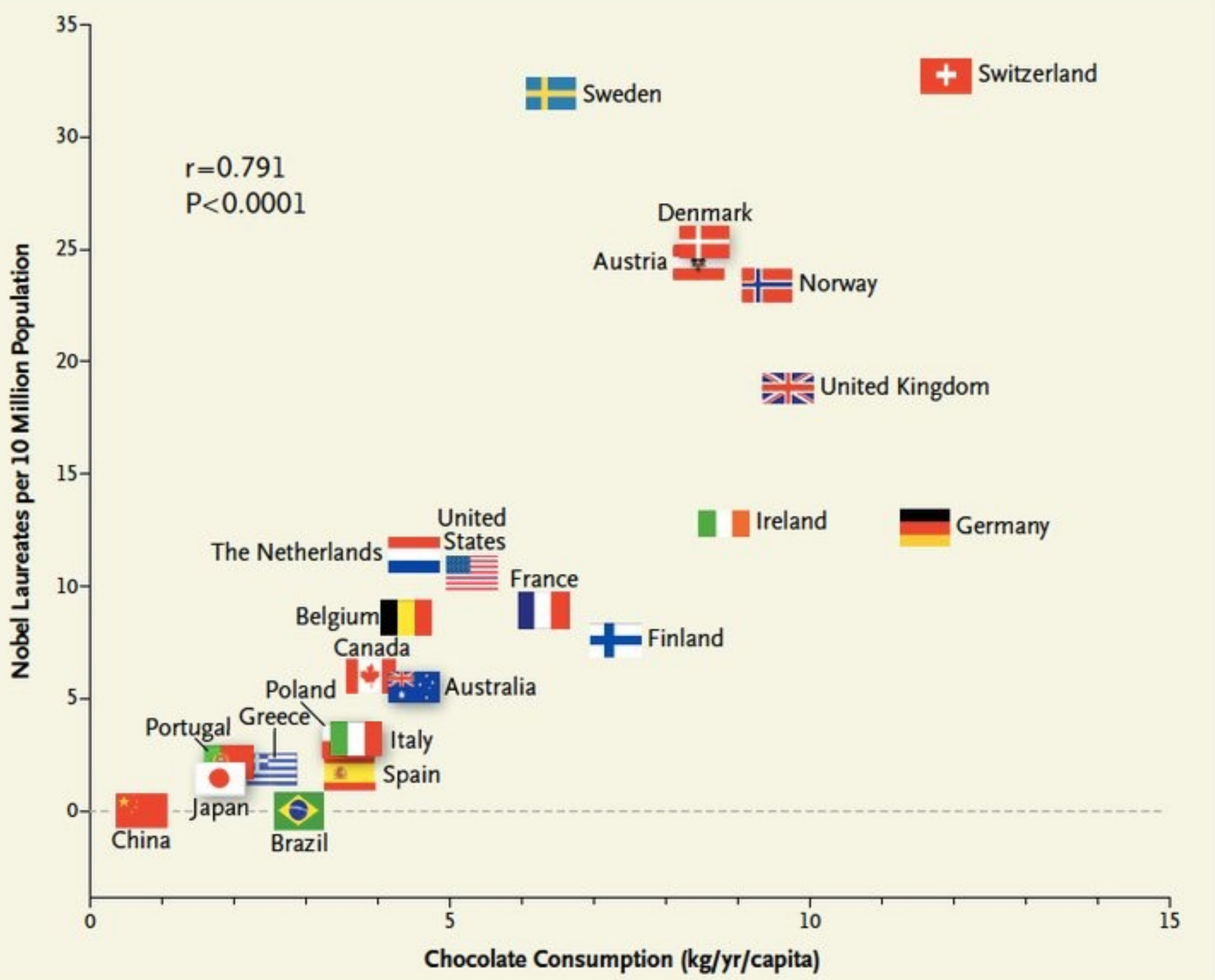


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Learning from interventional data - intuition

Hypothetical world: we perform the experiment and see these results:

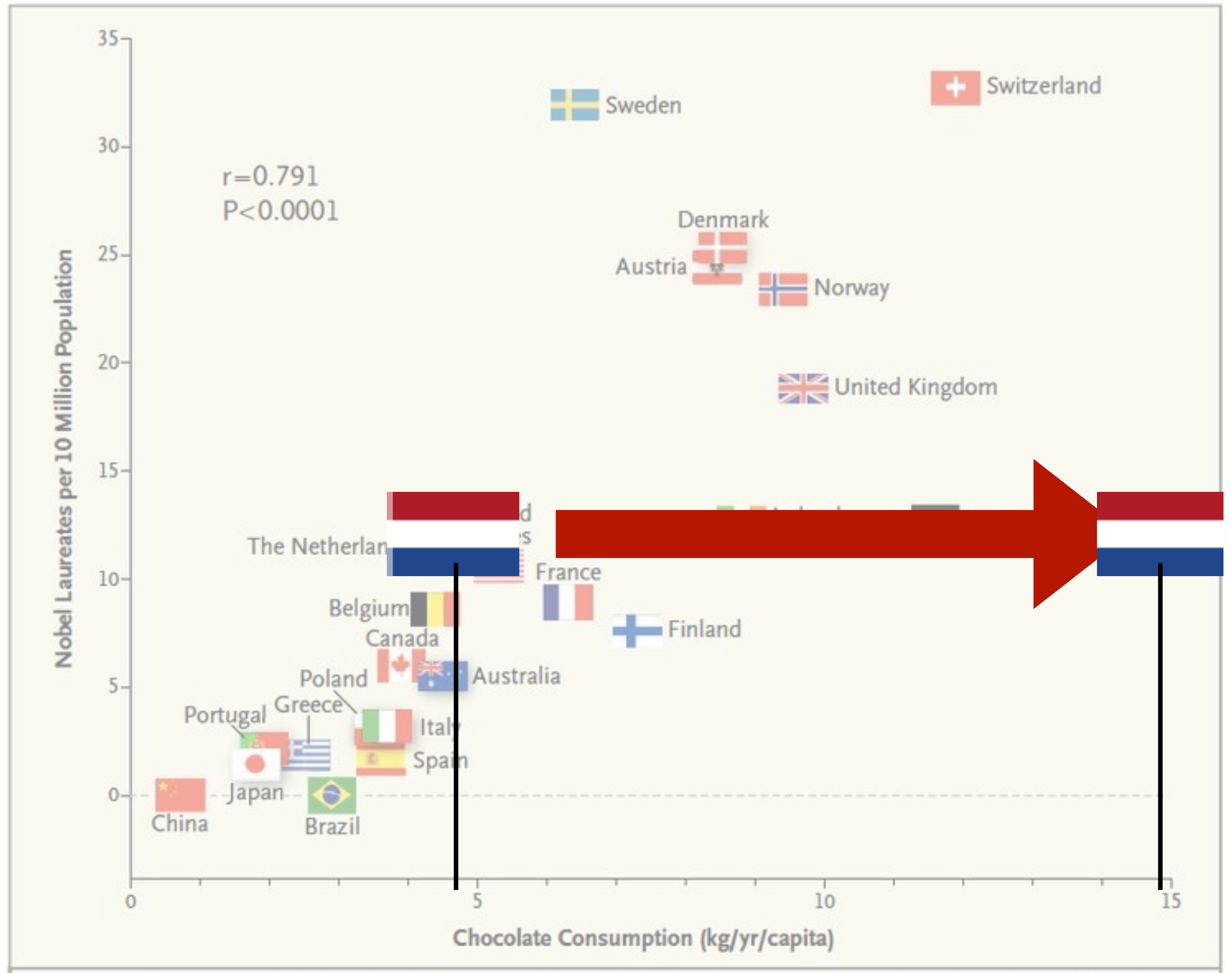


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

NL eats more chocolate => nothing changes

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Learning from interventional data - intuition

Hypothetical world: we perform the experiment and see these results:

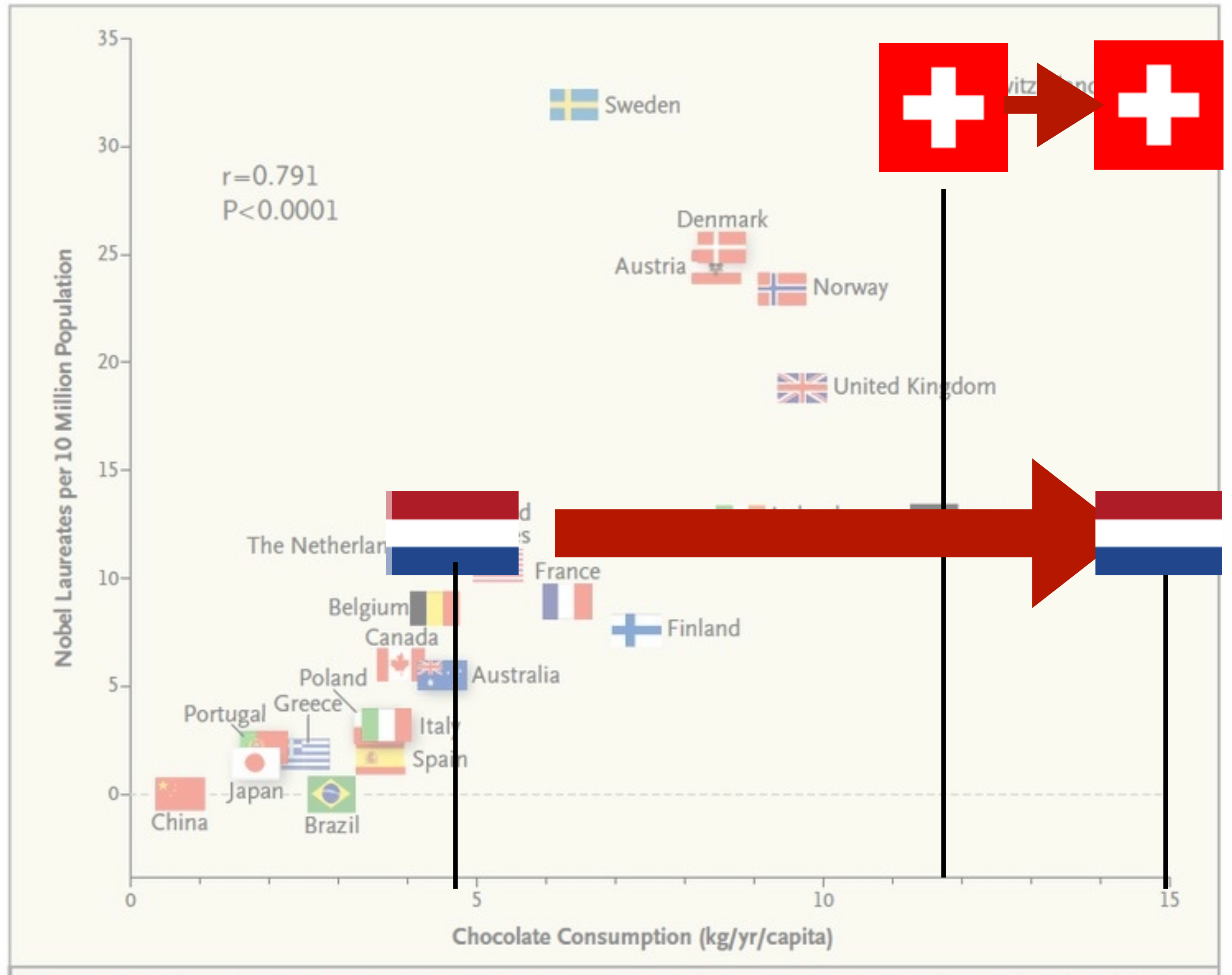


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

NL eats more chocolate => nothing changes

… and similarly for other countries (and other values)

Chocolate does not cause Nobel prizes

[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Learning from interventional data - intuition

Hypothetical world: we perform the experiment and see these results:

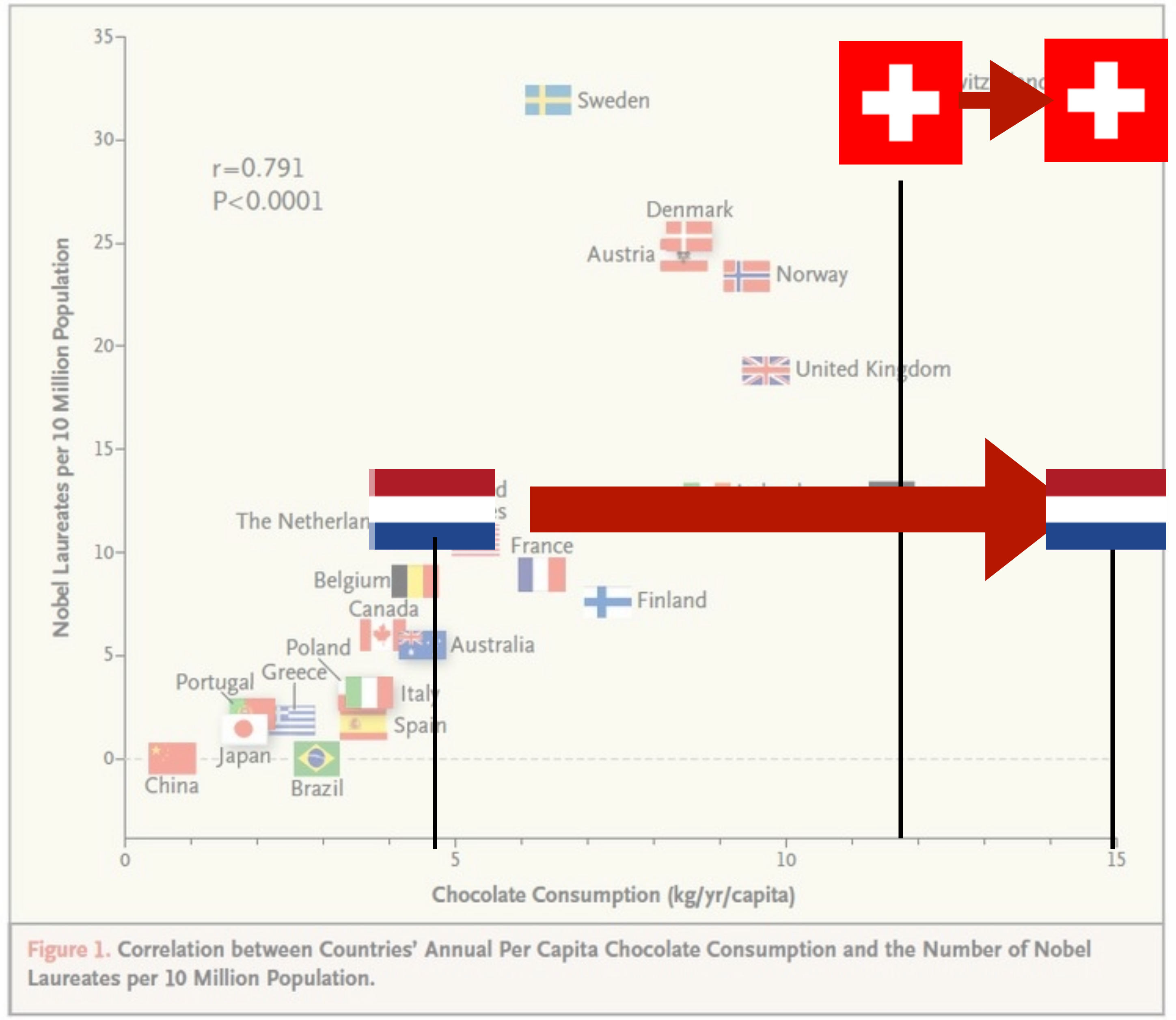
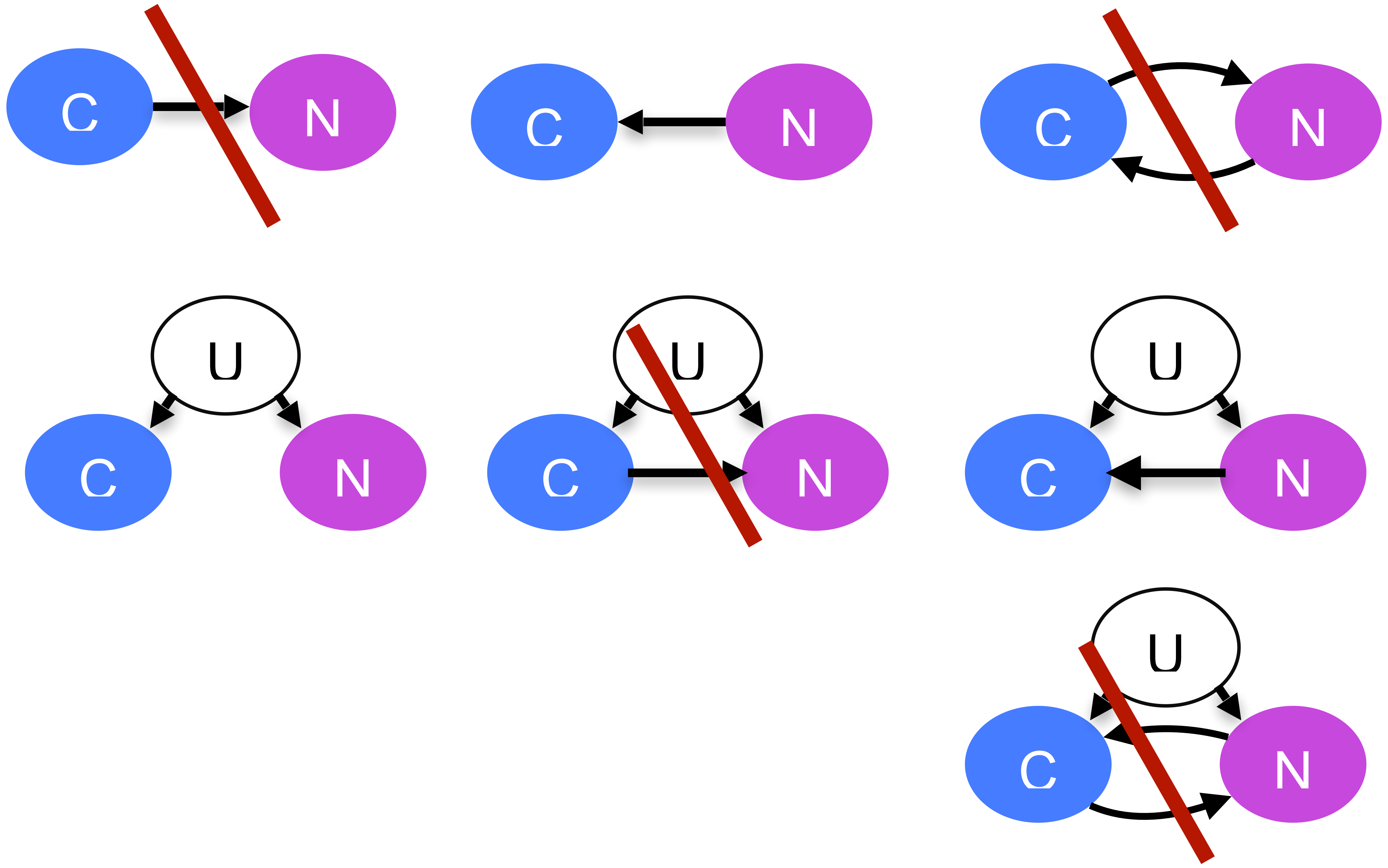


