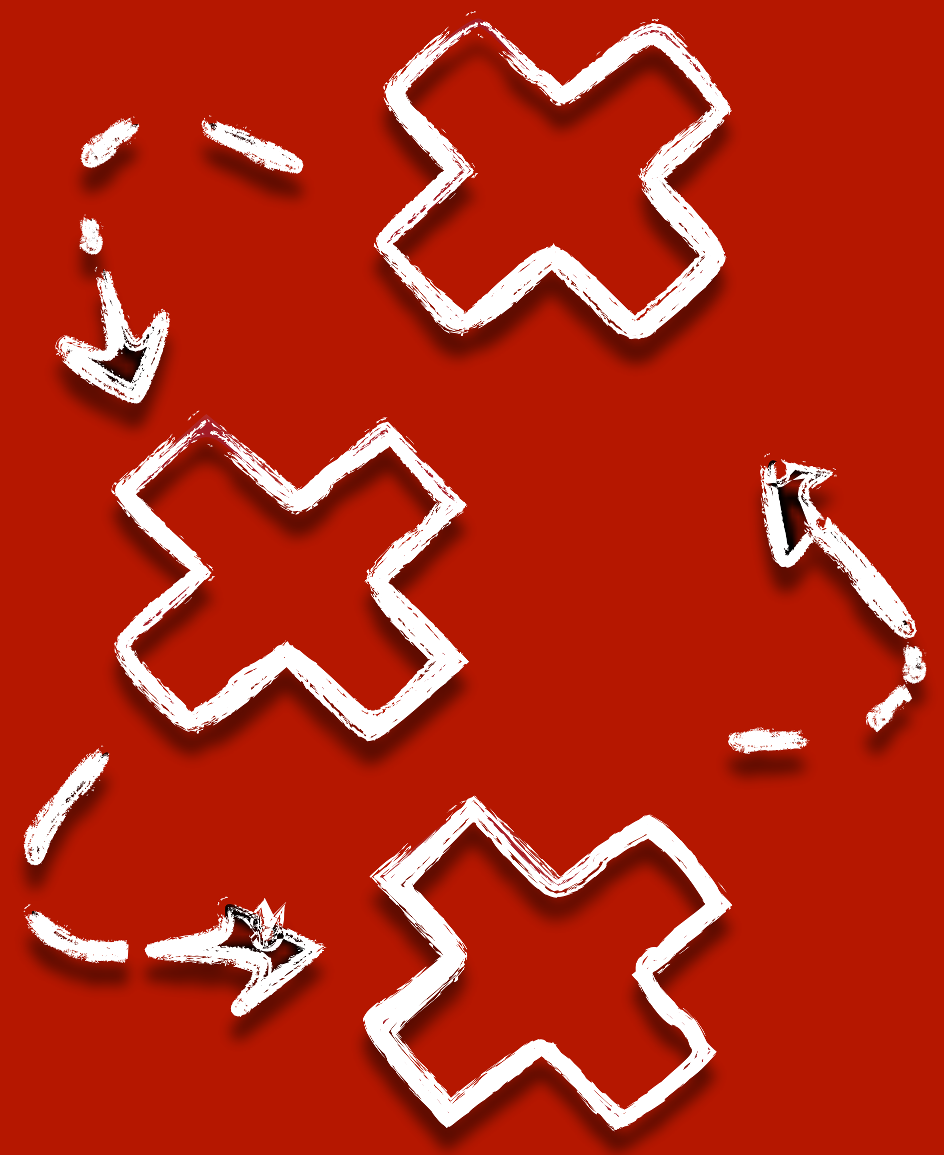


Causal Data Science

Lecture 4:1 Causal graphs

Lecturer: Sara Magliacane

UvA - Spring 2024

Last lecture: Why should we care about Bayesian networks?

- We have a set of random variables $X_1, \ldots, X_p$ with joint $p(X_1, \ldots, X_p)$
- We have a DAG G, s.t. **each random variable** X_i is represented by **node** i

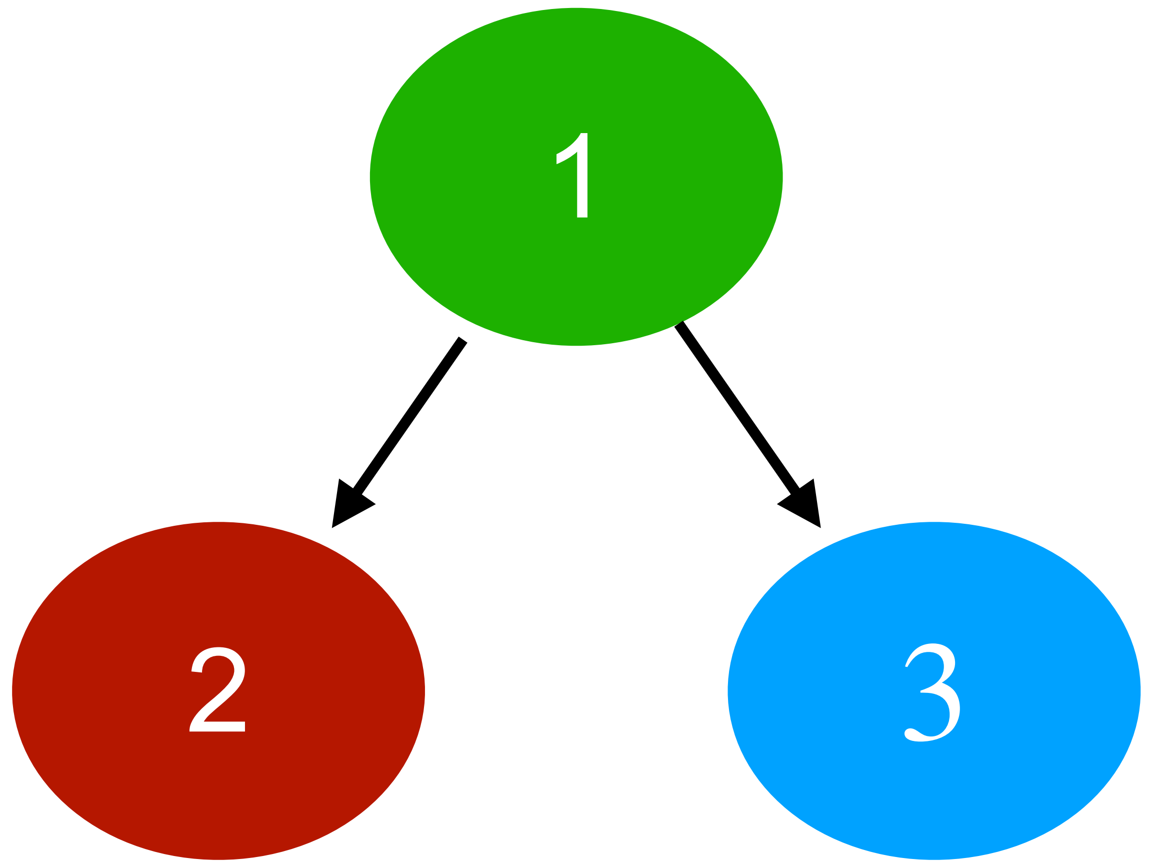
- We then say $p(X_1, \ldots, X_p)$ **factorizes over G** if

$$p(X_1, \ldots, X_p) = \prod_{i \in V} p(X_i \,|\, \mathbf{X}_{\mathrm{pa}(i)})$$

They can help simplify the factorisation

We can easily read conditional independences

They can represent causal models



Last lecture: Why should we care about Bayesian networks?

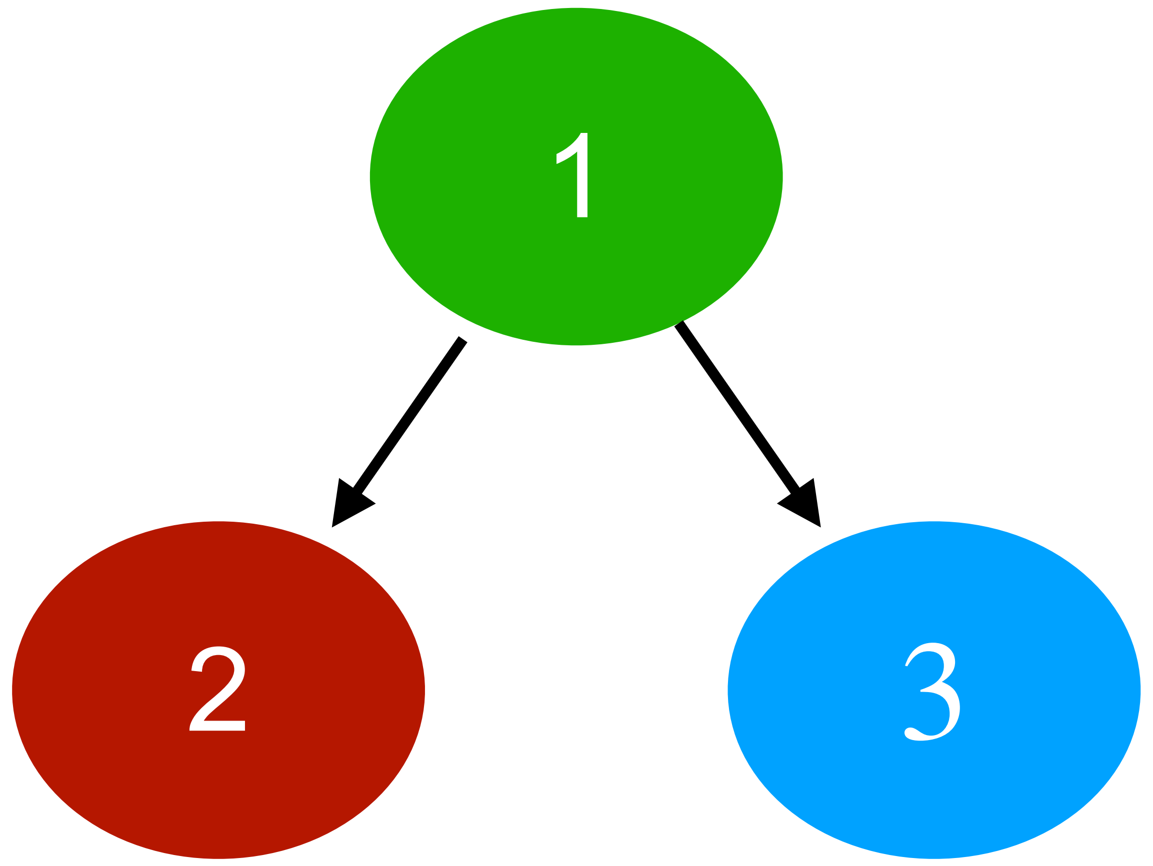
- We have a set of random variables $X_1, \ldots, X_p$ with joint $p(X_1, \ldots, X_p)$
- We have a DAG G, s.t. **each random variable** X_i is represented by **node** i
- We then say $p(X_1, \ldots, X_p)$ **factorizes over G** if

$$p(X_1, \ldots, X_p) = \prod_{i \in V} p(X_i \mid \mathbf{X}_{\mathrm{pa}(i)})$$

They can help simplify the factorisation

We can easily read conditional independences

They can represent causal models



This lecture: Why should we care about Bayesian networks?