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

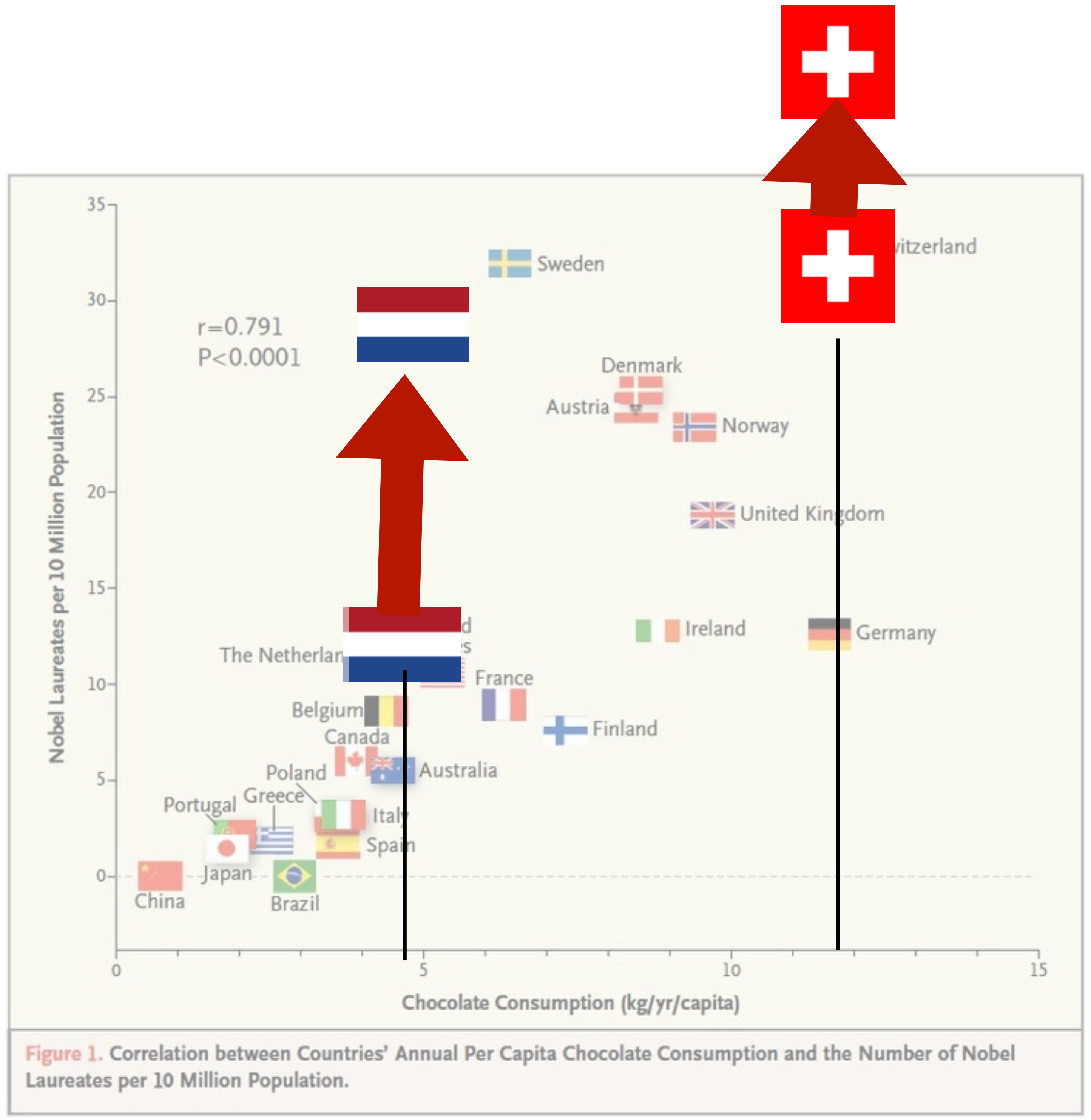
[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Chocolate does not cause Nobel prizes

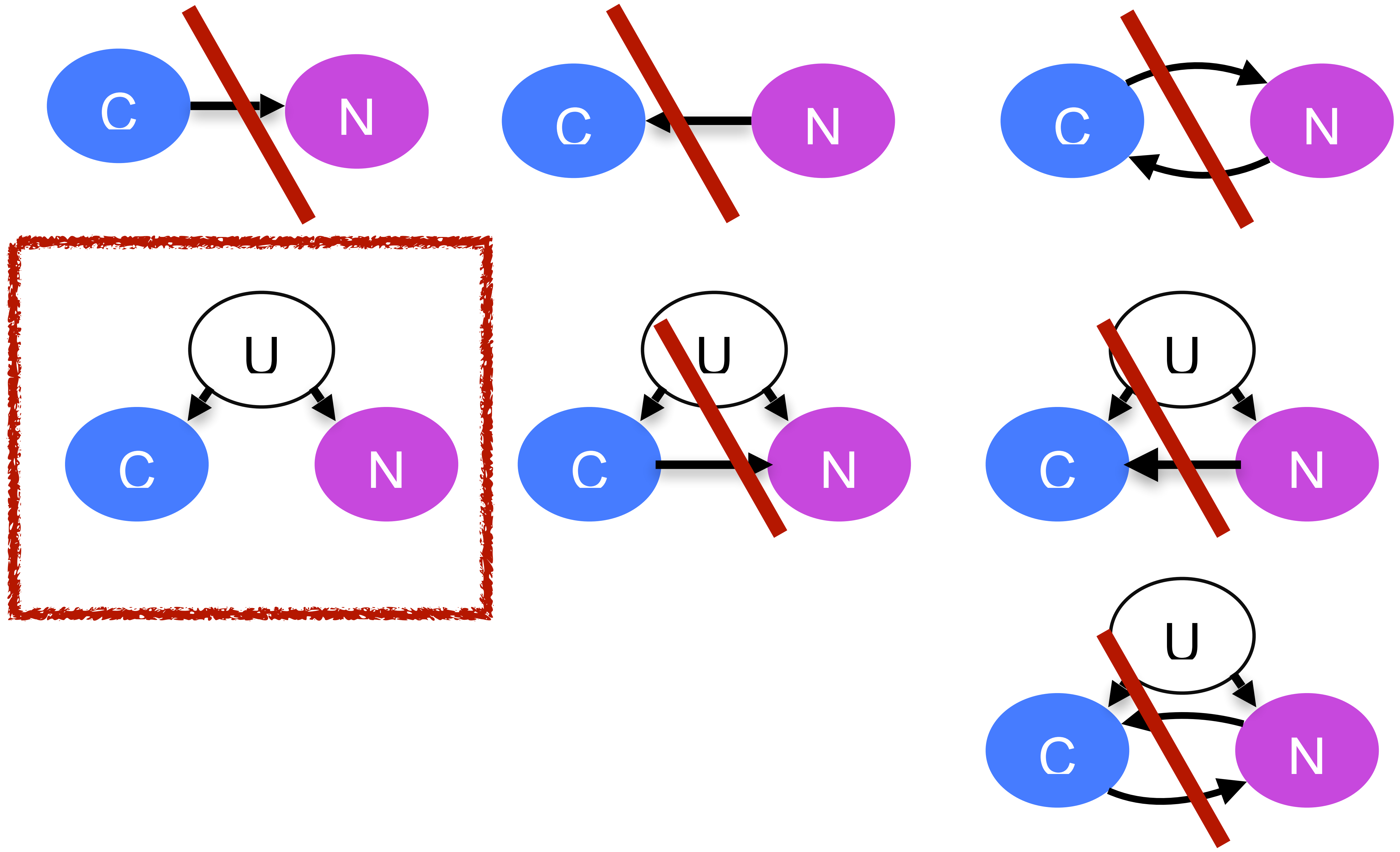


Learning from interventional data - intuition

Hypothetical world: we perform another experiment and see these results:



Nobel does not cause chocolate



[Messerli, 2012] https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Interventions on a variable identify its parents and descendants

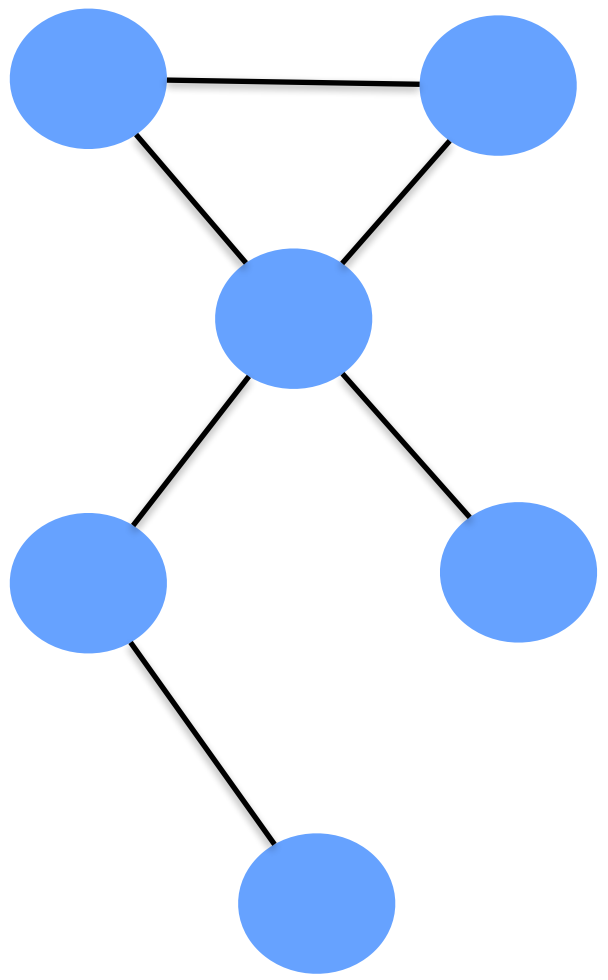
- The skeleton (and v-structures) can be identified from observational data
- Intervening on a node i identifies its parents and descendants:
  - For all j **adjacent** i **in** G **(** $j - i$ **in** G **):**
    - If j is not adjacent i in $G_{do(i)}$ then $j \in \mathrm{Pa}(i)$.

- If $P(X_j \,|\, \mathrm{do}(X_i)) \neq P(X_j)$, then $j \in \mathrm{Desc}(i)$. If also $j - i$ in G, $j \in \mathrm{Ch}(i)$.

$$P(A \,|\, \mathrm{do}(S)) \neq P(A) \wedge \; S - A \text{ in } G \implies A \in \mathrm{Ch}(S)$$

Intervention design/Experiment selection

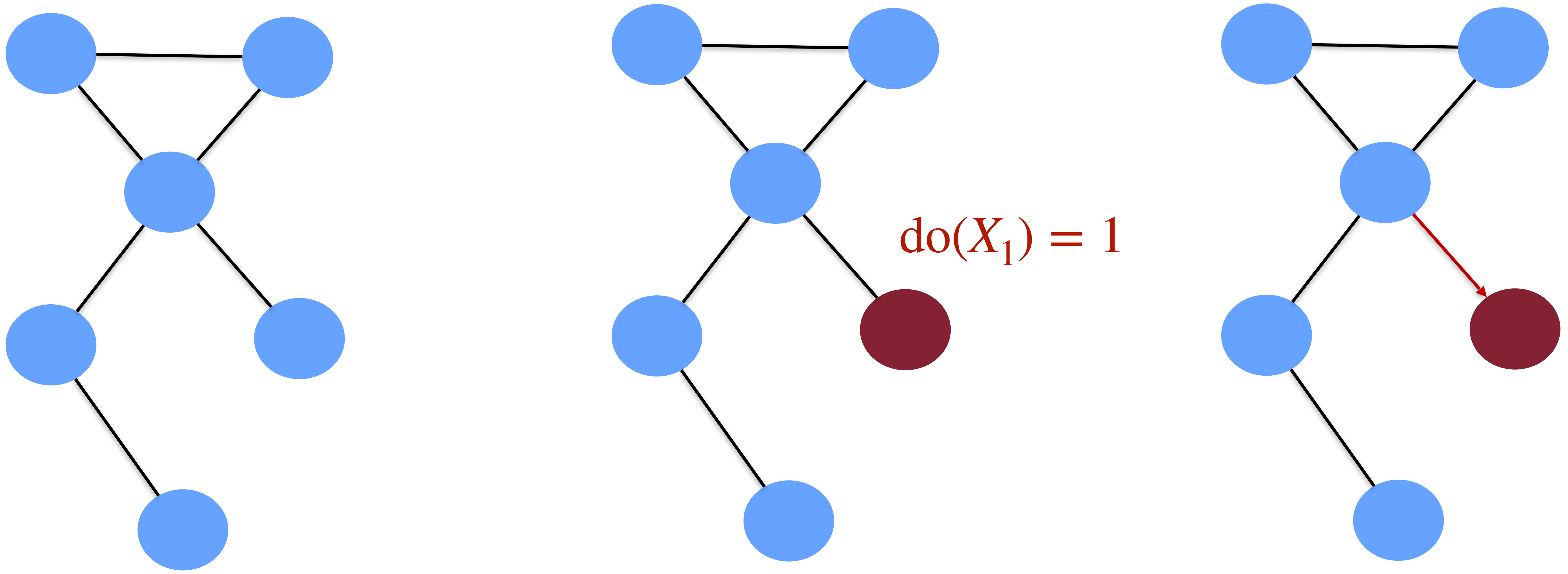
Design a set of interventions, so that we can **accurately** reconstruct **as much as possible** the causal graph **with the least samples,** also when **noisy**



https://papers.nips.cc/paper/2019/hash/5ee5605917626676f6a285fa4c10f7b0-Abstract.html https://arxiv.org/abs/2011.00641

Intervention design/Experiment selection

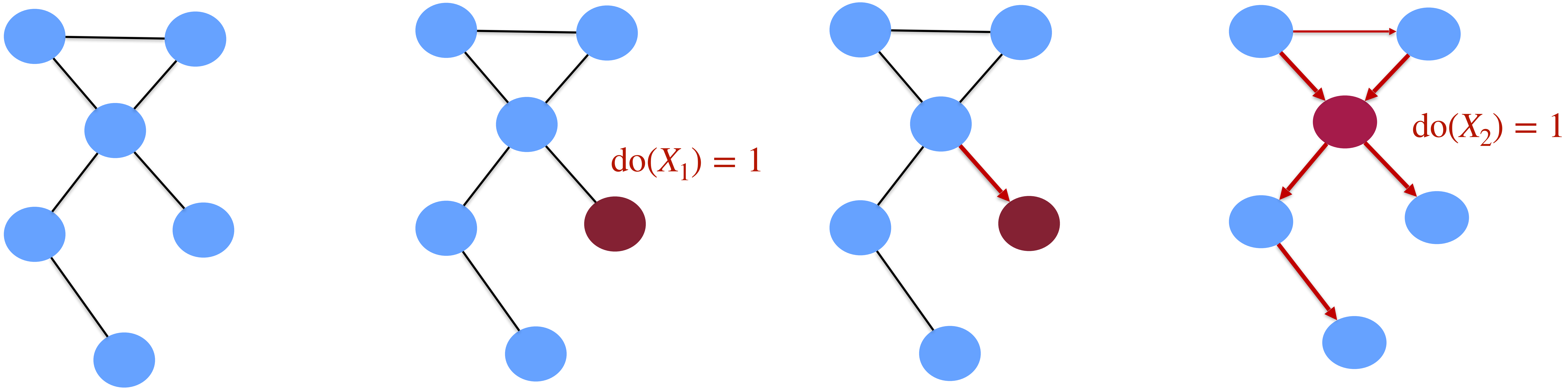
Design a set of interventions, so that we can **accurately** reconstruct **as much as possible** the causal graph **with the least samples,** also when **noisy**



https://papers.nips.cc/paper/2019/hash/5ee5605917626676f6a285fa4c10f7b0-Abstract.html https://arxiv.org/abs/2011.00641

Intervention design/Experiment selection

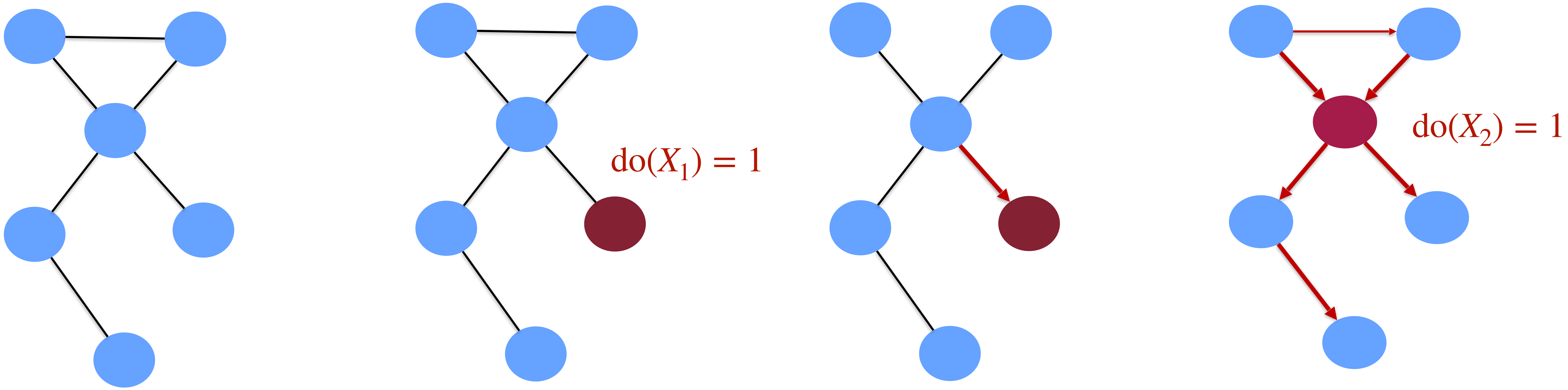
Design a set of interventions, so that we can **accurately** reconstruct **as much as possible** the causal graph **with the least samples,** also when **noisy**



https://papers.nips.cc/paper/2019/hash/5ee5605917626676f6a285fa4c10f7b0-Abstract.html https://arxiv.org/abs/2011.00641

Intervention design/Experiment selection

Design a set of interventions, so that we can **accurately** reconstruct **as much as possible** the causal graph **with the least samples,** also when **noisy**



We formalised an algorithm/policy based on the concept of **central node** for forests with noisy interventions and for DAGs with noiseless interventions.

https://papers.nips.cc/paper/2019/hash/5ee5605917626676f6a285fa4c10f7b0-Abstract.html https://arxiv.org/abs/2011.00641

Learning from multiple contexts

- In **intervention design** we could decide which intervention to perform
  - We usually also had **known intervention targets**, e.g. $\mathrm{do}(S = 1)$
- What if we cannot decide interventions, but instead somebody gives us a **set of data from multiple contexts?**
  - Possibly **with unknown intervention targets**
  - Possibly **soft interventions** instead of **perfect interventions**

Invariant Causal Prediction example

$$\begin{cases} X_1 = \varepsilon_1 \\ Y = X_1 + \varepsilon_Y \\ X_2 = Y + \varepsilon_{X_2} \end{cases}$$

$X_1 \to Y \to X_2$

$\varepsilon_1, \varepsilon_Y \sim N(0,1)$, $\varepsilon_{X_2} \sim N(0, 0.01)$

M1: $Y \sim X_1$ M2: $Y \sim X_2$ M2 has smaller error

Invariant Causal Prediction example

$$\begin{cases} X_1 = \varepsilon_1 \\ Y = X_1 + \varepsilon_Y \\ X_2 = Y + \varepsilon_{X_2} \end{cases}$$

$\varepsilon_1, \varepsilon_Y \sim N(0,1)$, $\varepsilon_{X_2} \sim N(0, 0.01)$

$X_1 \to Y \to X_2$

$X_1 \to Y \quad X_2 \quad do(X_2)$

$M1: Y \sim X_1$ $\quad$ $M2: Y \sim X_2$

M2 has smaller error

but it fails in $do(X_2)$

Invariant Causal Prediction example

$$\begin{cases} X_1 = \varepsilon_1 \\ Y = X_1 + \varepsilon_Y \\ X_2 = Y + \varepsilon_{X_2} \end{cases}$$

$\varepsilon_1, \varepsilon_Y \sim N(0,1)$, $\varepsilon_{X_2} \sim N(0, 0.01)$

$X_1 \to Y \to X_2$

M1: $Y \sim X_1$ M2: $Y \sim X_2$

Using causal parents of Y as predictors is robust under distributions shifts.

Can we reverse somehow this statement to find the causal parents from multiple distributions?

Invariant Causal Prediction (ICP) [Peters et al 2016]

- Given a target variable Y and features $(X_1, \ldots, X_p)$, we want to **find the causal parents of Y, i.e. $\mathrm{Pa}(Y)$**

Invariant Causal Prediction (ICP) [Peters et al 2016]

- Given a target variable Y and features $(X_1, \ldots, X_p)$, we want to **find the causal parents of Y, i.e. $\mathrm{Pa}(Y)$**
- We assume we have access to **a set of different environments E** (e.g. interventional or observational data), s.t. for $e \in E$, $(X_1^e, \ldots, X_p^e, Y^e) \sim P^e$

Invariant Causal Prediction (ICP) [Peters et al 2016]

- Given a target variable Y and features $(X_1, \dots, X_p)$, we want to **find the causal parents of Y, i.e. $\mathrm{Pa}(Y)$**
- We assume we have access to **a set of different environments E** (e.g. interventional or observational data), s.t. for $e \in E$, $(X_1^e, \dots, X_p^e, Y^e) \sim P^e$
- We further assume that in **none of the environments Y is intervened upon**
- We can then show that $e, f \in E : P^e(Y^e \mid \mathrm{Pa}(Y^e)) = P^f(Y^f \mid \mathrm{Pa}(Y^f))$

Invariant Causal Prediction example

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

Invariant Causal Prediction example

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 1 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 10Y + \epsilon_Y \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

Invariant Causal Prediction example

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases} \qquad E = 0$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 1 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases} \qquad E = 1$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 10Y + \epsilon_Y \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases} \qquad E = 2$$

Invariant Causal Prediction example

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases} \quad E = 0$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 1 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases} \quad E = 1$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 10Y + \epsilon_Y \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases} \quad E = 2$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = \begin{cases} -2Y + \epsilon_2 & \text{if } E = 0 \\ 1 & \text{if } E = 1 \\ 10Y + \epsilon_Y & \text{if } E = 2 \end{cases} \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

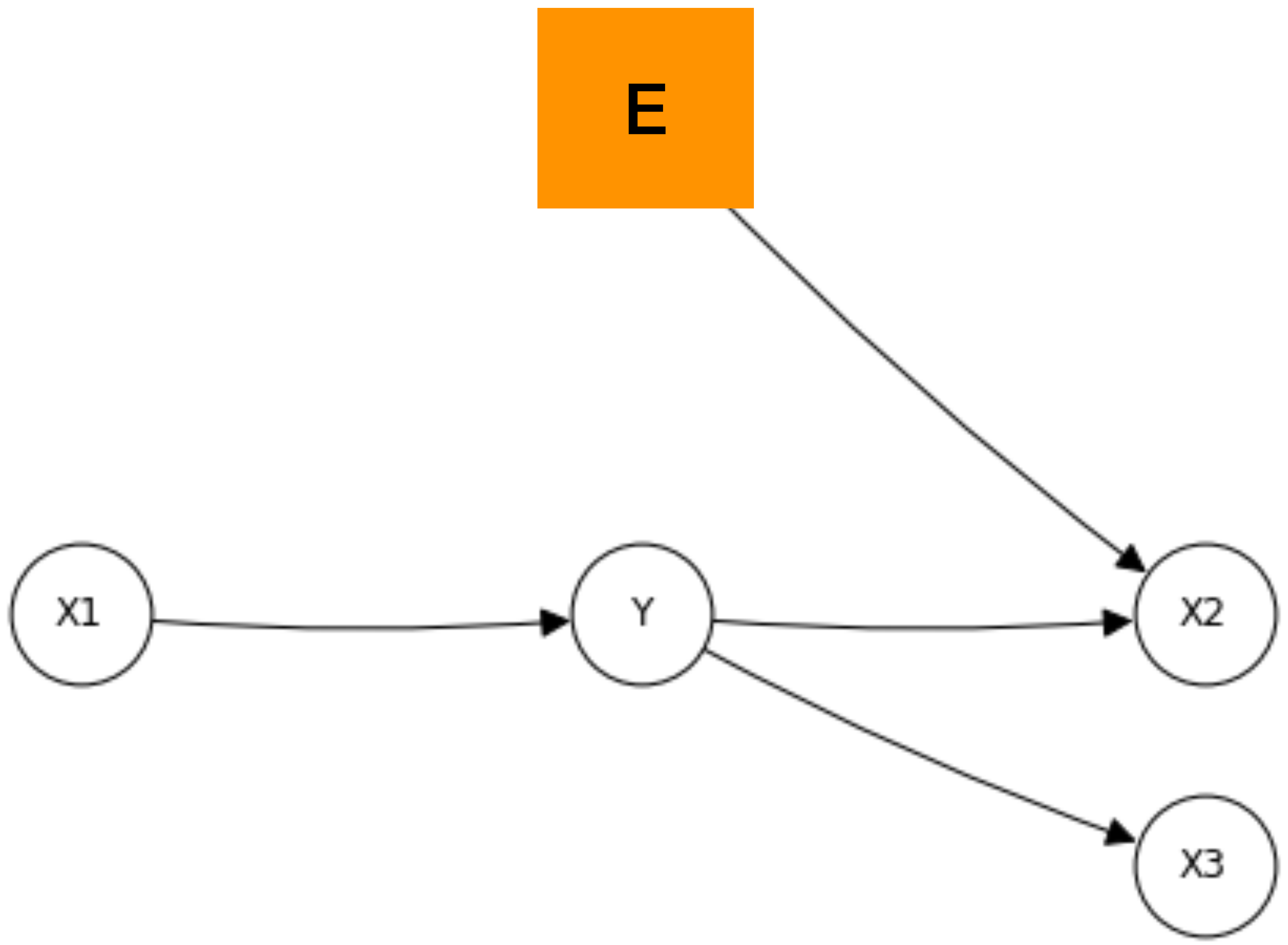
Invariant Causal Prediction example

$$E = 0: \quad \begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$$E = 1: \quad \begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 1 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$$E = 2: \quad \begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 10Y + \epsilon_Y \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = \begin{cases} -2Y + \epsilon_2 & \text{if } E = 0 \\ 1 & \text{if } E = 1 \\ 10Y + \epsilon_Y & \text{if } E = 2 \end{cases} \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$



Invariant Causal Prediction example 2

env e_1

$$\begin{cases} X_1 = \epsilon_1 \\ Y = X_1 + \epsilon_Y \\ X_2 = Y + \epsilon_2 \\ \epsilon_1, \epsilon_Y \sim \mathcal{N}(0,1) \\ \epsilon_2 \sim \mathcal{N}(0,0.01) \end{cases}$$

$X_1 \rightarrow Y \rightarrow X_2$

env e_2

$$\begin{cases} X_1 = 1 \\ Y = X_1 + \epsilon_Y \\ X_2 = 1 \\ \epsilon_1, \epsilon_Y \sim \mathcal{N}(0,1) \\ \epsilon_2 \sim \mathcal{N}(0,0.01) \end{cases}$$