- We have a set of random variables $X_1, \ldots, X_p$ with joint $p(X_1, \ldots, X_p)$
- We have a DAG G, s.t. **each random variable** X_i is represented by **node** i

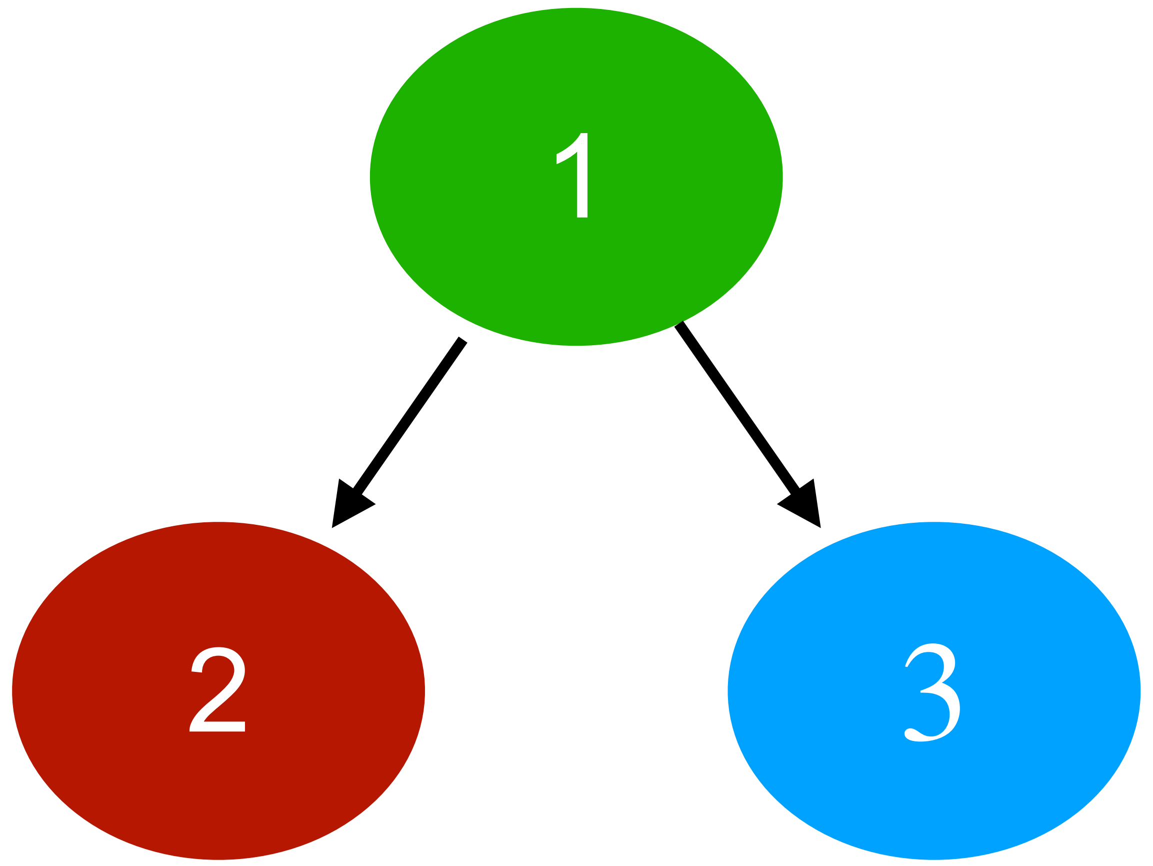
- We then say $p(X_1, \ldots, X_p)$ **factorizes over G** if

$$p(X_1, \ldots, X_p) = \prod_{i \in V} p(X_i \mid \mathbf{X}_{\mathrm{pa}(i)})$$

They can help simplify the factorisation

We can easily read conditional independences

They can represent causal models



BNs vs causal BNs - example 1

- Fire (F) and Alarm (A) with $p(F, A)$ and $A \not\perp\!\!\!\perp F$ can be factorized as:

BNs vs causal BNs - example 1

- Fire (F) and Alarm (A) with $p(F, A)$ and $A \not\perp\!\!\!\perp F$ can be factorized as:

$$p(F, A) = p(F)\, p(A \mid F)$$

$$p(F, A) = p(A)\, p(F \mid A)$$

$$F \longrightarrow A$$

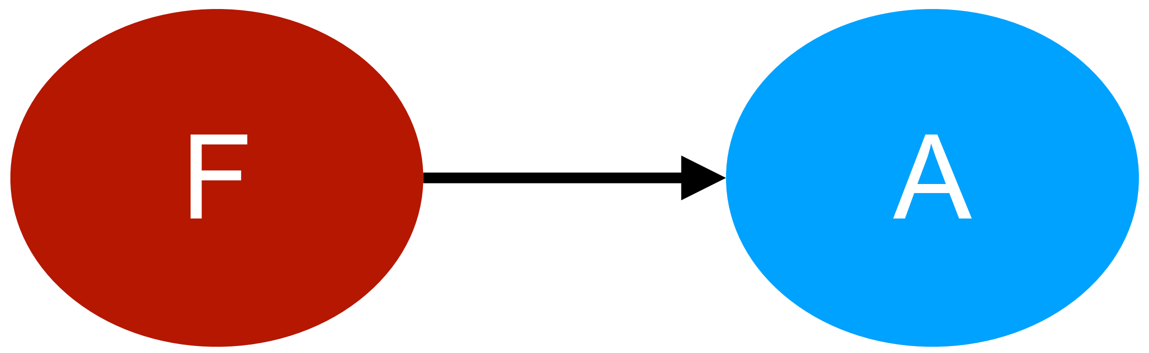
$$F \longleftarrow A$$

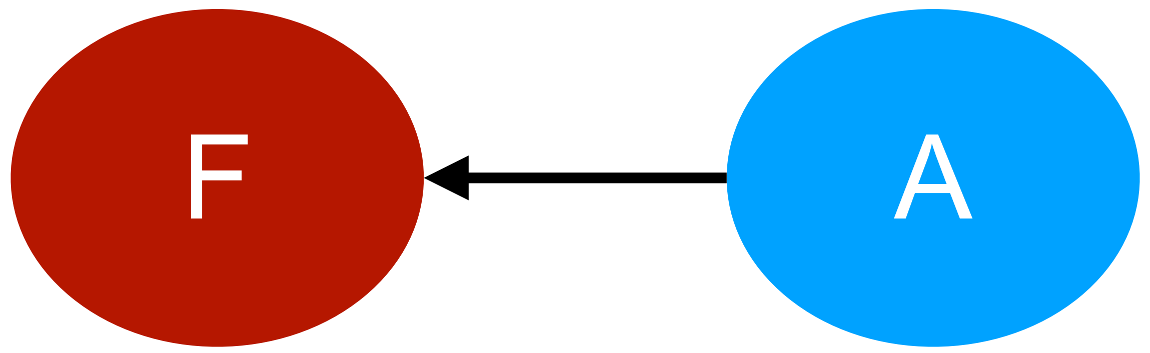
BNs vs causal BNs - example 1

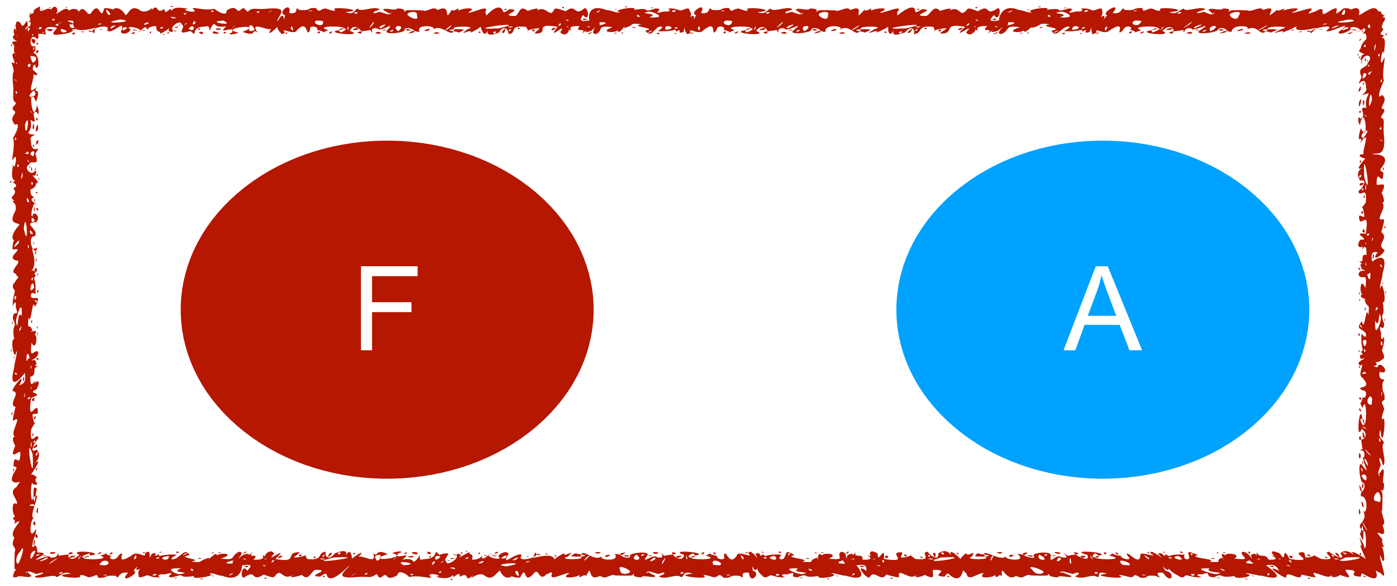
- Fire (F) and Alarm (A) with $p(F, A)$ and $A \not\perp\!\!\!\perp F$ can be factorized as:

$p(F,A) = p(F)\, p(A|F)$

$p(F,A) = p(A)\, p(F|A)$







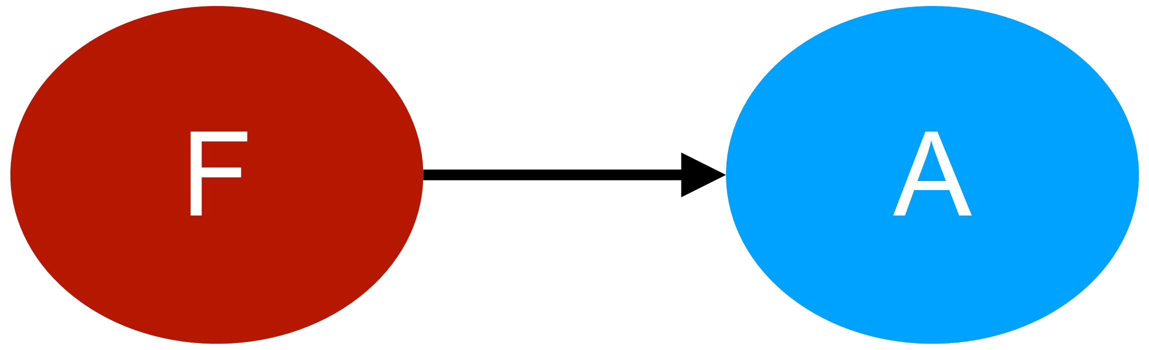
$p(F,A) = p(F) \cdot p(A)$

?