$X_1 \rightarrow Y \quad X_2$

Invariant Causal Prediction example 2

env e_1

$$\begin{cases} X_1 = \epsilon_1 \\ Y = X_1 + \epsilon_Y \\ X_2 = Y + \epsilon_2 \\ \epsilon_1, \epsilon_Y \sim \mathcal{N}(0,1) \\ \epsilon_2 \sim \mathcal{N}(0,0.01) \end{cases}$$

$\rightarrow \mathrm{do}(X_1 = 1) \rightarrow$

No intervention on Y by assumption

$\rightarrow \mathrm{do}(X_2 = 1) \rightarrow$

env e_2

$$\begin{cases} X_1 = 1 \\ Y = X_1 + \epsilon_Y \\ X_2 = 1 \\ \epsilon_1, \epsilon_Y \sim \mathcal{N}(0,1) \\ \epsilon_2 \sim \mathcal{N}(0,0.01) \end{cases}$$

$$P^{e_1}(Y^{e_1} \mid \mathrm{Pa}(Y^{e_1})) = P^{e_2}(Y^{e_2} \mid \mathrm{Pa}(Y^{e_2}))$$

$$P^{e_1}(Y \mid X_1 = x_1) = x_1 + N(0,1) = P^{e_2}(Y \mid X_1 = x_1)$$

Invariant Causal Prediction example 2

env e_1

$$\begin{cases} X_1 = \epsilon_1 \\ Y = X_1 + \epsilon_Y \\ X_2 = Y + \epsilon_2 \\ \epsilon_1, \epsilon_Y \sim \mathcal{N}(0,1) \\ \epsilon_2 \sim \mathcal{N}(0,0.01) \end{cases}$$

$X_1 \to Y \to X_2$

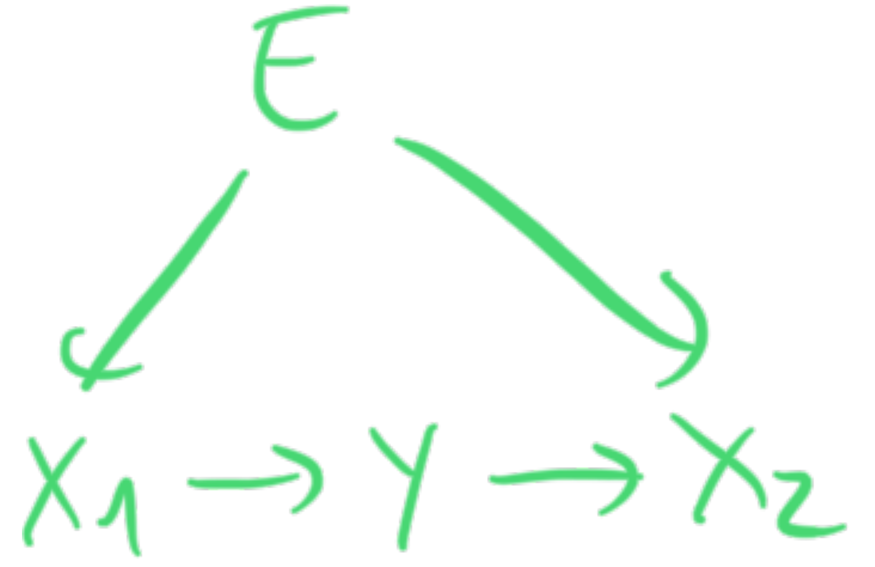
env e_2

$$\begin{cases} X_1 = 1 \\ Y = X_1 + \epsilon_Y \\ X_2 = 1 \\ \epsilon_1, \epsilon_Y \sim \mathcal{N}(0,1) \\ \epsilon_2 \sim \mathcal{N}(0,0.01) \end{cases}$$

$X_1 \to Y \quad X_2$

$e_1 + e_2$

$$\begin{cases} X_1 = \begin{cases} \epsilon_1 \text{ if } E = e_1 \\ 1 \text{ if } E = e_2 \end{cases} \\ Y = X_1 + \epsilon_Y \\ X_2 = \begin{cases} Y + \epsilon_2 \text{ if } E = e_1 \\ 1 \qquad\quad \text{ if } E = e_2 \end{cases} \\ \epsilon_1, \epsilon_Y \sim \mathcal{N}(0,1) \\ \epsilon_2 \sim \mathcal{N}(0,0.01) \end{cases}$$



Invariant Causal Prediction example 2

$e_1 + e_2$

$$
\begin{cases}
X_1 = \begin{cases} \epsilon_1 & \text{if } E = e_1 \\ 1 & \text{if } E = e_2 \end{cases} \\
Y = X_1 + \epsilon_Y \\
X_2 = \begin{cases} Y + \epsilon_2 & \text{if } E = e_1 \\ 1 & \text{if } E = e_2 \end{cases} \\
\epsilon_1, \epsilon_Y \sim \mathcal{N}(0,1) \\
\epsilon_2 \sim \mathcal{N}(0,0.01)
\end{cases}
$$

$$
\begin{tikzcd}
 & E \arrow[dl] \arrow[dr] & \\
X_1 \arrow[r] & Y \arrow[r] & X_2
\end{tikzcd}
$$

E is similar to intervention variables, but now it can be associated with multiple distributions and multiple causal variables (and we potentially want to learn which)

$$P^{e_1}(Y^{e_1} \mid \mathrm{Pa}(Y^{e_1})) = P^{e_2}(Y^{e_2} \mid \mathrm{Pa}(Y^{e_2}))$$

Invariant Causal Prediction example 2

$e_1 + e_2$

$$\begin{cases} X_1 = \begin{cases} \epsilon_1 & \text{if } E = e_1 \\ 1 & \text{if } E = e_2 \end{cases} \\ Y = X_1 + \epsilon_Y \\ X_2 = \begin{cases} Y + \epsilon_2 & \text{if } E = e_1 \\ 1 & \text{if } E = e_2 \end{cases} \\ \epsilon_1, \epsilon_Y \sim \mathcal{N}(0,1) \\ \epsilon_2 \sim \mathcal{N}(0,0.01) \end{cases}$$

$$\begin{tikzcd} & E \arrow[dl] \arrow[dr] & \\ X_1 \arrow[r] & Y \arrow[r] & X_2 \end{tikzcd}$$

Assuming the Markov and faithfulness assumptions, this corresponds to d-separations in the true graph

$$P^{e_1}(Y^{e_1} \mid \mathrm{Pa}(Y^{e_1})) = P^{e_2}(Y^{e_2} \mid \mathrm{Pa}(Y^{e_2}))$$

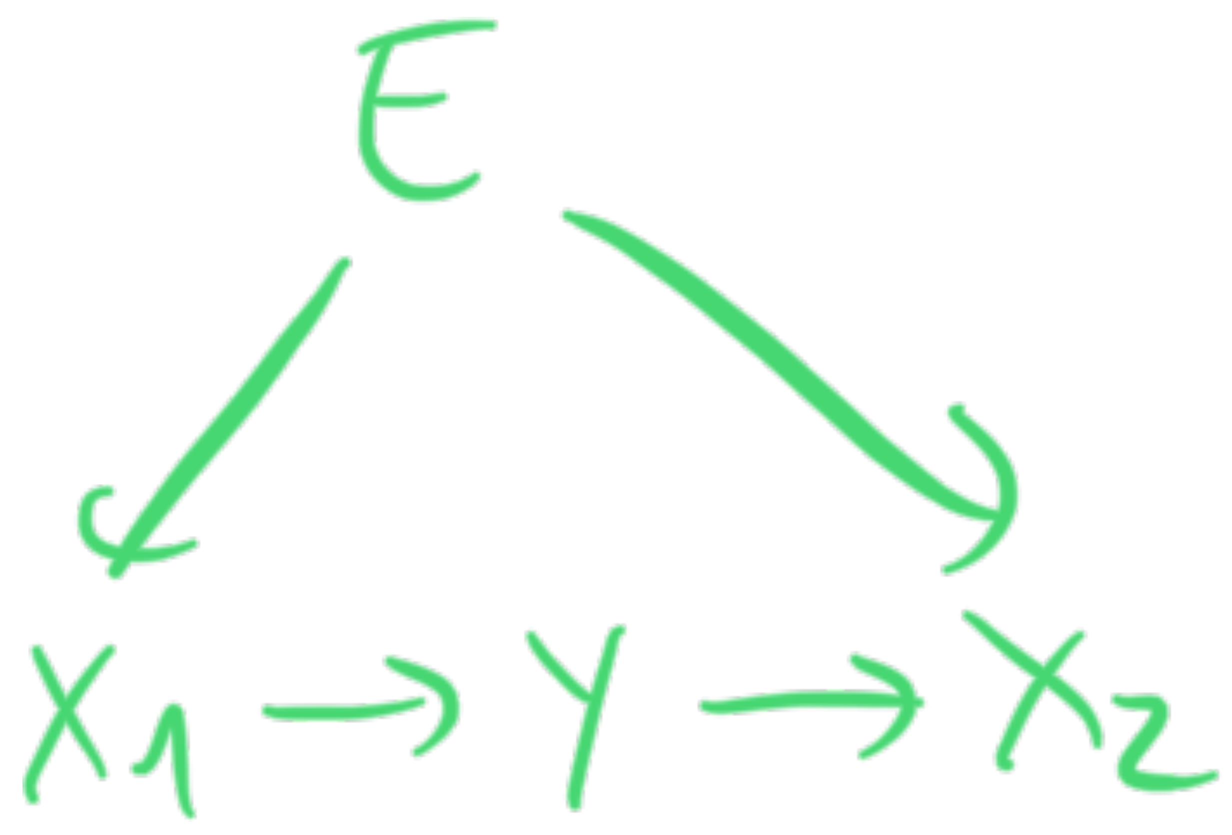
$$P(Y \mid \mathrm{Pa}(Y), E = e_1) = P(Y \mid \mathrm{Pa}(Y), E = e_2)$$

$$Y \perp\!\!\!\perp E \mid \mathrm{Pa}(Y) \qquad Y \perp\!\!\!\perp E \mid X_1$$

Invariant Causal Prediction example 2

$e_1 + e_2$

$$
\begin{cases}
X_1 = \begin{cases} \epsilon_1 & \text{if } E = e_1 \\ 1 & \text{if } E = e_2 \end{cases} \\
Y = X_1 + \epsilon_Y \\
X_2 = \begin{cases} Y + \epsilon_2 & \text{if } E = e_1 \\ 1 & \text{if } E = e_2 \end{cases} \\
\epsilon_1, \epsilon_Y \sim \mathcal{N}(0,1) \\
\epsilon_2 \sim \mathcal{N}(0,0.01)
\end{cases}
$$



$$Y \perp\!\!\!\perp E \mid \mathrm{Pa}(Y)$$

$$\exists \mathbf{S} : Y \perp\!\!\!\perp E \mid \mathbf{S} \implies \mathbf{S} = \mathrm{Pa(Y)}?$$

Invariant Causal Prediction example 3

env e_1

$$\begin{cases} X_1 = \epsilon_1 \\ Y = X_1 + \epsilon_Y \\ X_2 = Y + \epsilon_2 \\ \epsilon_1, \epsilon_Y \sim \mathcal{N}(0,1) \\ \epsilon_2 \sim \mathcal{N}(0,0.01) \end{cases}$$

env e_2

$\rightarrow do(X_1 = 1)$

$X_1 \rightarrow Y \rightarrow X_2$ env e_1

$X_1 \rightarrow Y \rightarrow X_2$ env e_2

$$\begin{tikzcd} & E \arrow[dl] & \\ X_1 \arrow[r] & Y \arrow[r] & X_2 \end{tikzcd}$$

$$\exists \mathbf{S} : Y \perp\!\!\!\perp E \,|\, \mathbf{S} \not\Longrightarrow \mathbf{S} = \mathrm{Pa(Y)}$$

$Y \perp\!\!\!\perp E | X_1, X_2$

In general, being a robust predictor to distribution shifts across some environments E is not enough to be a parent

Invariant Causal Prediction (ICP)

- We assume we have access to **a set of different environments E** (e.g. interventional or observational data), s.t. for $e \in E$, $(X_1^e, \ldots, X_p^e, Y^e) \sim P^e$
- We further assume that in **none of the environments Y is intervened upon**

Invariant Causal Prediction (ICP)

- We assume we have access to **a set of different environments E** (e.g. interventional or observational data), s.t. for $e \in E$, $(X_1^e, \ldots, X_p^e, Y^e) \sim P^e$
- We further assume that in **none of the environments Y is intervened upon**
- We represent the environment index with the random variable E
- If there are no latent confounders, one can prove that:

$$\bigcap_{\mathbf{S} \subseteq \{1,\ldots,p\} \text{ s.t. } Y \perp\!\!\!\perp E | \mathbf{S}} \mathbf{S} \subseteq \mathrm{Pa}(Y)$$

Invariant Causal Prediction example 3

env e_1

$$\begin{cases} X_1 = \epsilon_1 \\ Y = X_1 + \epsilon_Y \\ X_2 = Y + \epsilon_2 \\ \epsilon_1, \epsilon_Y \sim \mathcal{N}(0,1) \\ \epsilon_2 \sim \mathcal{N}(0,0.01) \end{cases}$$

env e_2

$\rightarrow do(X_1 = 1)$

$X_1 \rightarrow Y \rightarrow X_2$ env e_1

$X_1 \rightarrow Y \rightarrow X_2$ env e_2

E

$X_1 \rightarrow Y \rightarrow X_2$

$$\bigcap_{S \subseteq \{1,\ldots,p\} \text{ s.t. } Y \perp\!\!\!\perp E | S} S \subseteq \mathrm{Pa}(Y)$$

$Y \perp\!\!\!\perp E | X_1$

$Y \not\perp\!\!\!\perp E | X_1, X_2$

$\{X_1\} \cap \{X_1, X_2\} = \{X_1\}$

Invariant Causal Prediction example 4

env e_1

$$\begin{cases} X_1 = \epsilon_1 \\ Y = X_1 + \epsilon_Y \\ X_2 = Y + \epsilon_2 \\ \epsilon_1, \epsilon_Y \sim \mathcal{N}(0,1) \\ \epsilon_2 \sim \mathcal{N}(0,0.01) \end{cases}$$

env e_2

$\rightarrow do(X_1 = 1)$

$X_1 \rightarrow Y \rightarrow X_2$ env e_1

$X_1 \rightarrow Y \rightarrow X_2$ env e_2

E

$X_1 \rightarrow Y \rightarrow X_2$

$pa(X_2) = ?$

$$\bigcap_{S \subseteq \{1,\ldots,p\} \text{ s.t. } X_2 \perp\!\!\!\perp E | S} S \subseteq \mathrm{Pa}(X_2)$$

$X_2 \perp\!\!\!\perp E | X_1$

$X_2 \perp\!\!\!\perp E | X_1, Y$

$X_2 \perp\!\!\!\perp E | Y$

$\{X_1\} \cap \{X_1, Y\} \cap \{Y\} = \emptyset$

$\emptyset \subseteq Pa(X_2)$

Invariant Causal Prediction (ICP)

- We assume we have access to **a set of different environments E** (e.g. interventional or observational data), s.t. for $e \in E$, $(X_1^e, \ldots, X_p^e, Y^e) \sim P^e$
- We further assume that in **none of the environments Y is intervened upon**
- We represent the environment index with the random variable E
- If there are no latent confounders, one can prove that:

$$\bigcap_{\mathbf{S} \subseteq \{1,\ldots,p\} \text{ s.t. } Y \perp\!\!\!\perp E | \mathbf{S}} \mathbf{S} \subseteq \mathrm{Pa}(Y)$$

ICP finds subsets of parents (not necessarily all, it also depends on the available environments)

ICP improves with more interventions

$e_1 + e_2$

$E \to X_1 \to X_2 \to Y$

$$\left.\begin{array}{l} E \perp\!\!\!\perp Y \mid X_1 \\ E \not\perp\!\!\!\perp Y \mid X_2 \\ E \perp\!\!\!\perp Y \mid X_1, X_2 \end{array}\right\} \cap \quad \phi$$

$$\bigcap_{\mathbf{S} \subseteq \{1,\ldots,p\} \text{ s.t. } Y \perp\!\!\!\perp E \mid \mathbf{S}} \mathbf{S} = \varnothing$$

+ new enviroment e3

$X_2 = g^{new}(X_1, \varepsilon_2)$

$E \to X_1 \to X_2 \to Y$, $E \to X_2$

$$\left.\begin{array}{l} E \perp\!\!\!\perp Y \mid X_2 \\ E \perp\!\!\!\perp Y \mid X_2, X_1 \end{array}\right. \cap \quad X_2 \in Pa(Y)$$

$$\bigcap_{\mathbf{S} \subseteq \{1,\ldots,p\} \text{ s.t. } Y \perp\!\!\!\perp E \mid \mathbf{S}} \mathbf{S} = X_2 \in \mathrm{Pa}(Y)$$

ICP improves with more interventions

- We assume we have access to **a set of different environments E** (e.g. interventional or observational data), s.t. for $e \in E$, $(X_1^e, \ldots, X_p^e, Y^e) \sim P^e$
- We further assume that in **none of the environments Y is intervened upon**
- We represent the environment index with E, no latent confounders
- If the **environment variable causes all** $(X_1, \ldots, X_p)$

$$\mathrm{Pa}(Y) = \bigcap_{\mathbf{S} \subseteq \{1,\ldots,p\} \text{ s.t. } Y \perp\!\!\!\perp E | \mathbf{S}} \mathbf{S}$$

I.e. we have "enough" different experiments

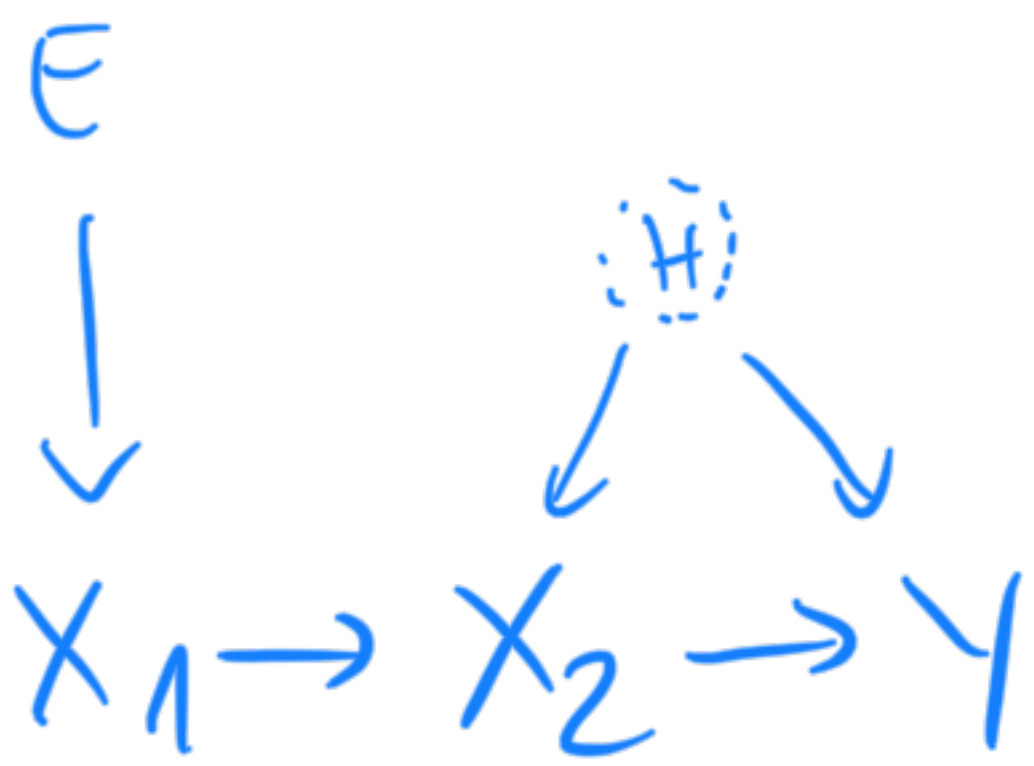
Invariant Causal Prediction in practice

- We assume we have access to **a set of different environments E** (e.g. interventional or observational data), s.t. for $e \in E$, $(X_1^e, \ldots, X_p^e, Y^e) \sim P^e$
- We further assume that in **none of the environments Y is intervened upon**
- Approximate test on residuals for each $\mathbf{S} \subseteq \{1, \ldots, p\}$
  - Fit linear regression with $\mathbf{S}$ and let $R = Y - \hat{f}(X_{\mathbf{S}})$
  - Test null hypothesis that mean and variance of R are the same across E
  - Combine the two p-values and reject $\mathbf{S}$ if the combined p-value $\leq \alpha$

Invariant Causal Prediction - latent confounders

- If there are latent confounders, one can prove that:

$$\bigcap_{\mathbf{S}\subseteq\{1,\dots,p\} \text{ s.t. } Y \perp\!\!\!\perp E | \mathbf{S}} \mathbf{S} \subseteq \mathrm{Anc}(Y)$$



$Y \not\!\perp\!\!\!\perp E \mid X_2$

We cannot find the parent X_2

$Y \perp\!\!\!\perp E \mid X_1 \Rightarrow X_1 \in \mathrm{Anc}(Y)$

$Y \perp\!\!\!\perp E \mid X_1, X_2$

We now find the ancestor X_1

Learning from multiple contexts

- In **intervention design** we could decide which intervention to perform
  - We usually also had **known intervention targets**, e.g. $\mathrm{do}(S = 1)$
- What if we cannot decide interventions, but instead somebody gives us a **set of data from multiple contexts?**
  - Possibly **with unknown intervention targets**
  - Possibly **soft interventions** instead of **perfect interventions**

ICP finds subsets of parents, what about finding (an equivalence class of) the causal graph?

Joint Causal Inference from Multiple Contexts

Joris M. Mooij, Sara Magliacane, Tom Claassen

- We represent different distributions (including interventional) as an **unknown** **joint causal graph** *(possibly cyclic or with latent confounders)*

- We **add context variables** so we can **disentangle** changes in distribution across the datasets

https://arxiv.org/abs/1611.10351

Joint Causal Inference intuition

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

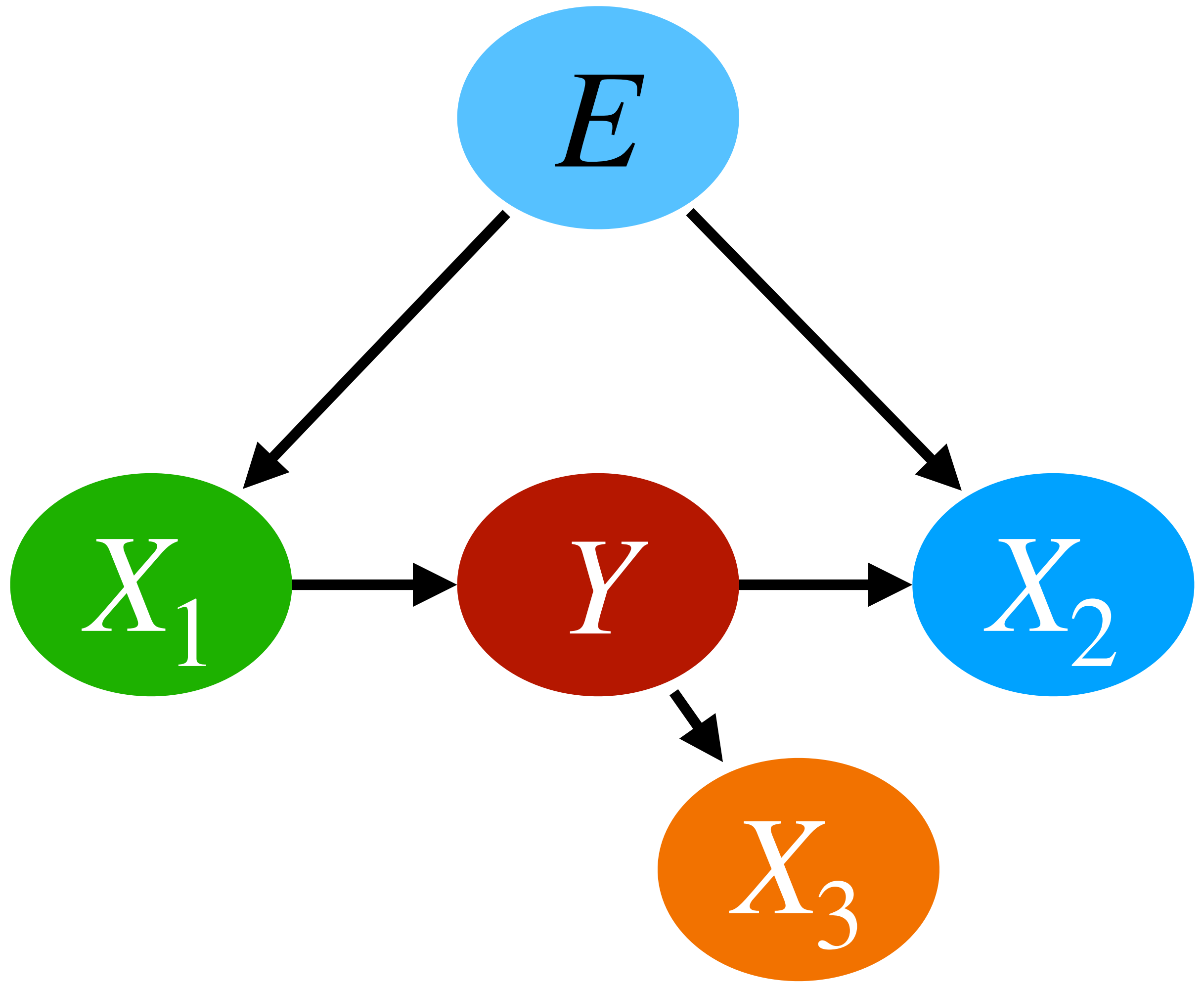
$E = 0$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 100 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$E = 1$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 10Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$E = 2$



Joint Causal Inference intuition

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$E = 0$

$C_1 = 0$
$C_2 = 0$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 100 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

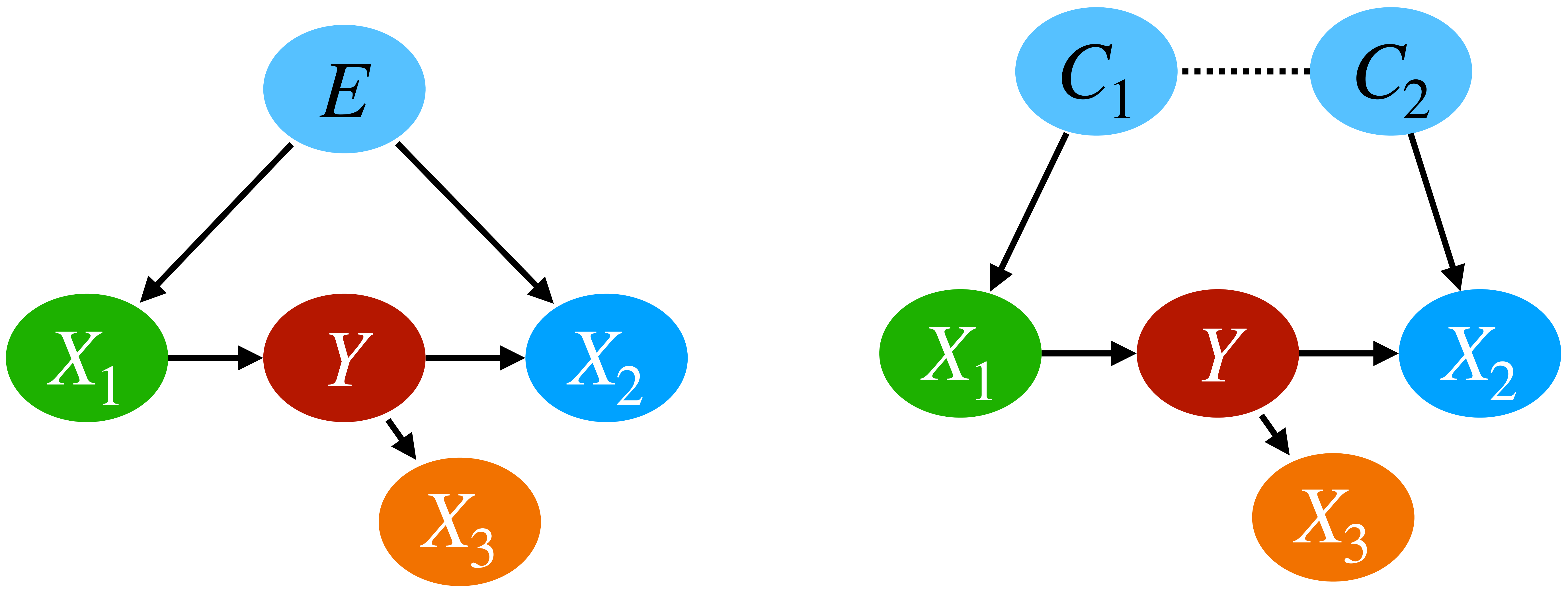
$E = 1$

$C_1 = 1$
$C_2 = 0$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 10Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$E = 2$

$C_1 = 0$
$C_2 = 1$



Adding context variables C_1 and C_2 helps disentangle the changes in each environment

Joint Causal Inference intuition

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$E = 0$

$C_1 = 0$
$C_2 = 0$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 100 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$E = 1$