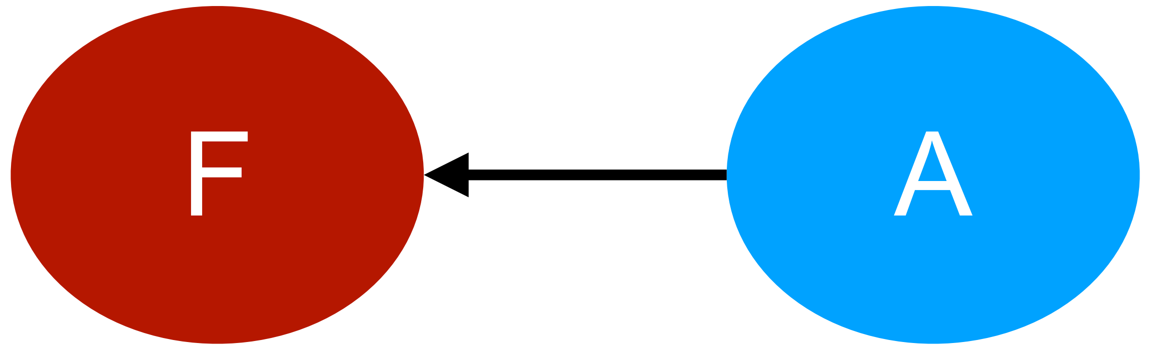
BNs vs causal BNs - example 1

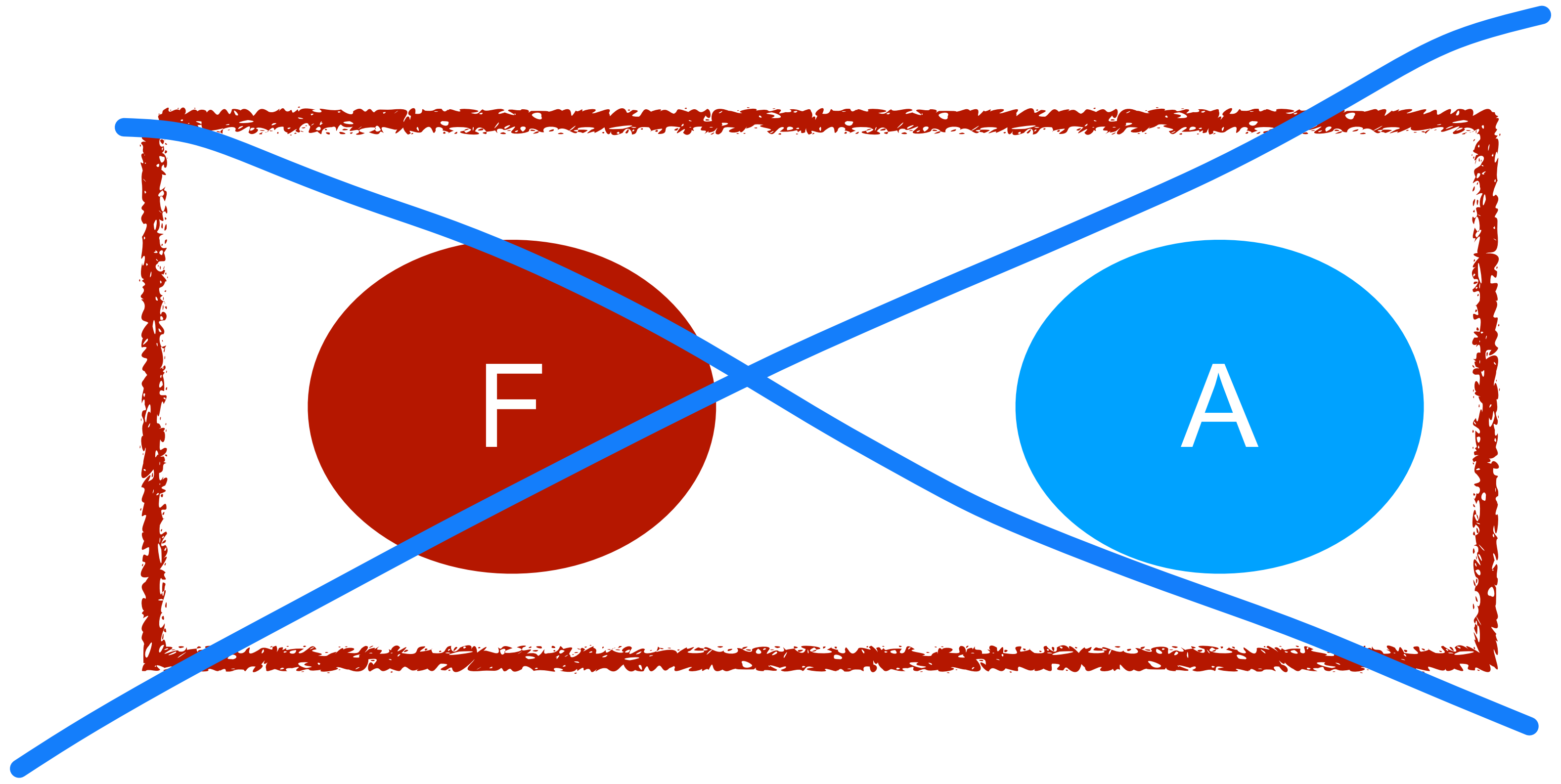
- Fire (F) and Alarm (A) with $p(F, A)$ and $A \not\perp\!\!\!\perp F$ can be factorized as:

$p(F, A) = p(F)\, p(A|F)$



$p(F, A) = p(A)\, p(F|A)$





$p(F, A) = p(F)\, p(A)$

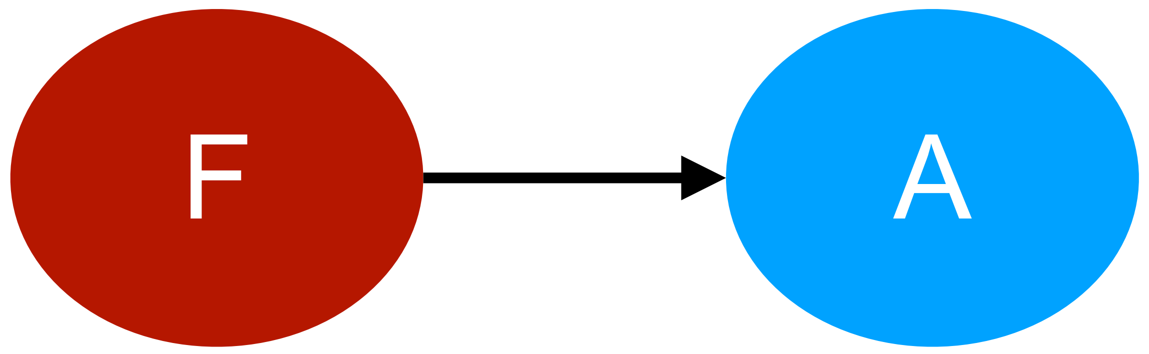
$\Rightarrow A \perp\!\!\!\perp F$

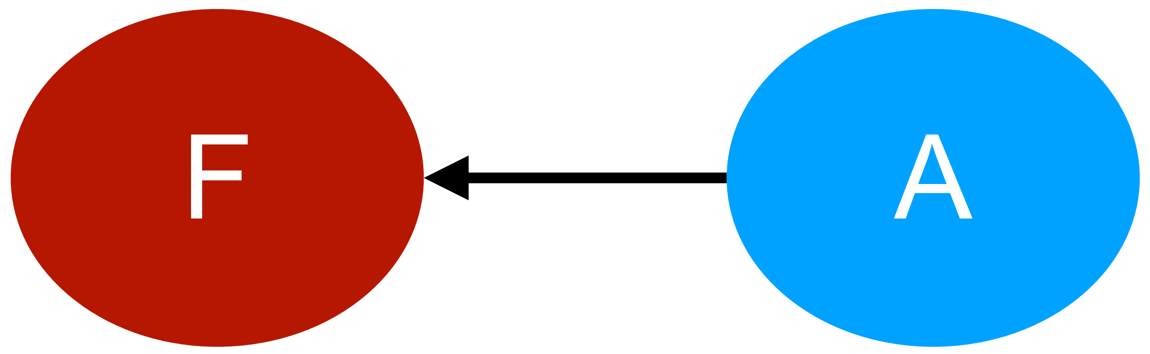
BNs vs causal BNs - example 1

- Fire (F) and Alarm (A) with $p(F, A)$ and $A \not\perp\!\!\!\perp F$ can be factorized as:

$p(F, A) = p(F)\, p(A|F)$

$p(F, A) = p(A)\, p(F|A)$





BNs vs causal BNs - example 1

- Fire (F) and Alarm (A) with $p(F, A)$ and $A \not\perp\!\!\!\perp F$ can be factorized as:

$p(F,A) = p(F)\, p(A|F)$

$F \rightarrow A$

CAUSAL

(lighting a fire triggers alarm)

$p(F,A) = p(A)\, p(F|A)$

$F \leftarrow A$

NOT-CAUSAL

(triggering alarm does not light a fire)

BNs vs causal BNs - example 1

A working definition of causality in machine learning

Informal definition: A variable X causes another variable Y, if changing (the distribution of) X, e.g. by fixing its value, changes (the distribution of) Y

$F \rightarrow A$

CAUSAL

(lighting a fire triggers alarm)

$F \leftarrow A$

NOT-CAUSAL

(triggering alarm does not light a fire)