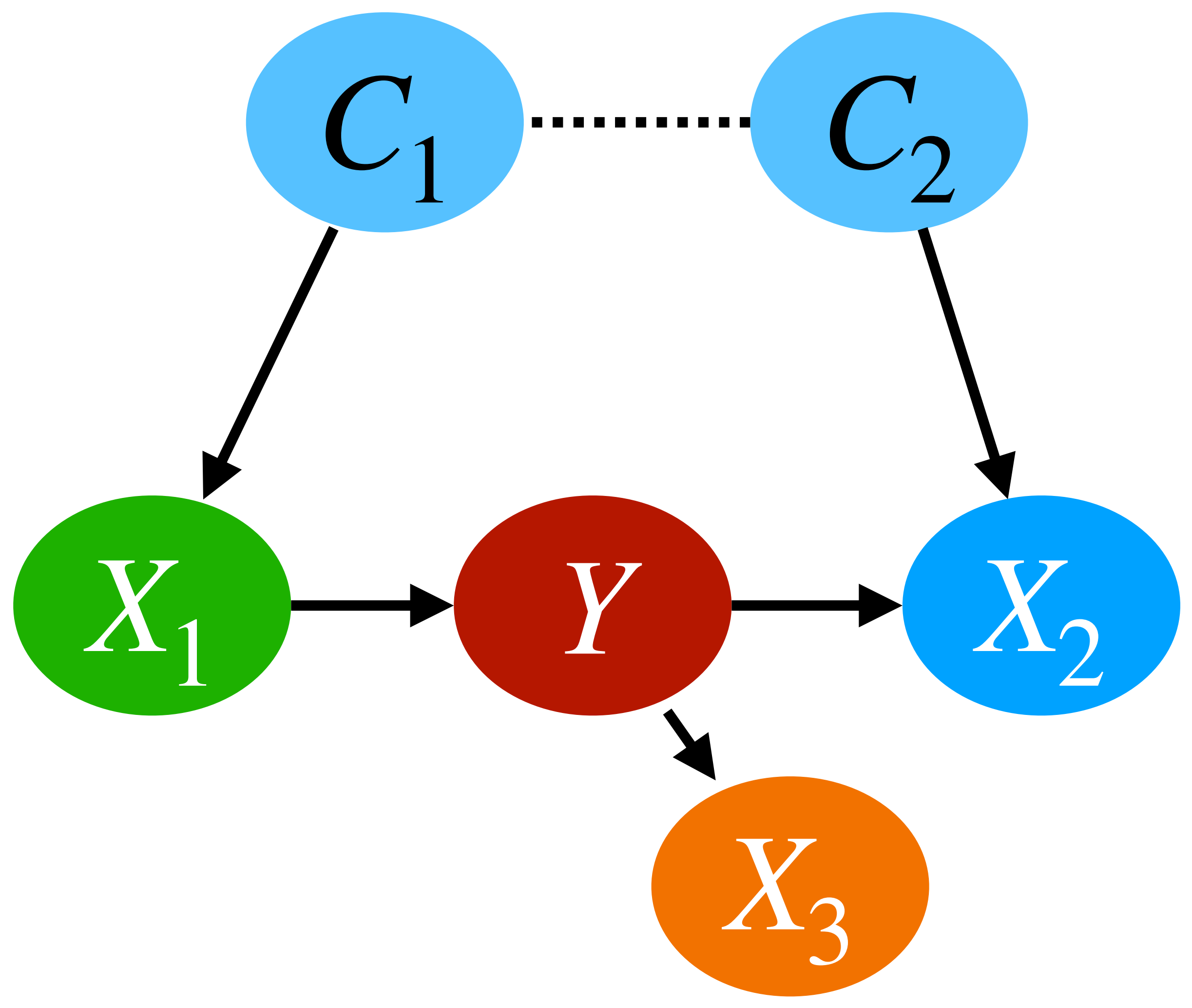
$C_1 = 1$
$C_2 = 0$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 10Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$E = 2$

$C_1 = 0$
$C_2 = 1$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = \begin{cases} 10 + \epsilon_1 \text{ if } C_1 = 0 \\ 100 + \epsilon_1 \text{ if } C_1 = 1 \end{cases} \\ Y = 3X_1 + \epsilon_Y \\ X_2 = \begin{cases} -2Y + \epsilon_2 \text{ if } C_2 = 0 \\ 10Y + \epsilon_2 \text{ if } C_2 = 1 \end{cases} \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$



Joint Causal Inference intuition

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$E = 0$

$C_1 = 10$
$C_2 = -2$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 100 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = -2Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

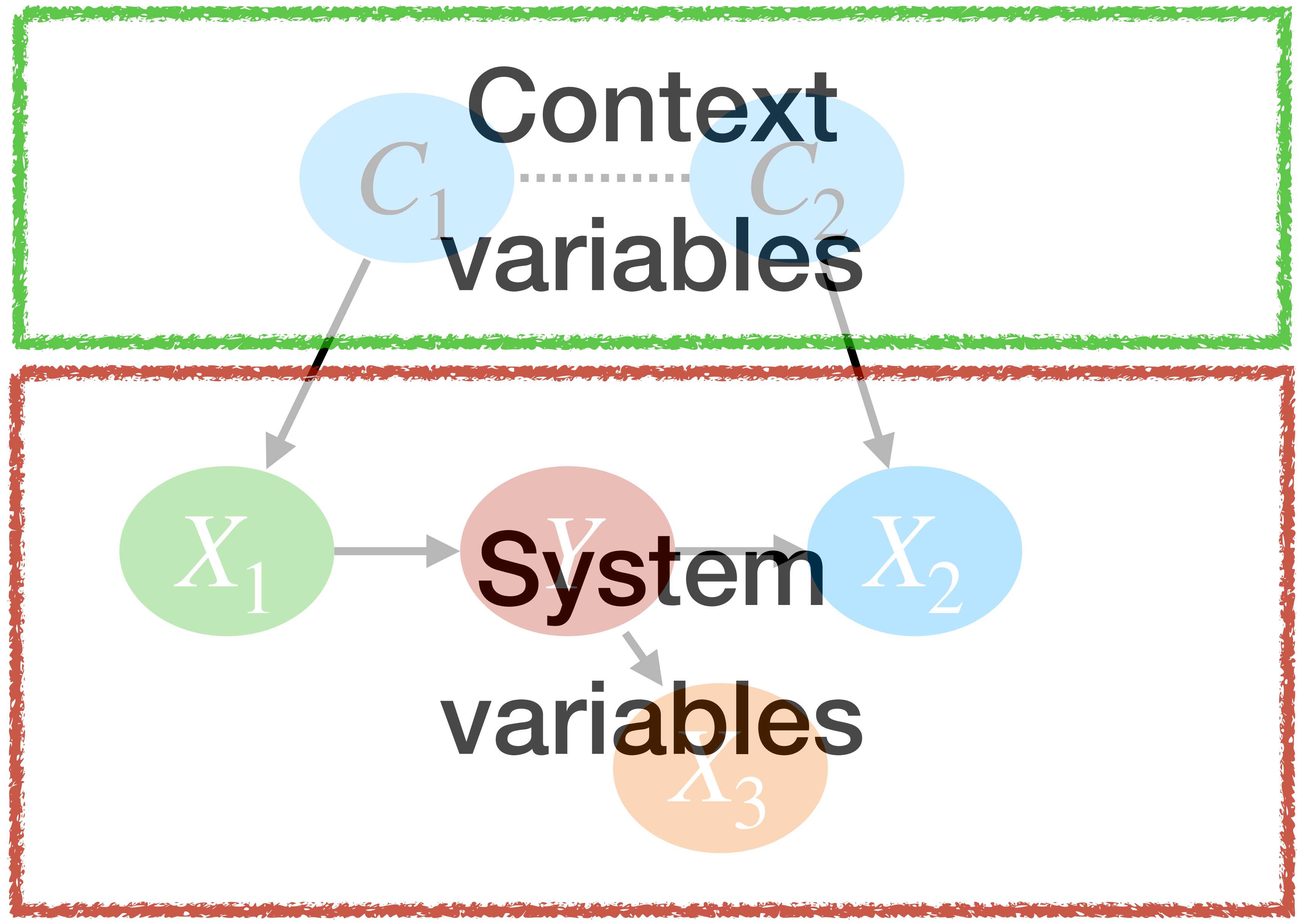
$E = 1$

$C_1 = 100$
$C_2 = -2$

$$\begin{cases} \epsilon_1, \epsilon_2, \epsilon_3, \epsilon_Y \sim \mathcal{N}(0,1) \\ X_1 = 10 + \epsilon_1 \\ Y = 3X_1 + \epsilon_Y \\ X_2 = 10Y + \epsilon_2 \\ X_3 = 2Y + 0.1\epsilon_3 \end{cases}$$

$E = 2$

$C_1 = 10$
$C_2 = 10$



The context variables C_1 and C_2 can be also descriptive of the intervention in each environment

Joint Causal Inference from Multiple Contexts

Joris M. Mooij, Sara Magliacane, Tom Claassen

- We represent different distributions (including interventional) as an **unknown** **joint causal graph** (possibly cyclic or with latent confounders)

- We **add context variables** so we can **disentangle** changes in distribution across the datasets

	X1	X2	X3
Normal	0,1	2	0
Normal	0,2	3	0

	X1	X2	X3
Gene A	3,1	2	1
Gene A	3,2	3	1
			1
			1

	X1	X2	X3
Gene B	0.2	1	0
Gene B	0.3	1	1
Gene B	0.3	2	1
Gene B	0.4	1	1

C1	C2	X1	X2	X3
0	0	0,1	2	0
0	0	0,2	3	0
0	0	1,1	2	1
0	0	0,1	3	0
1	0	3,1	2	1
1	0	3,2	3	1
1	0	4	1	1
1	0	3,2	3	1
0	1	0,2	1	0
0	1	0,3	1	1
0	1	0,3	2	1
0	1	0,4	1	1

Joint Causal Inference from Multiple Contexts

Joris M. Mooij, Sara Magliacane, Tom Claassen

- We **add context variables** so we can **disentangle** changes in distribution across the datasets (and optionally background knowledge, e.g. context variables are uncaused)
- We can reuse **any standard method for observational data** that fits any chosen assumptions

C1	C2	X1	X2	X3
0	0	0,1	2	0
0	0	0,2	3	0
0	0	1,1	2	1
0	0	0,1	3	0
1	0	3,1	2	1
1	0	3,2	3	1
1	0	4	1	1
1	0	3,2	3	1
0	1	0,2	1	0
0	1	0,3	1	1
0	1	0,3	2	1
0	1	0,4	1	1

Joint Causal Inference from Multiple Contexts

Joris M. Mooij, Sara Magliacane, Tom Claassen

- We **add context variables** so we can **disentangle** changes in distribution across the datasets (and optionally background knowledge, e.g. context variables are uncaused)

- We can reuse **any standard method for observational data** that fits any chosen assumptions

C1	C2	X1	X2	X3
0	0	0,1	2	0
0	0	0,2	3	0
0	0	1,1	2	1
0	0	0,1	3	0
1	0	3,1	2	1
1	0	3,2	3	1
1	0	4	1	1
1	0	3,2	3	1
0	1	0,2	1	0
0	1	0,3	1	1
0	1	0,3	2	1
0	1	0,4	1	1

$X_2 \not\perp\!\!\!\perp C_2$

$X_1 \perp\!\!\!\perp C_2 \mid C_1$

$X_2 \perp\!\!\!\perp C_1 \mid X_3$

$\dots$

Joint Causal Inference from Multiple Contexts

Joris M. Mooij, Sara Magliacane, Tom Claassen

- We **add context variables** so we can **disentangle** changes in distribution across the datasets (and optionally background knowledge, e.g. context variables are uncaused)

- We can reuse **any standard method for observational data** that fits any chosen assumptions

C1	C2	X1	X2	X3
0	0	0,1	2	0
0	0	0,2	3	0
0	0	1,1	2	1
0	0	0,1	3	0
1	0	3,1	2	1
1	0	3,2	3	1
1	0	4	1	1
1	0	3,2	3	1
0	1	0,2	1	0
0	1	0,3	1	1
0	1	0,3	2	1
0	1	0,4	1	1

$X_2 \not\perp\!\!\!\perp C_2$

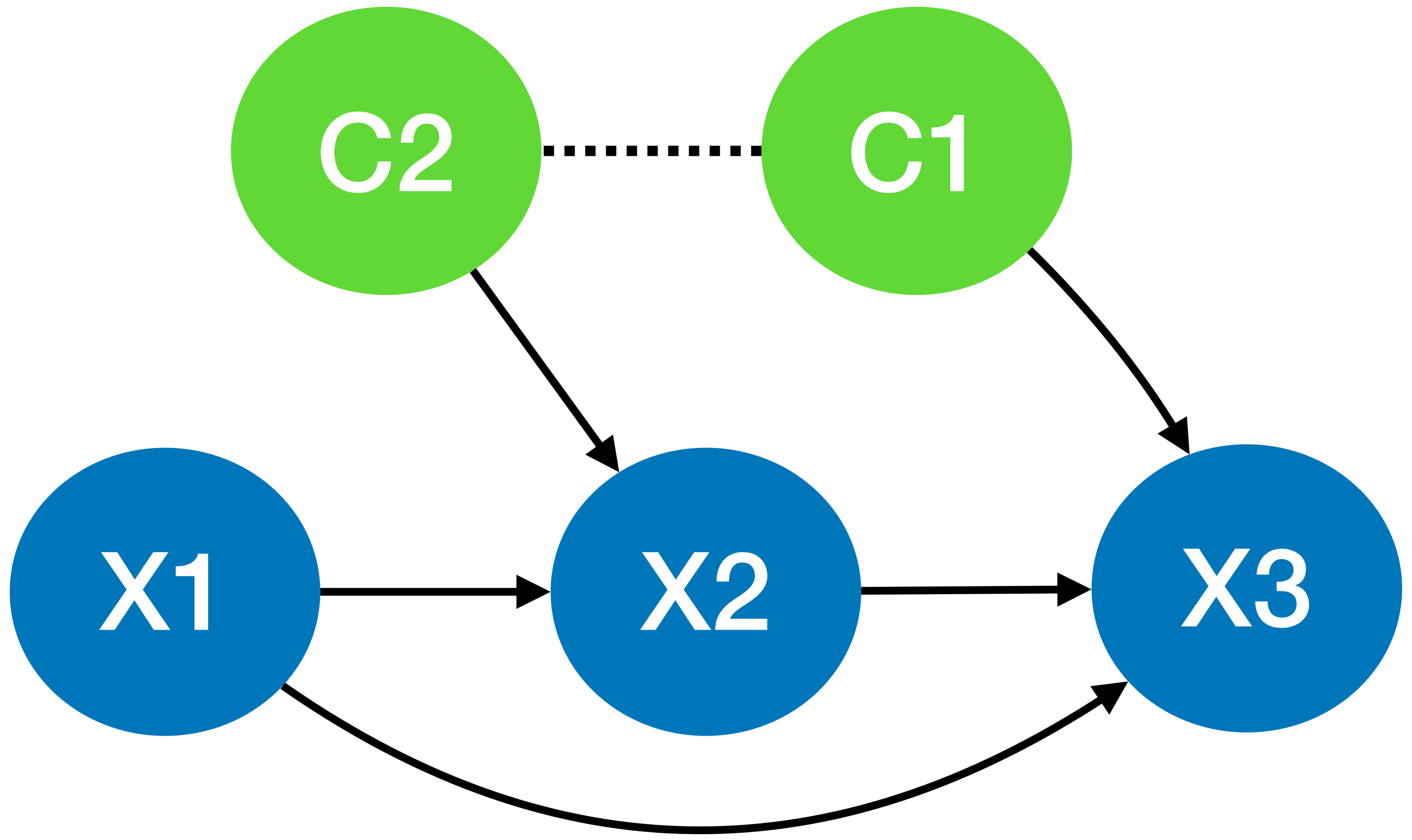
$X_1 \perp\!\!\!\perp C_2 \mid C_1$

$X_2 \perp\!\!\!\perp C_1 \mid X_3$

$\dots$

PC-JCI

FCI-JCI

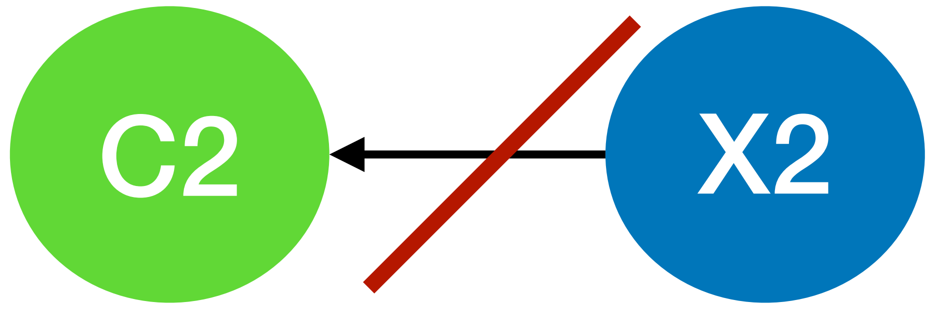


Joint Causal Inference from Multiple Contexts

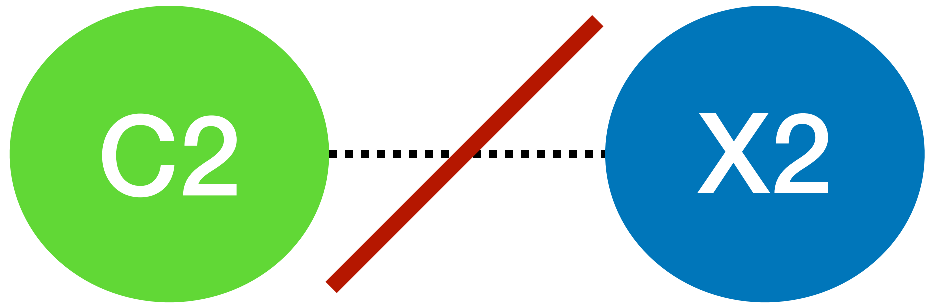
Joris M. Mooij, Sara Magliacane, Tom Claassen

- Additional optional background knowledge based on assumptions:

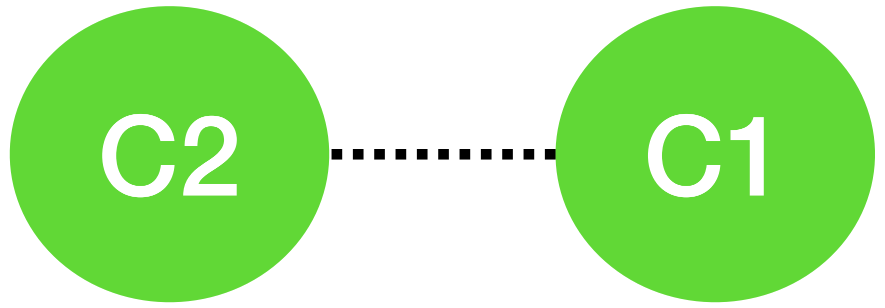
  1. No system variable causes any context variable.