BNs vs causal BNs - example 2

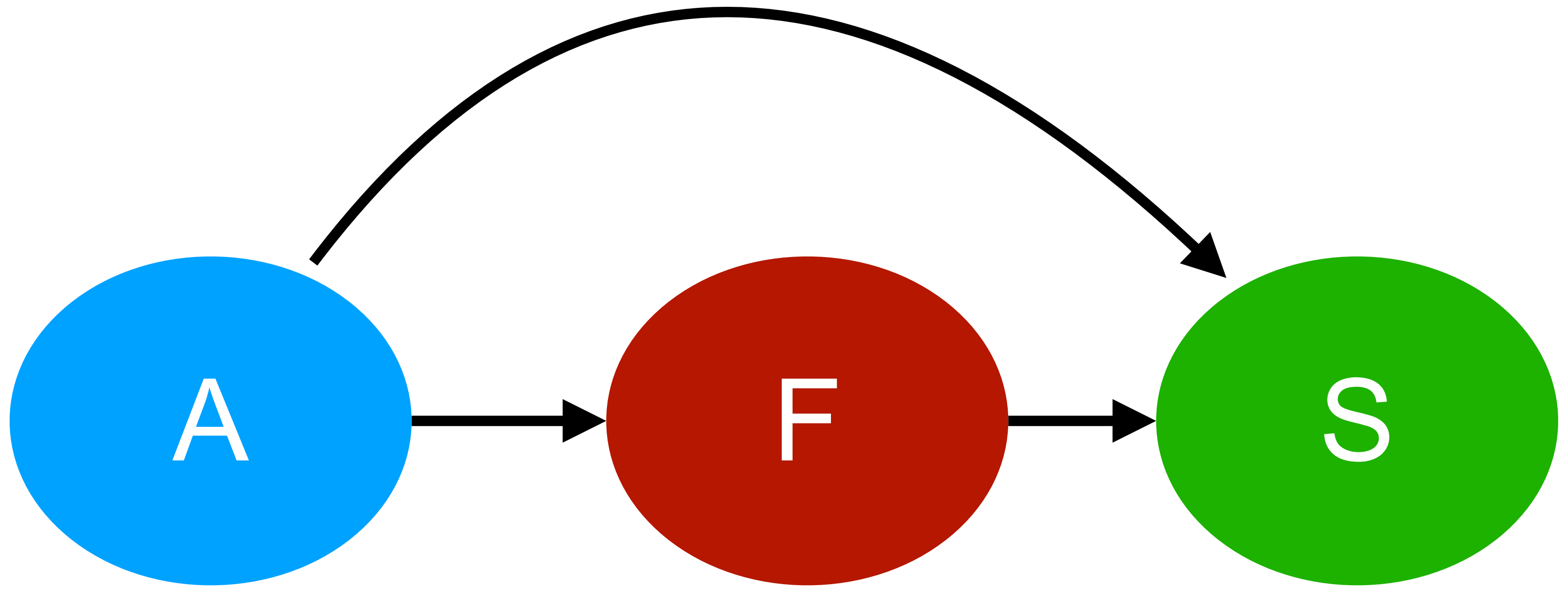
- Fire (F), Smoke (S), Alarm (A) with $p(F, S, A)$ and $A \perp\!\!\!\perp F \mid S$

$P(A, F, S) =$

BNs vs causal BNs - example 2

- Fire (F), Smoke (S), Alarm (A) with $p(F, S, A)$ and $A \perp\!\!\!\perp F \mid S$

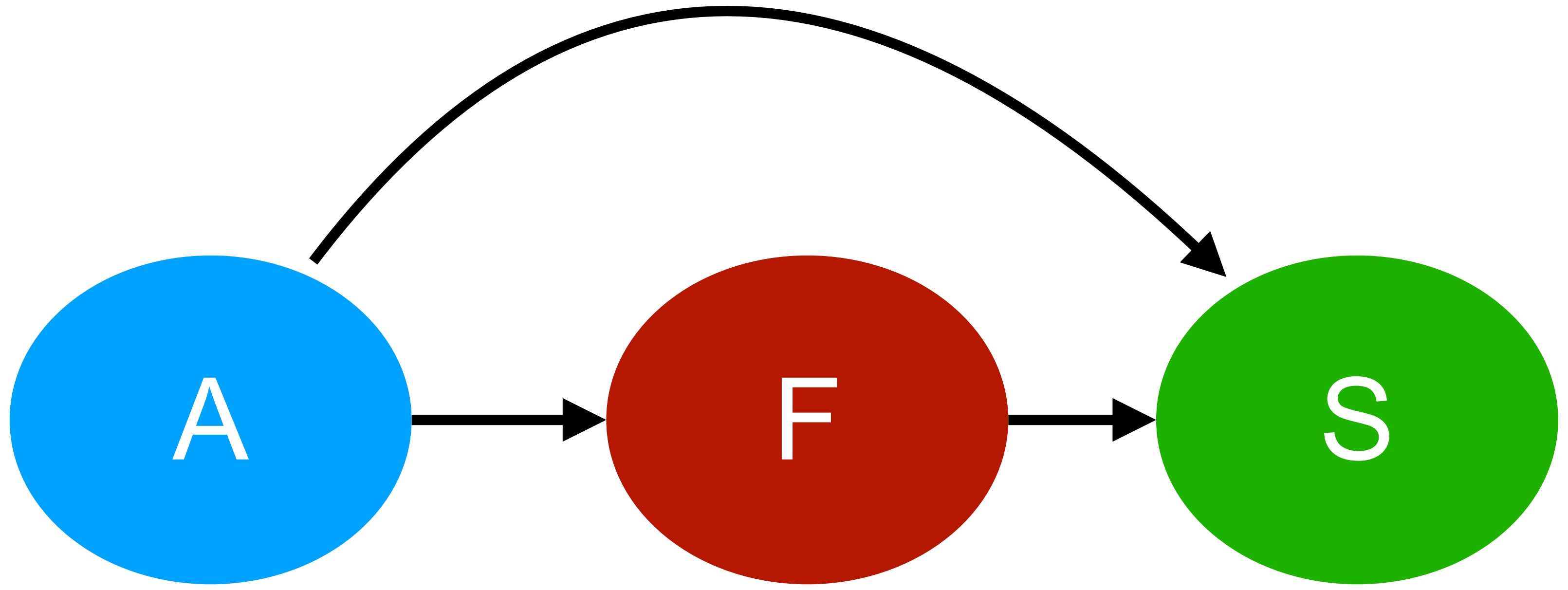
$$P(A,F,S) = P(A) \cdot P(F|A) \cdot P(S|F,A)$$



BNs vs causal BNs - example 2

- Fire (F), Smoke (S), Alarm (A) with $p(F, S, A)$ and $A \perp\!\!\!\perp F \mid S$

$$P(A,F,S) = P(A) \cdot P(F|A) \cdot P(S|F,A)$$



In the course, we will assume:

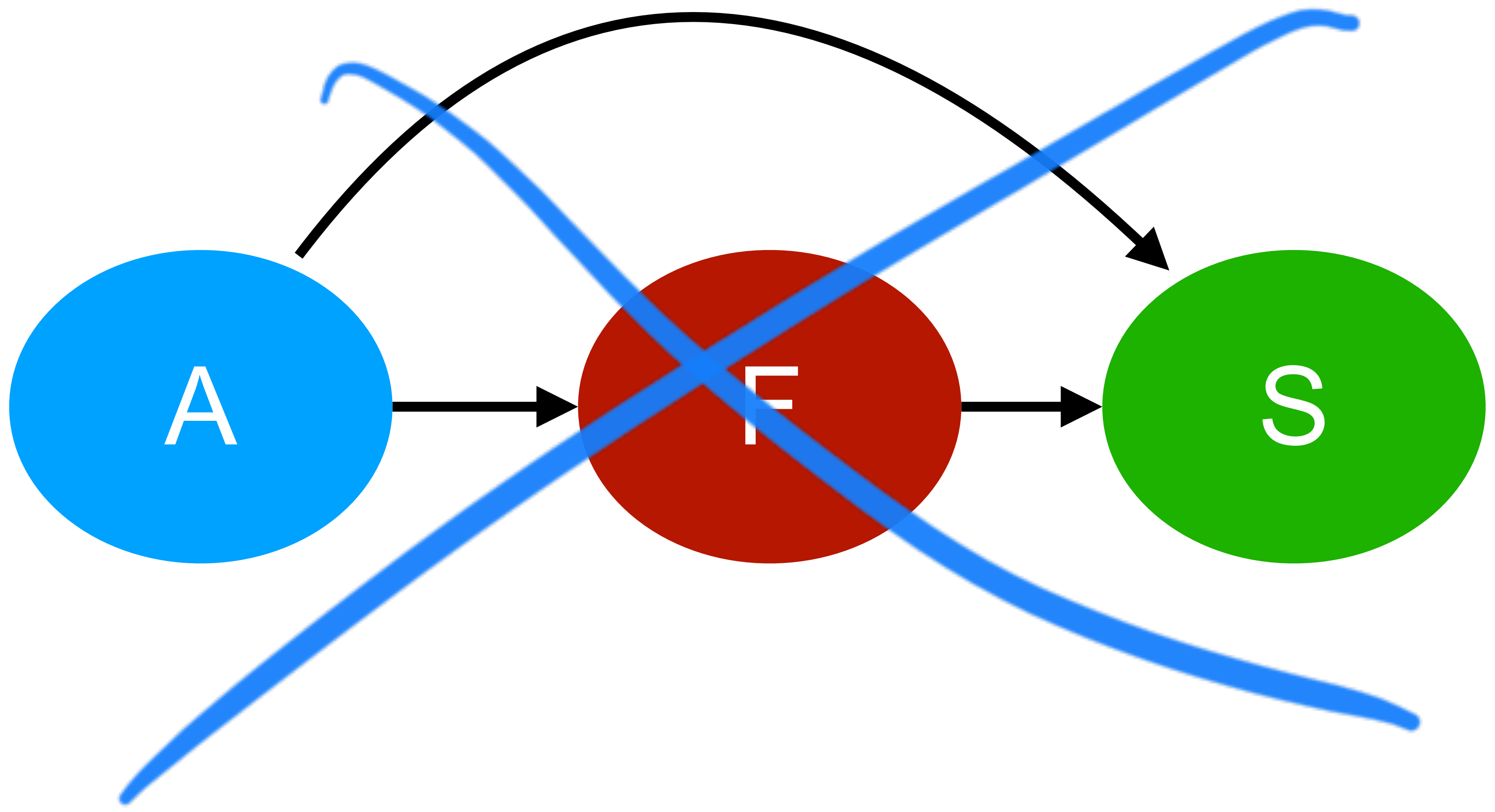
$\mathbf{A} \perp \mathbf{B} \mid \mathbf{C} \iff X_{\mathbf{A}} \perp\!\!\!\perp X_{\mathbf{B}} \mid X_{\mathbf{C}}$

$A \perp\!\!\!\perp F \mid S$

BNs vs causal BNs - example 2

- Fire (F), Smoke (S), Alarm (A) with $p(F, S, A)$ and $A \perp\!\!\!\perp F \mid S$

$$P(A, F, S) = P(A) \cdot P(F \mid A) \cdot P(S \mid F, A)$$



In the course, we will assume:
$\mathbf{A} \perp \mathbf{B} \mid \mathbf{C} \iff X_{\mathbf{A}} \perp\!\!\!\perp X_{\mathbf{B}} \mid X_{\mathbf{C}}$

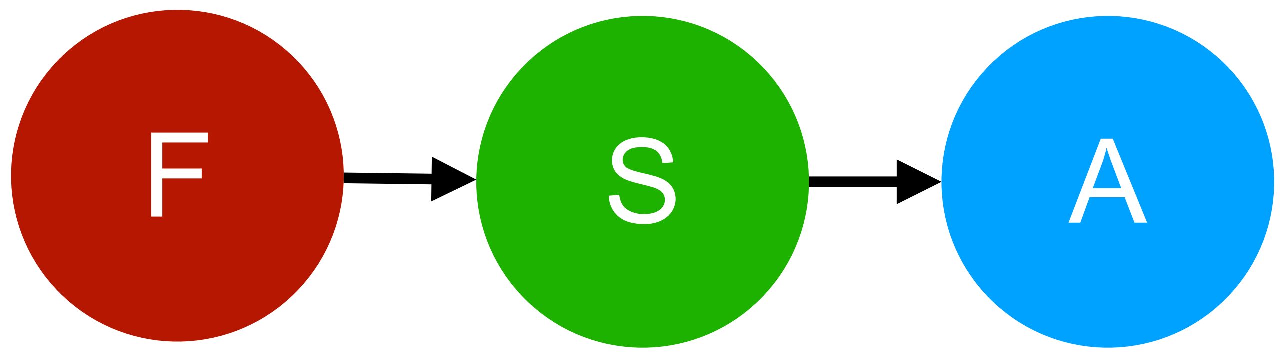
$$A \perp\!\!\!\perp F \mid S \not\Leftrightarrow A \not\perp F \mid S$$

This is a BN, but not a faithful BN

BNs vs causal BNs - example 2

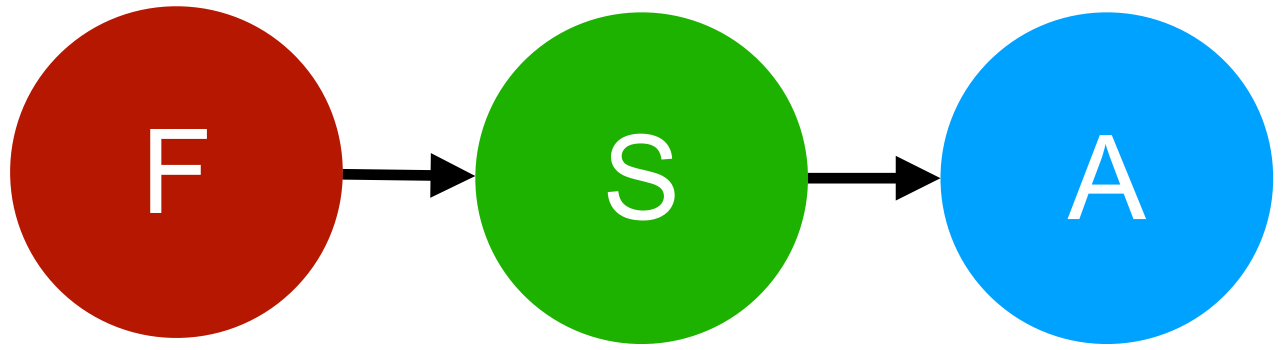
- Fire (F), Smoke (S), Alarm (A) with $p(F, S, A)$ and $A \perp\!\!\!\perp F \mid S$

$$p(F,S,A) = p(F) \cdot p(S \mid F) \cdot p(A \mid F,S)$$
$$\overset{!}{=} p(F) \cdot p(S \mid F) \cdot p(A \mid S)$$



BNs vs causal BNs - example 2

- Fire (F), Smoke (S), Alarm (A) with $p(F, S, A)$ and $A \perp\!\!\!\perp F \mid S$



$P(F) \cdot P(S|F) \cdot P(A|S)$

BNs vs causal BNs - example 2

- Fire (F), Smoke (S), Alarm (A) with $p(F, S, A)$ and $A \perp\!\!\!\perp F \mid S$

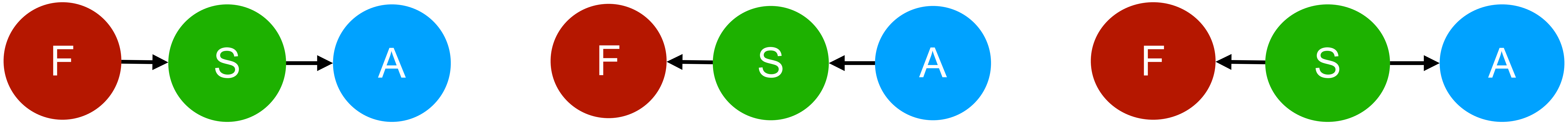
$F \rightarrow S \rightarrow A$ $\qquad$ $F \leftarrow S \leftarrow A$ $\qquad$ $F \leftarrow S \rightarrow A$

$P(F) \cdot P(S|F) \cdot P(A|S)$ $\qquad$ $P(A) \cdot P(S|A) \cdot P(F|S)$ $\qquad$ $P(S) \cdot P(F|S) \cdot P(A|S)$

BNs vs causal BNs - example 2

- Fire (F), Smoke (S), Alarm (A) with $p(F, S, A)$ and $A \perp\!\!\!\perp F \mid S$

$\boxed{\text{These are all faithful BNs, they have the same d-separation statements}}$

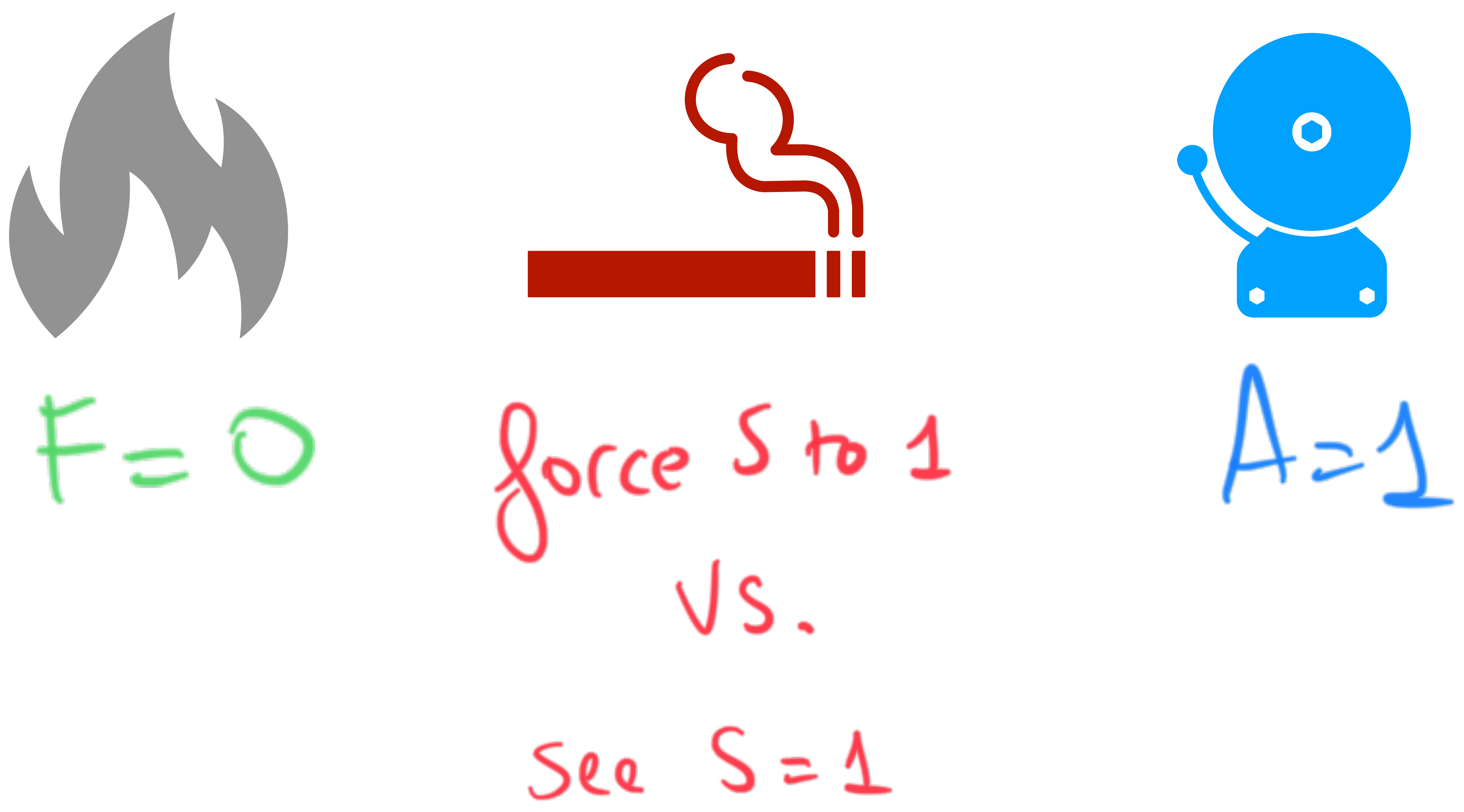


$P(F) \cdot P(S|F) \cdot P(A|S)$ — CAUSAL

$P(A) \cdot P(S|A) \cdot P(F|S)$ — NON-CAUSAL

$P(S) \cdot P(F|S) \cdot P(A|S)$ — NON-CAUSAL

Seeing is not doing



The $do(X = x)$ operator [Pearl 2009]

- We introduce a new operator that can represent a **hypothetical intervention** on the whole population, e.g. a perturbation of the system that fixes the value of X to x:

$$\mathrm{do}(X = x)$$

The $do(X = x)$ operator [Pearl 2009]

- We introduce a new operator that can represent a **hypothetical intervention** on the whole population, e.g. a perturbation of the system that fixes the value of X to x:

$$\mathrm{do}(X = x)$$

- We can then define the **(post-) interventional distribution for $X_{\mathbf{V}}$:**

$P(X_{\mathbf{V}} \mid \mathrm{do}(X = x))$, which in general $\neq P(X_{\mathbf{V}})$ **observational**

- $P(\,\cdot \mid \mathrm{do}(X = x))$ seems like conditioning, but it's a different distribution

Conditioning vs intervening example

Conditioning vs intervening example

$F \rightarrow A$

$P(F=1) = 0.01$

$P(A=1 \mid F=0) = 0.1$

$P(A=1 \mid F=1) = 0.8$

$P(F=1 \mid A=1)$?