  2. No context variable is confounded with a system variable.



  3. The context variables do not cause each other and they are assumed to be confounded.



https://arxiv.org/abs/1611.10351

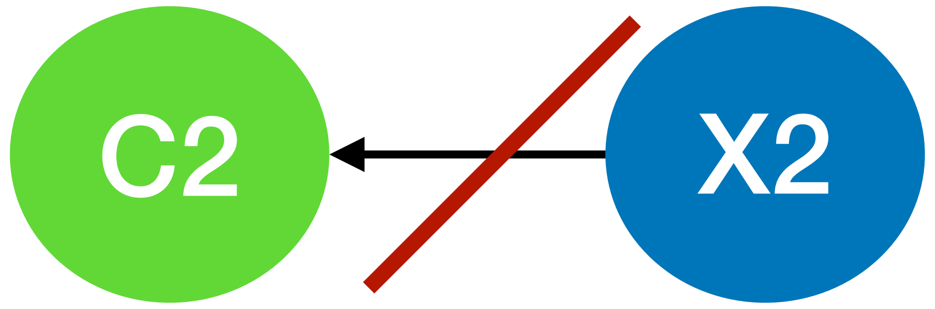
Joint Causal Inference from Multiple Contexts

Joris M. Mooij, Sara Magliacane, Tom Claassen

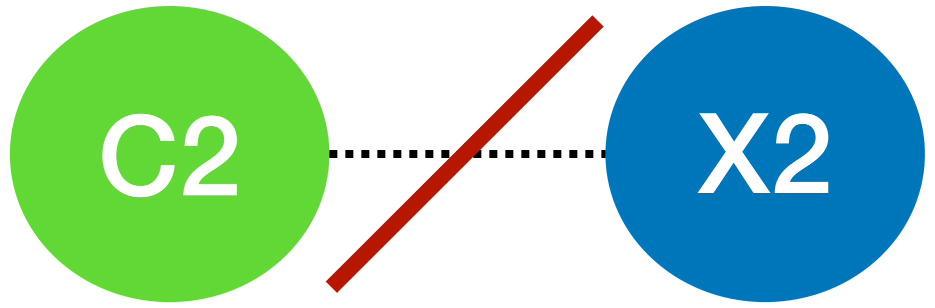
In our course, we assume that there are no latent confounders (except some dependence between the context variables)

- Additional (mptions:

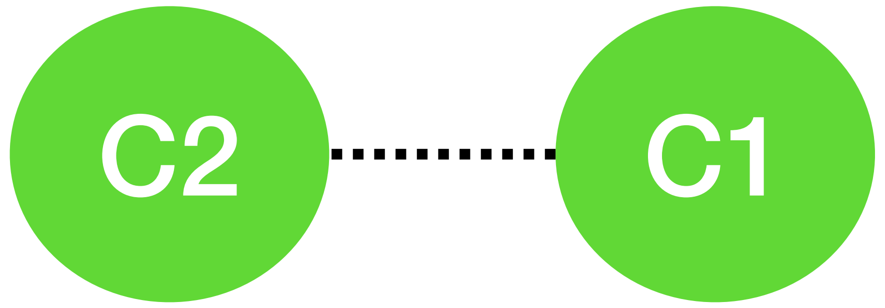
1. No system variable causes any context variable.



2. ~~No context variable is confounded with a system variable.~~

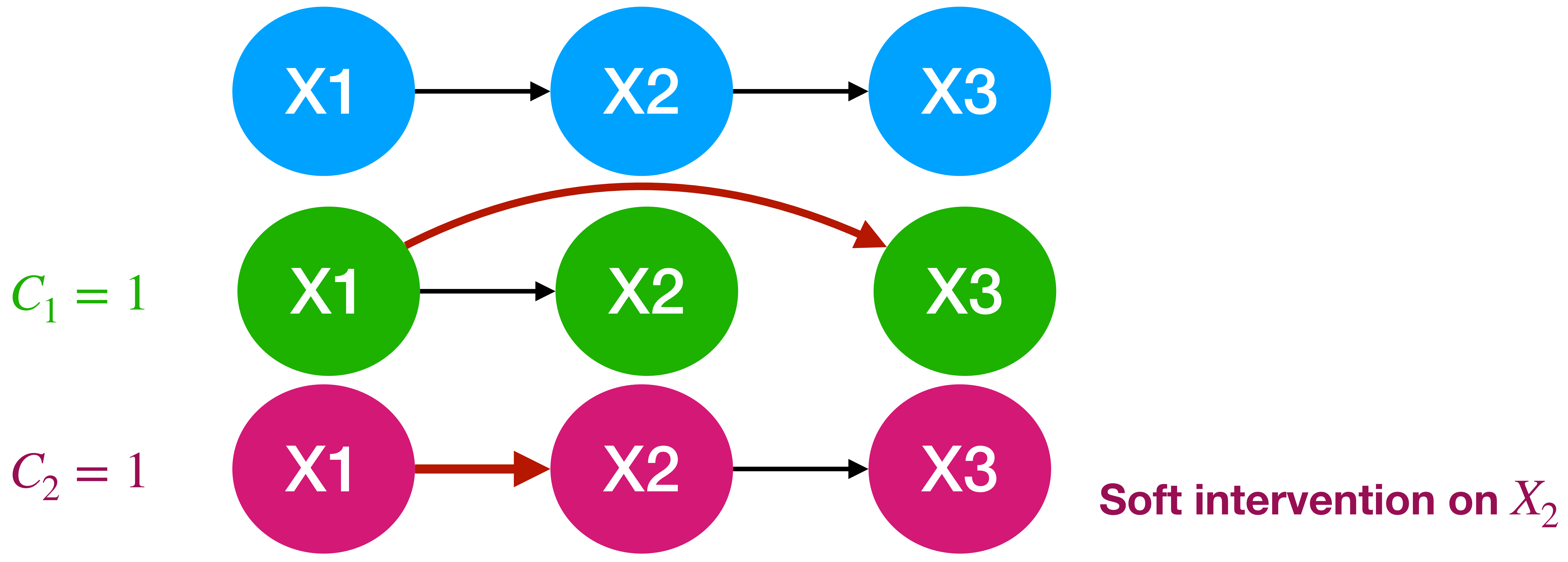


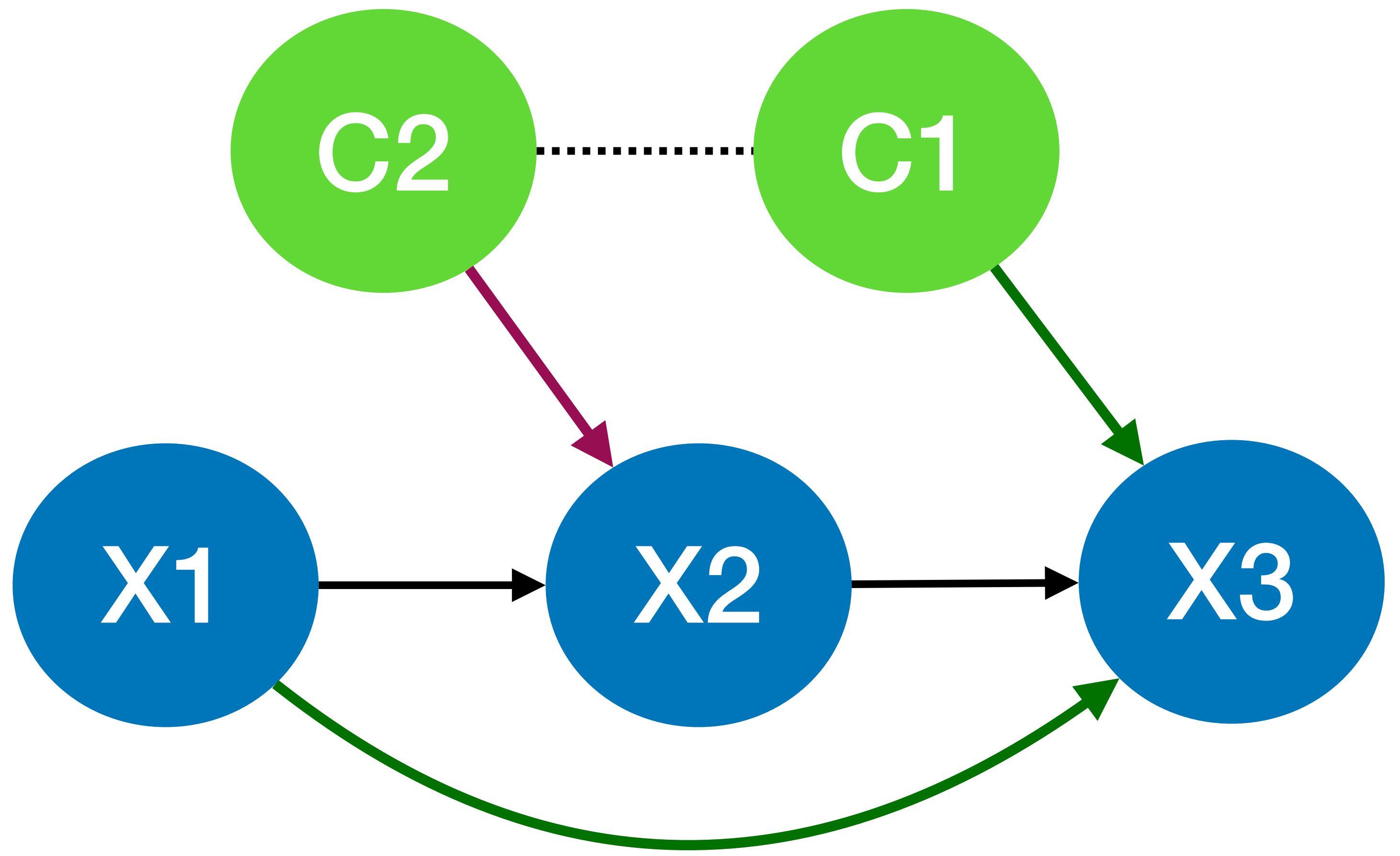
3. The context variables do not cause each other and they are assumed to be confounded.