Conditioning vs intervening example

$$F \rightarrow A$$

$$P(F=1) = 0.01$$

$$P(A=1 \mid F=0) = 0.1$$

$$P(A=1 \mid F=1) = 0.8$$

$$P(F=1 \mid A=1) = \frac{P(A=1 \mid F=1) \cdot P(F=1)}{P(A=1)}$$

Conditioning vs intervening example

$F \rightarrow A$

$P(F=1) = 0.01$

$P(A=1 \mid F=0) = 0.1$

$P(A=1 \mid F=1) = 0.8$

$$P(F=1 \mid A=1) = \frac{P(A=1 \mid F=1) \cdot P(F=1)}{P(A=1)}$$

$$P(A=1) = \overset{0.8}{P(A=1 \mid F=1)} \cdot \overset{0.01}{P(F=1)} + \underset{0.1}{P(A=1 \mid F=0)} \cdot \underset{0.99}{P(F=0)}$$

$$= 0.107$$

Conditioning vs intervening example

$F \rightarrow A$

$P(F=1) = 0.01$

$P(A=1 \mid F=0) = 0.1$

$P(A=1 \mid F=1) = 0.8$

$$P(F=1 \mid A=1) = \frac{P(A=1 \mid F=1) \cdot P(F=1)}{P(A=1)}$$

$P(A=1) = 0.107$

$$P(F=1 \mid A=1) = \frac{0.8 \cdot 0.01}{0.107} \simeq 0.075$$

Conditioning vs intervening example

$$F \rightarrow A$$

$$P(F=1) = 0.01$$

$$P(A=1 \mid F=0) = 0.1$$

$$P(A=1 \mid F=1) = 0.8$$

$$P(F=1 \mid A=1) \simeq 0.075$$

$$P(F=1 \mid do(A=1)) \ ?$$

Conditioning vs intervening example

$$F \rightarrow A$$

$$P(F=1) = 0.01$$

$$P(A=1 \mid F=0) = 0.1$$

$$P(A=1 \mid F=1) = 0.8$$

$$P(F=1 / A=1) \simeq 0.075$$

$$P(F=1 \mid do(A=1)) \ ?$$

intuitively

$$P(F=1 \mid do(A=1)) = P(F=1) = 0.01$$

A formal definition of causality

- X has a **causal effect** on Y iff

$$\exists x, x' : P(Y \mid \mathrm{do}(X = x)) \neq P(Y \mid \mathrm{do}(X = x'))$$

A formal definition of causality

- X has a **causal effect** on Y iff

$$\exists x, x' : P(Y \mid \mathrm{do}(X = x)) \neq P(Y \mid \mathrm{do}(X = x'))$$

- In the $F \to A$ example:

$$P(F = 1 \mid \mathrm{do}(A = 0)) = P(F = 1 \mid \mathrm{do}(A = 1)) = P(F = 1)$$
$$P(F = 0 \mid \mathrm{do}(A = 0)) = P(F = 0 \mid \mathrm{do}(A = 1)) = P(F = 0)$$

A formal definition of causality

- X has a **causal effect** on Y iff

$$\exists x, x' : P(Y \mid \mathrm{do}(X = x)) \neq P(Y \mid \mathrm{do}(X = x'))$$

- In the $F \to A$ example:

$$P(F = 1 \mid \mathrm{do}(A = 0)) = P(F = 1 \mid \mathrm{do}(A = 1)) = P(F = 1)$$

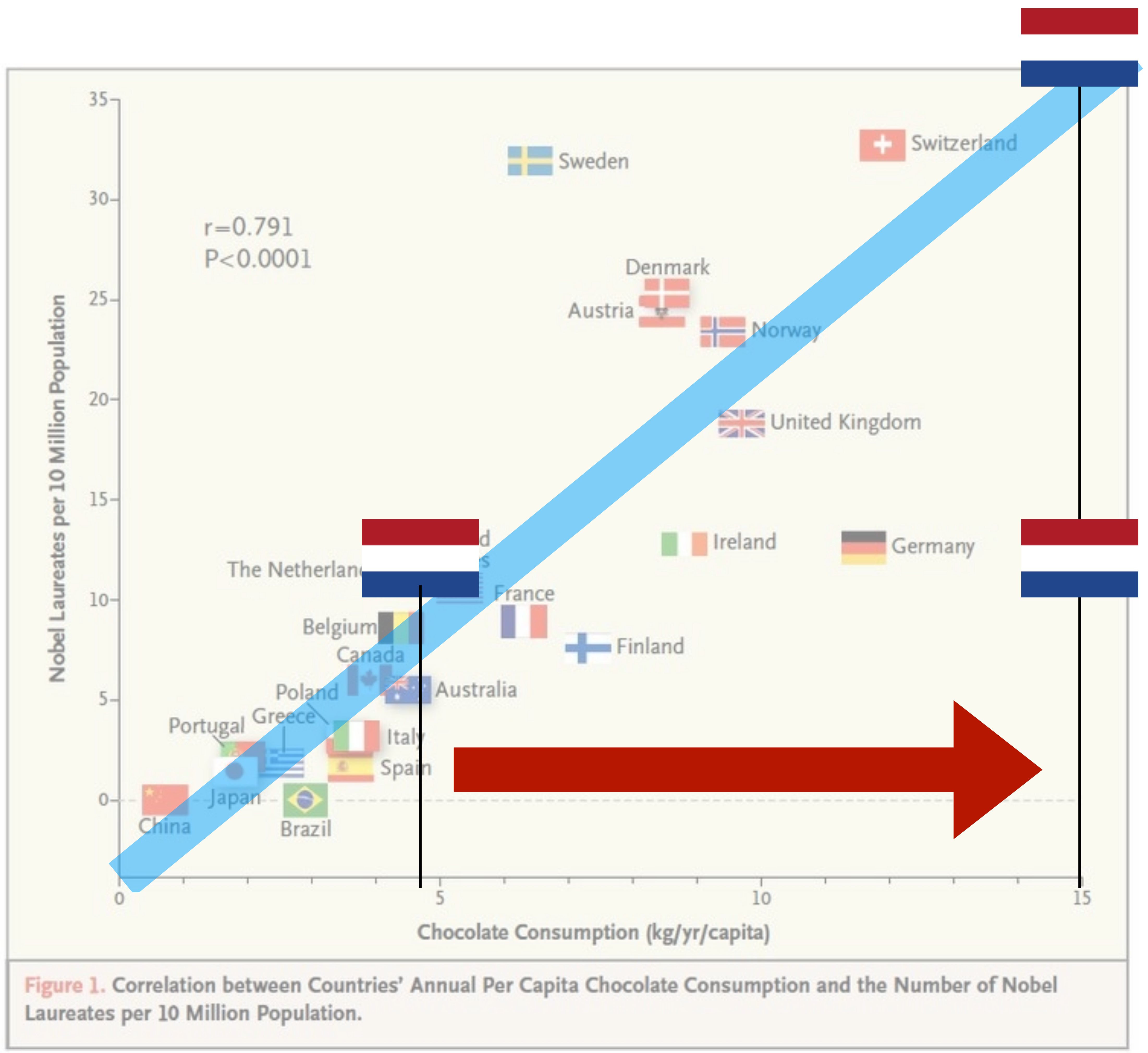
$$P(F = 0 \mid \mathrm{do}(A = 0)) = P(F = 0 \mid \mathrm{do}(A = 1)) = P(F = 0)$$

  - => A does not cause F

- We will discuss next week the **average causal effect:**

$$\mathrm{ACE}(x, x') = \mathbb{E}[Y \mid do(X = x)] - \mathbb{E}[Y \mid do(X = x')]$$

From Introduction: helping the Dutch government with decision making



$C \rightarrow N$

$C \not\rightarrow N$

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

Another example of causal effect vs no effect for $\mathrm{do}(X = 0)$

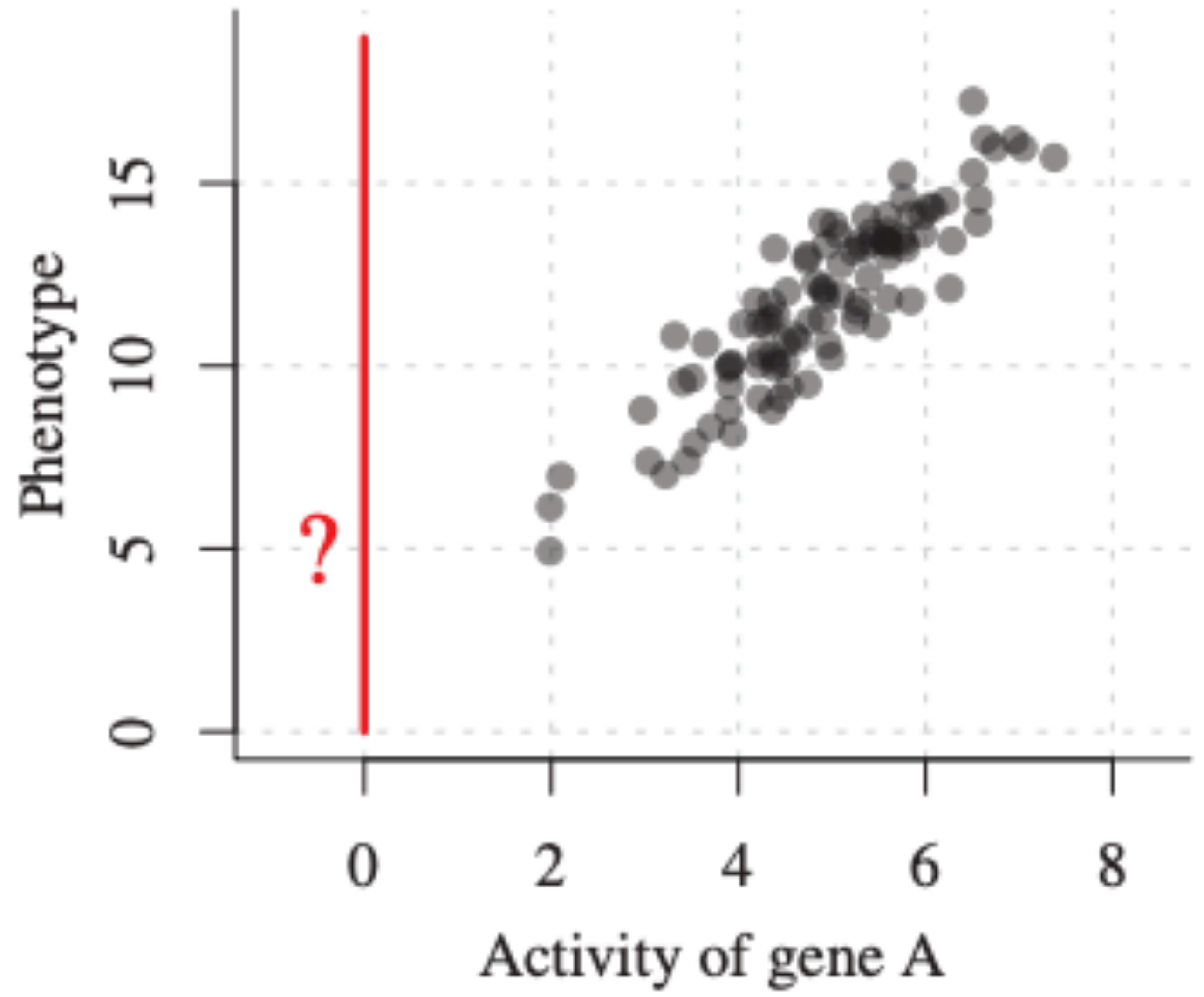


Fig 1.4 in Elements of Causal Inference (http://web.math.ku.dk/~peters/jonas_files/ElementsOfCausalInference.pdf)

Another example of causal effect vs no effect for $\mathrm{do}(X = 0)$

p(Phenotype | gene A)

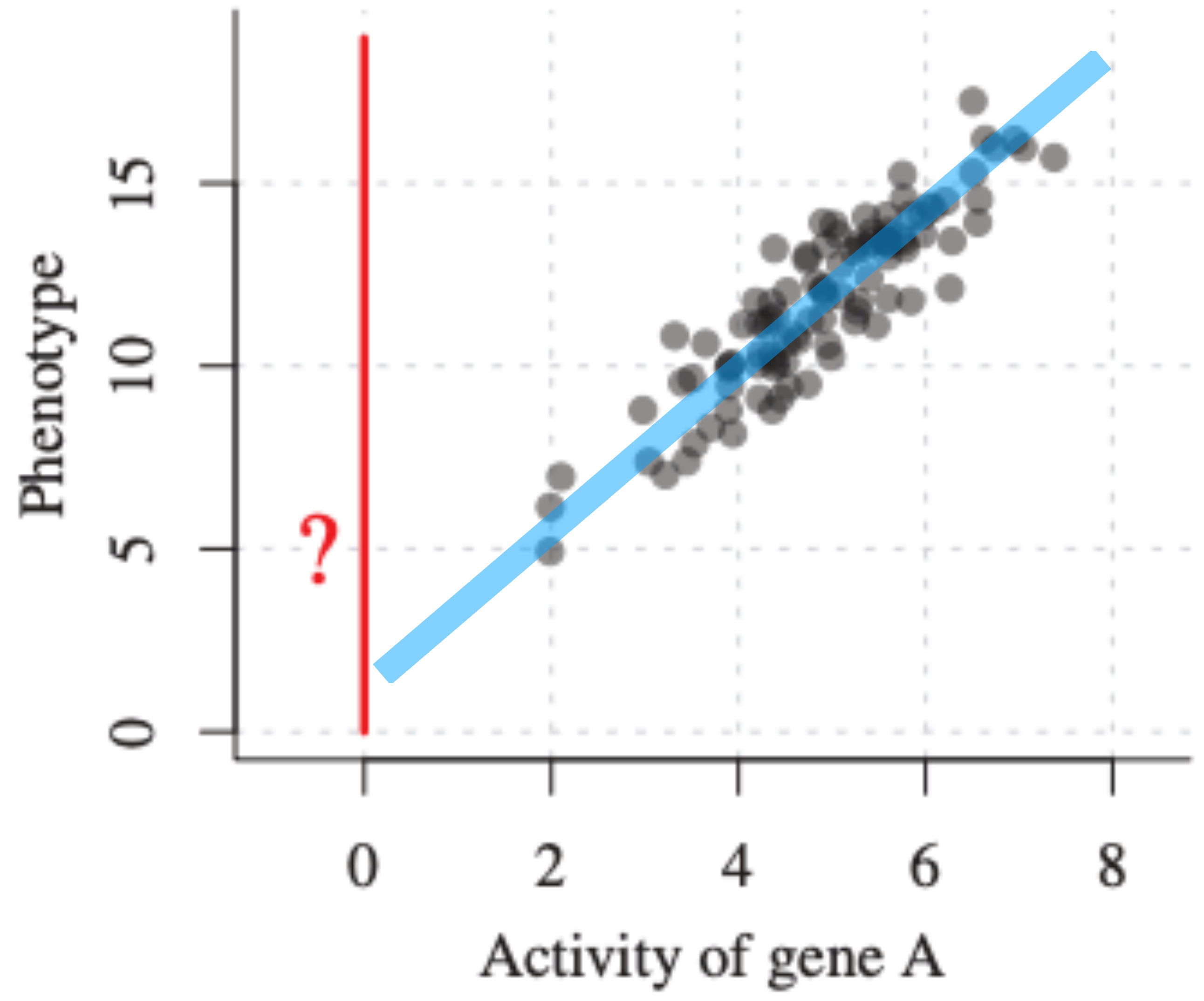
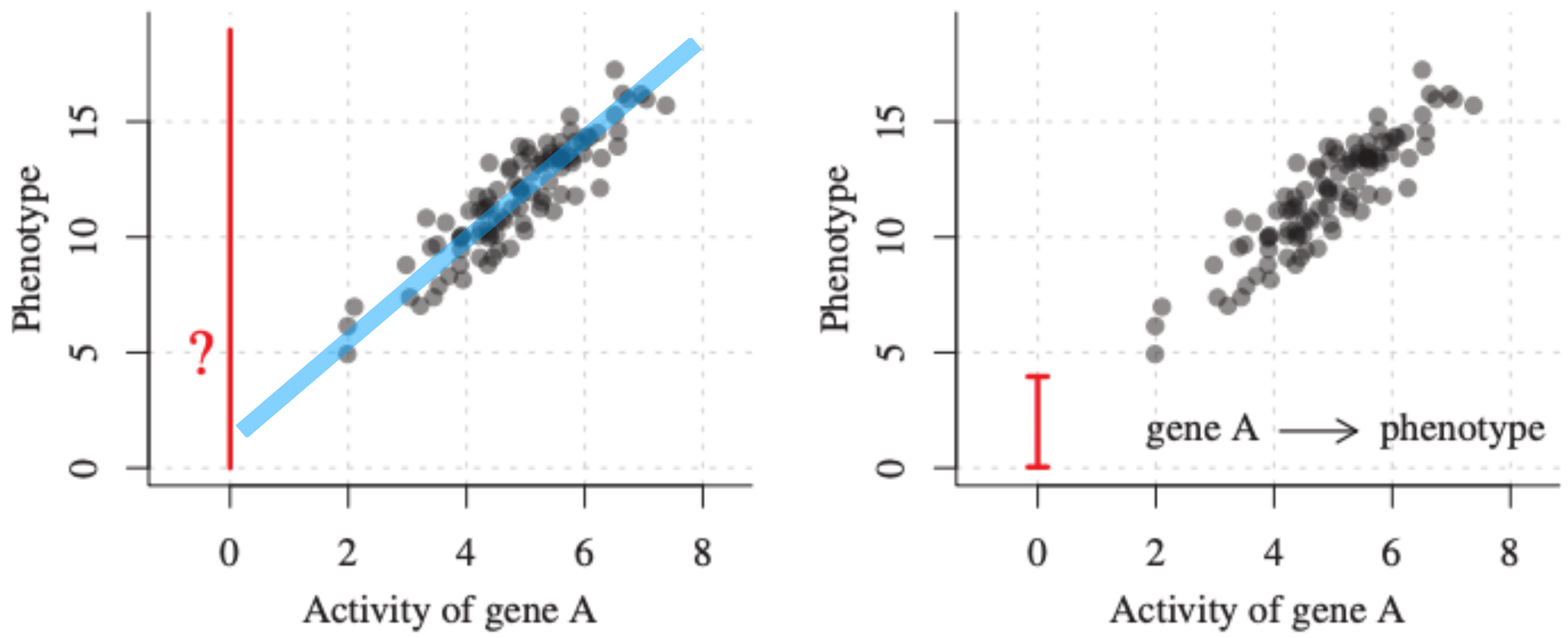


Fig 1.4 in Elements of Causal Inference (http://web.math.ku.dk/~peters/jonas_files/ElementsOfCausalInference.pdf)

Another example of causal effect vs no effect for $\mathrm{do}(X = 0)$

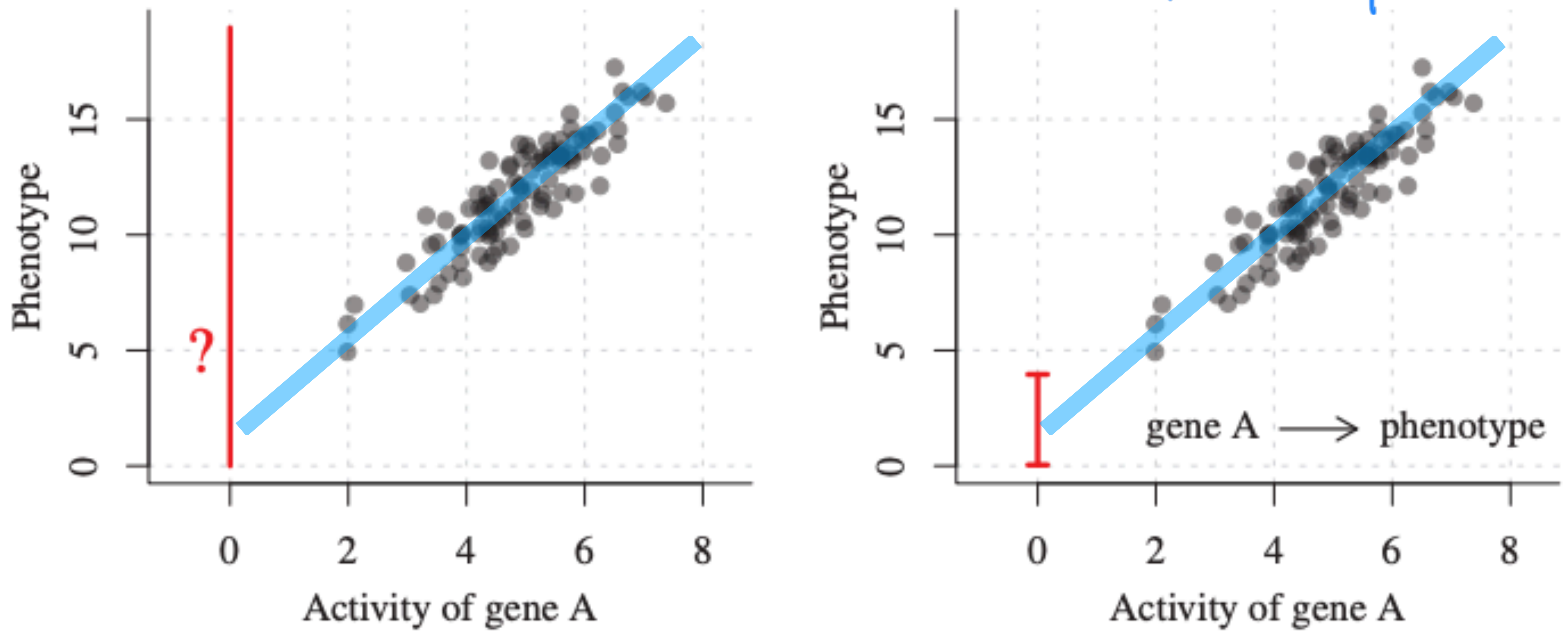


$\mathrm{do}(A = 0)$

Fig 1.4 in Elements of Causal Inference (http://web.math.ku.dk/~peters/jonas_files/ElementsOfCausalInference.pdf)