https://arxiv.org/abs/1611.10351

Joint Causal Inference example - setting

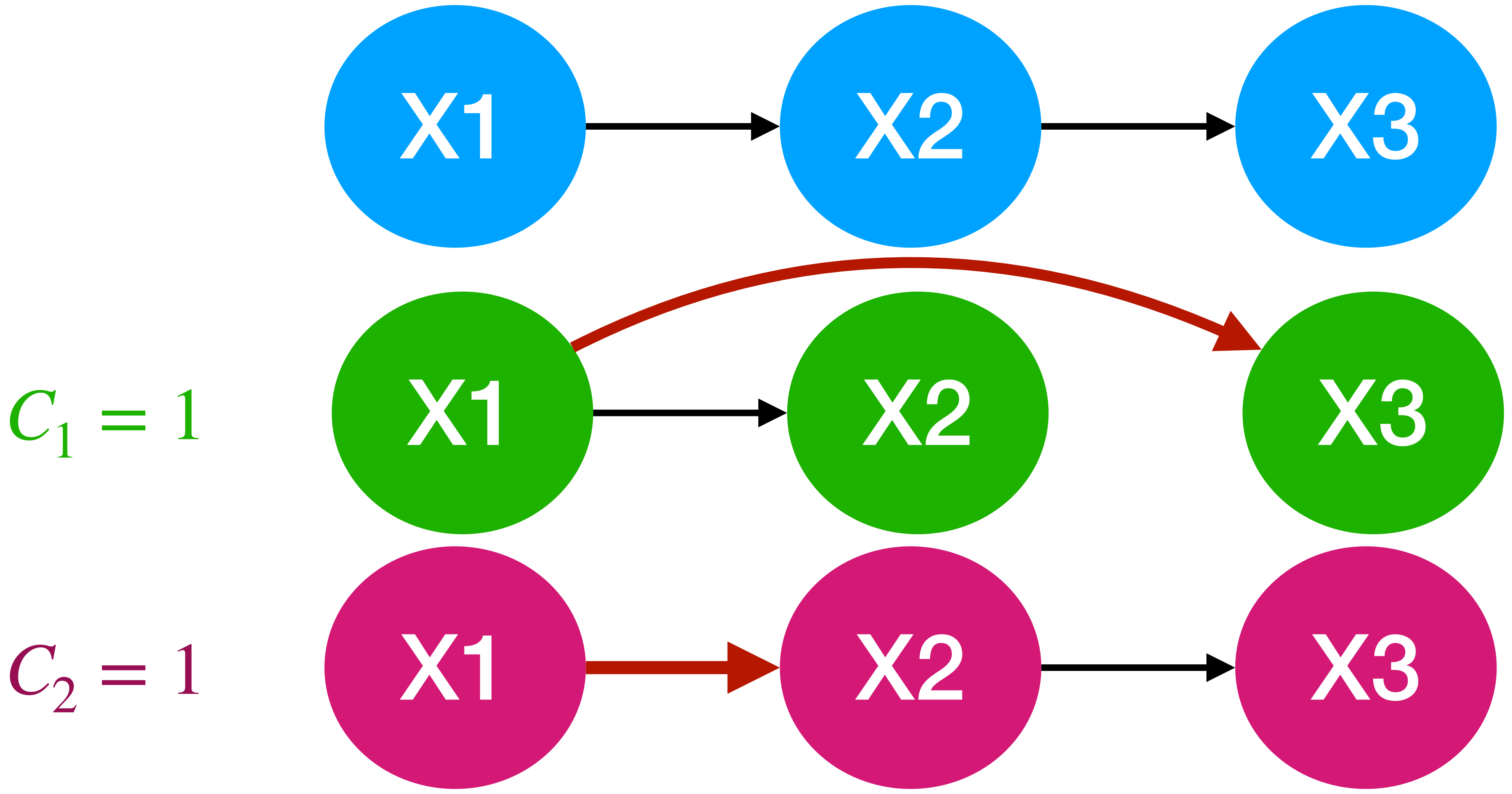




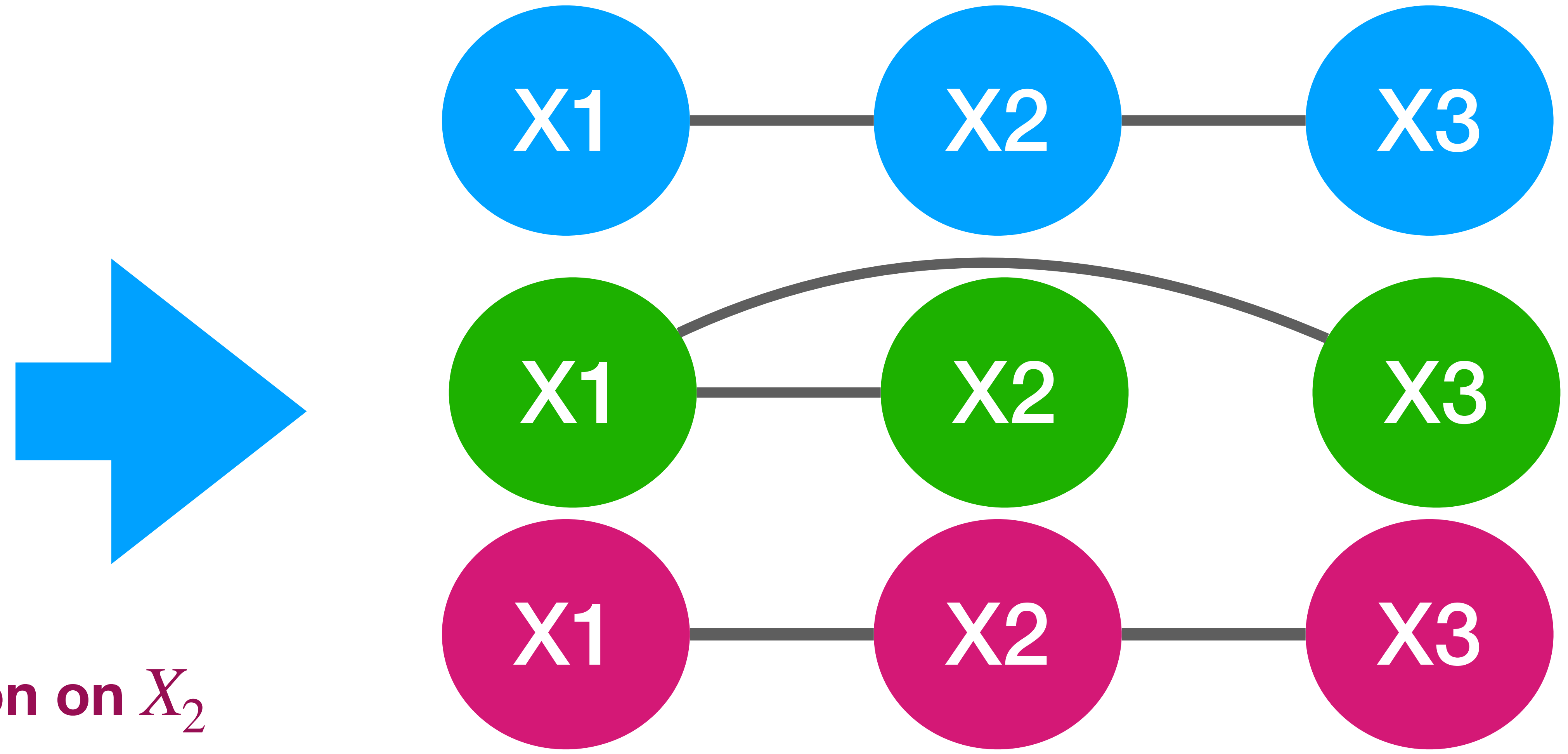
Single graphs in each environment

The joint graph is the union of the single graphs + edges from context variables for the causal mechanisms that change

Learning graphs separately in each environment with PC

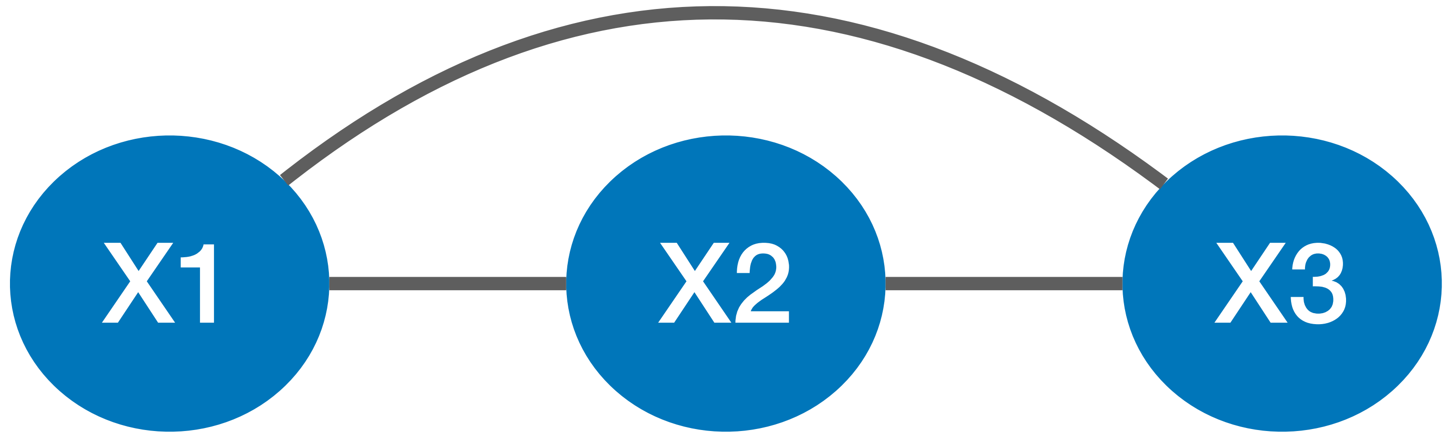


Soft intervention on X_2

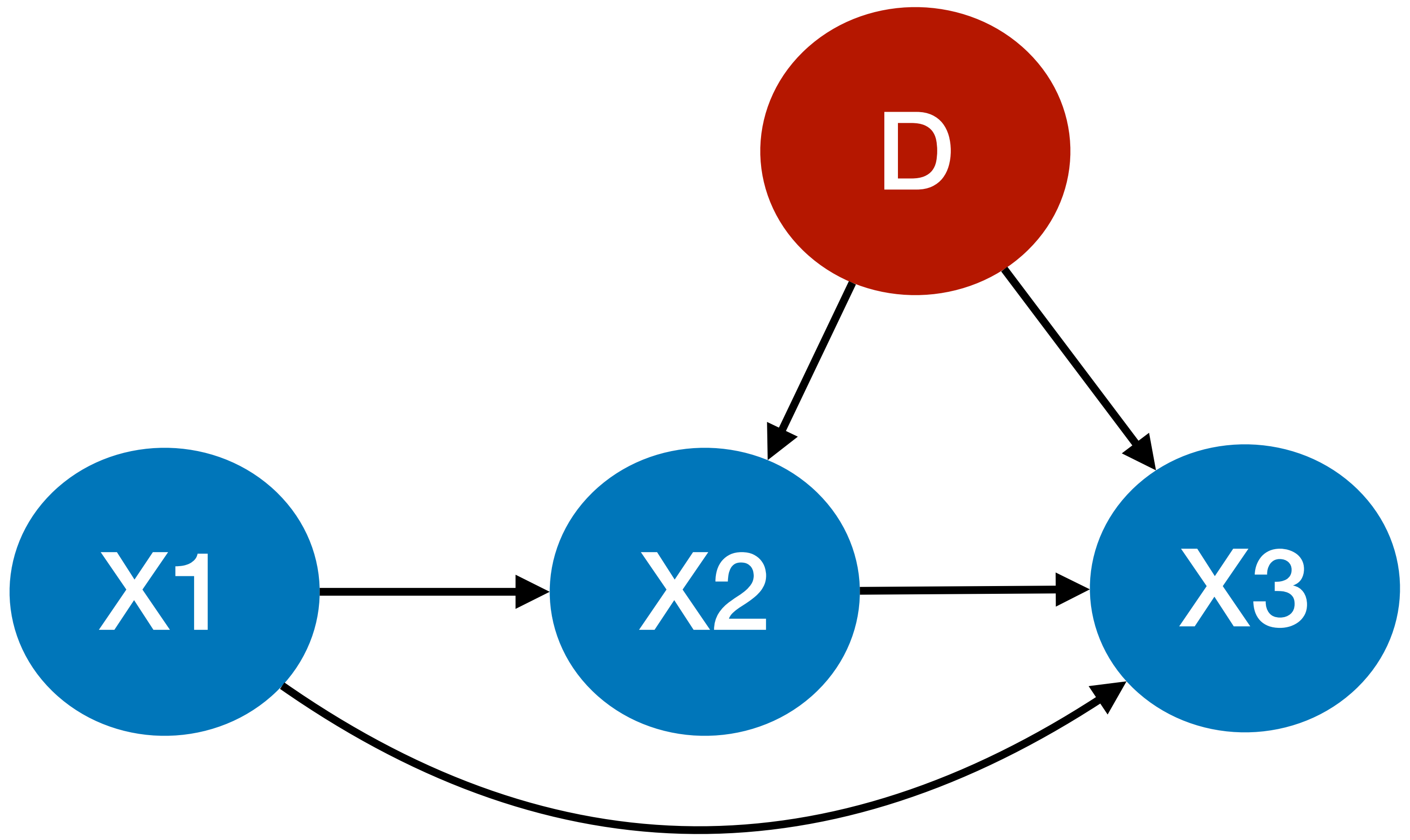


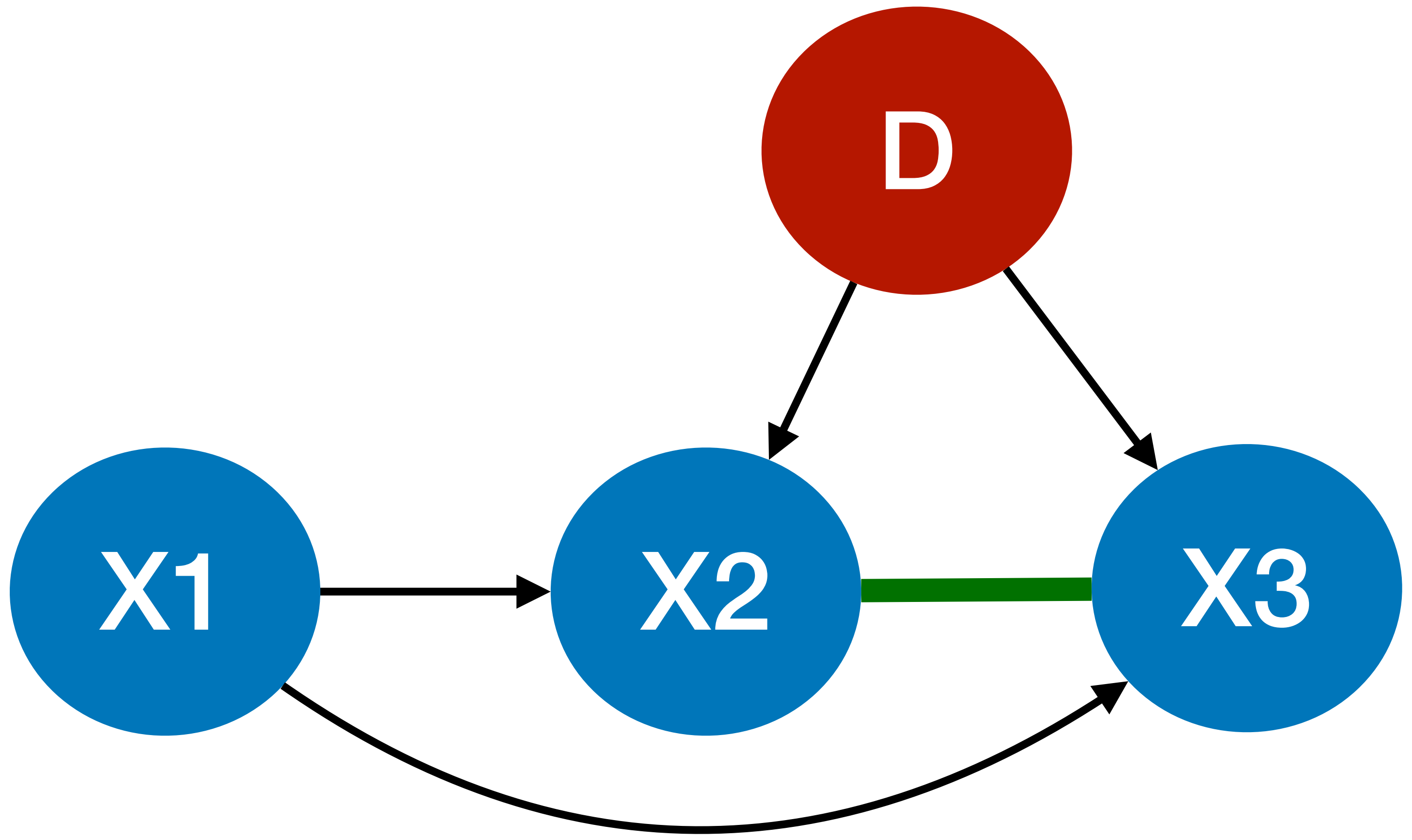
Single graphs in each environment

CPDAGs from each environment



Joint causal inference + PC with a single domain variable

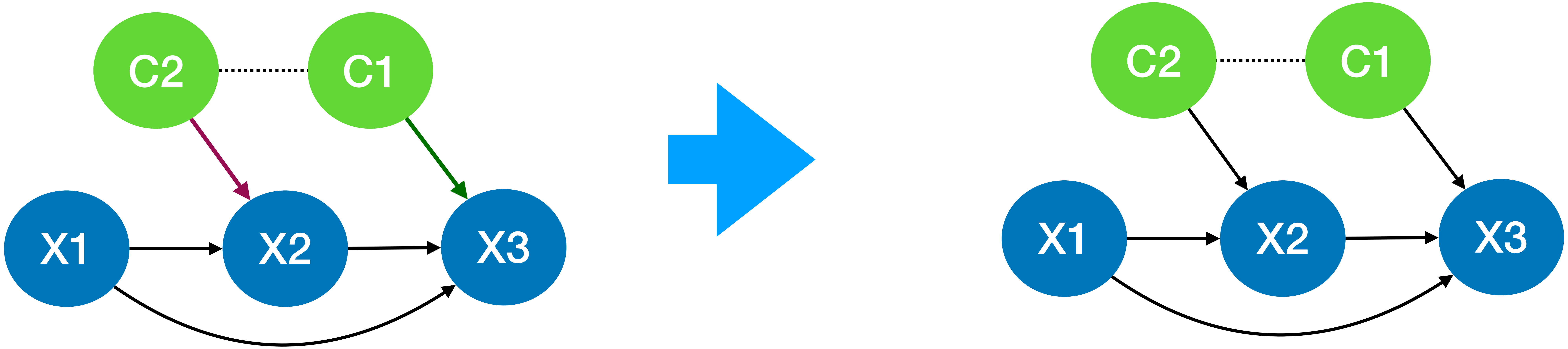




True underlying joint graph with domain variable

CPDAG with Joint Causal Inference and PC with a single domain variable

Joint causal inference + PC with multiple context variables



True underlying joint graph with multiple context variable

CPDAG with Joint Causal Inference and PC with multiple context variables