Another example of causal effect vs no effect for $\mathrm{do}(X = 0)$

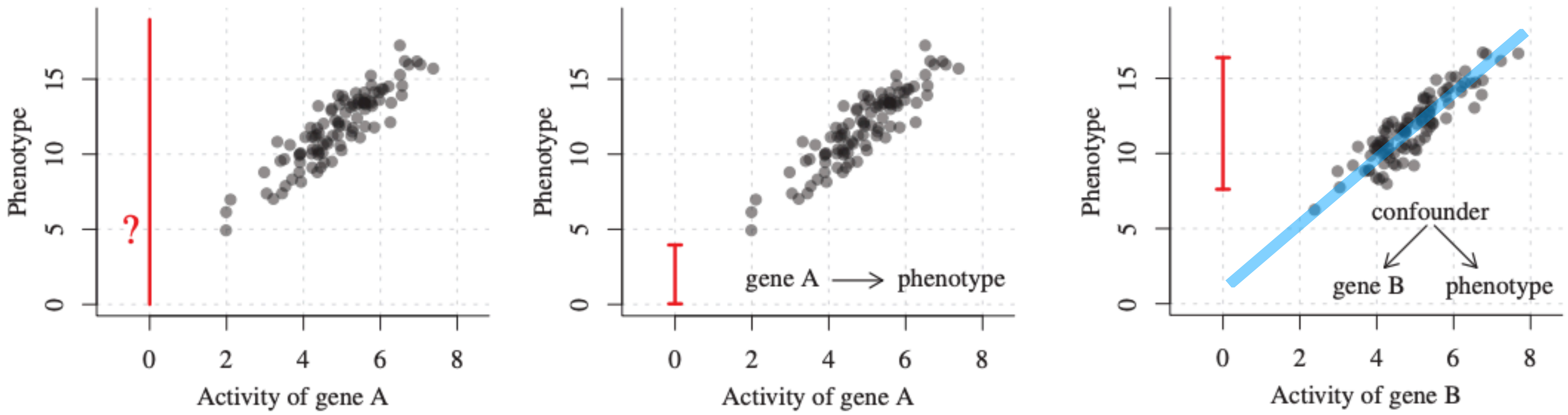
$p(\text{Phenotype} \mid \text{gene } A)$



$\mathrm{do}(A = 0)$

Fig 1.4 in Elements of Causal Inference (http://web.math.ku.dk/~peters/jonas_files/ElementsOfCausalInference.pdf)

Another example of causal effect vs no effect for $\mathrm{do}(X = 0)$



$\mathrm{do}(A = 0)$ $\mathrm{do}(B = 0)$

Fig 1.4 in Elements of Causal Inference (http://web.math.ku.dk/~peters/jonas_files/ElementsOfCausalInference.pdf)

Causal Bayesian networks

- Given DAG $G = (\mathbf{V}, \mathbf{E})$ and distribution p, (G, p) is a Bayesian network if

$$p(X_1, \ldots, X_p) = \prod_{i \in \mathbf{V}} p(X_i \mid X_{\mathrm{Pa}(i)})$$

Causal Bayesian networks

- Given DAG $G = (\mathbf{V}, \mathbf{E})$ and distribution p, (G, p) is a Bayesian network if

$$p(X_1, \ldots, X_p) = \prod_{i \in \mathbf{V}} p(X_i \mid X_{\mathrm{Pa}(i)})$$

- If for any $\mathbf{W} \subset \mathbf{V}$:

$$p(X_{\mathbf{V}} \mid \mathrm{do}(X_{\mathbf{W}} = x_{\mathbf{W}})) = \begin{cases} \prod_{i \in \mathbf{V} \setminus \mathbf{W}} p(X_i \mid X_{\mathrm{Pa}(i)}) & \text{if } \ X_{\mathbf{W}} = x_{\mathbf{W}} \\ 0 & \text{otherwise } \ (\text{if } \ X_{\mathbf{W}} \neq x_{\mathbf{W}}) \end{cases}$$

(G, p) is a **causal Bayesian network**

Causal Bayesian networks

- Given DAG $G = (\mathbf{V}, \mathbf{E})$ and distribution p, (G, p) is a Bayesian network if

$$p(X_1, \ldots, X_p) = \prod_{i \in V} p(X_i \mid \mathbf{X}_{\mathrm{pa}(i)})$$

- If for any $\mathbf{W} \subset \mathbf{V}$:

$\cancel{p(X_{\bar{w}} \mid X_{pa(\bar{w})})}$

$$p(X_{\mathbf{V}} \mid \mathrm{do}(X_{\mathbf{W}} = x_{\mathbf{W}})) = \prod_{i \in \bar{V} \setminus \bar{W}} p(X_i \mid X_{pa(i)}) \cdot \underbrace{\mathbb{1}(X_{\bar{w}} = x_{\bar{w}})}_{\text{indicator function}}$$

(G, p) is a **causal Bayesian network**

Causal Bayesian networks

- Given DAG $G = (\mathbf{V}, \mathbf{E})$ and distribution p, (G, p) is a Bayesian network if

$$p(X_1, \ldots, X_p) = \prod_{i \in V} p(X_i \mid \mathbf{X}_{\mathrm{pa}(i)})$$

- If for any $\mathbf{W} \subset \mathbf{V}$:

$$p(X_{\mathbf{V}} \mid \mathrm{do}(X_{\mathbf{W}} = x_{\mathbf{W}})) = \prod_{i \in \bar{V} \setminus \bar{W}} p(X_i \mid X_{pa(i)}) \cdot \mathbb{1}(x_{\bar{w}} = x_{\bar{w}})$$

(G, p) is a **causal Bayesian network**

$\boxed{\textbf{Parents are now direct causes}}$

Truncated factorisation formula [Pearl 2009]

- If for any $\mathbf{W} \subset \mathbf{V}$:

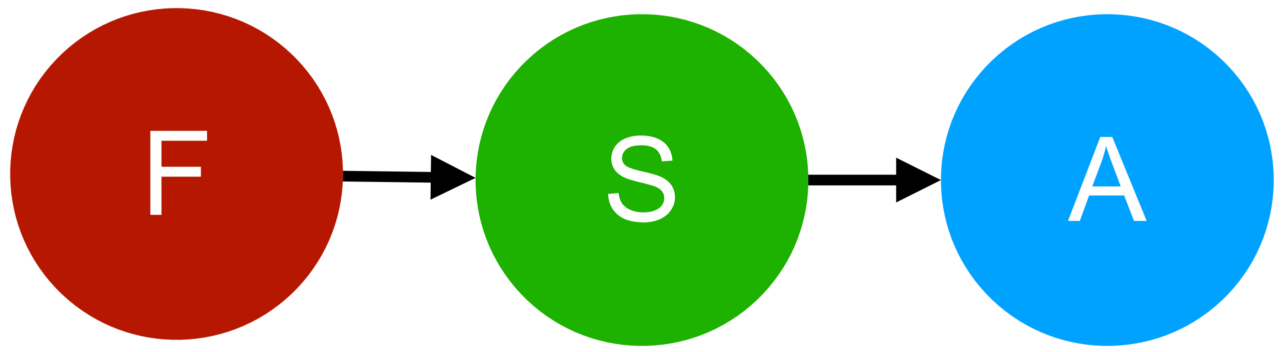
$$p(X_{\mathbf{V}} \,|\, \mathrm{do}(X_{\mathbf{W}} = x_{\mathbf{W}})) = \underbrace{\prod_{i \in \bar{V} \setminus \bar{W}} p(X_i \,|\, X_{pa(i)})}_{\text{doesn't change}} \cdot \underbrace{\mathbb{1}(X_{\bar{w}} = x_{\bar{w}})}_{\text{consistent with intervention}}$$

- Includes also observational data $\mathbf{W} = \emptyset$,
- Includes also multiple **intervention targets** $|\mathbf{W}| \geq 1$

Mutilated/manipulated graphs

Graphically $p(X_{\mathbf{V}} \,|\, \mathrm{do}(X_{\mathbf{W}} = x_{\mathbf{W}})) = \prod_{i \in \bar{V} \setminus \bar{W}} p(X_i \,|\, X_{pa(i)}) \cdot \mathbb{1}(X_{\bar{w}} = x_{\bar{w}})$

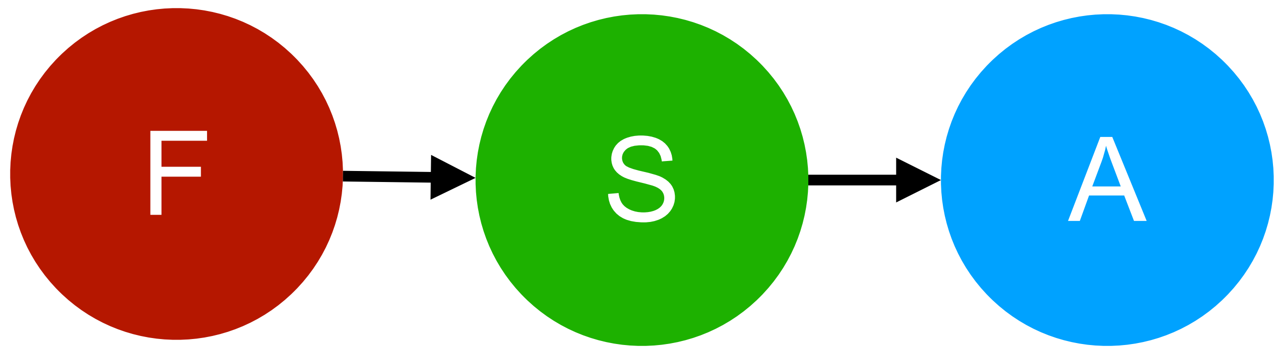
can be represented as cutting the incoming edges to $X_{\mathbf{W}}$

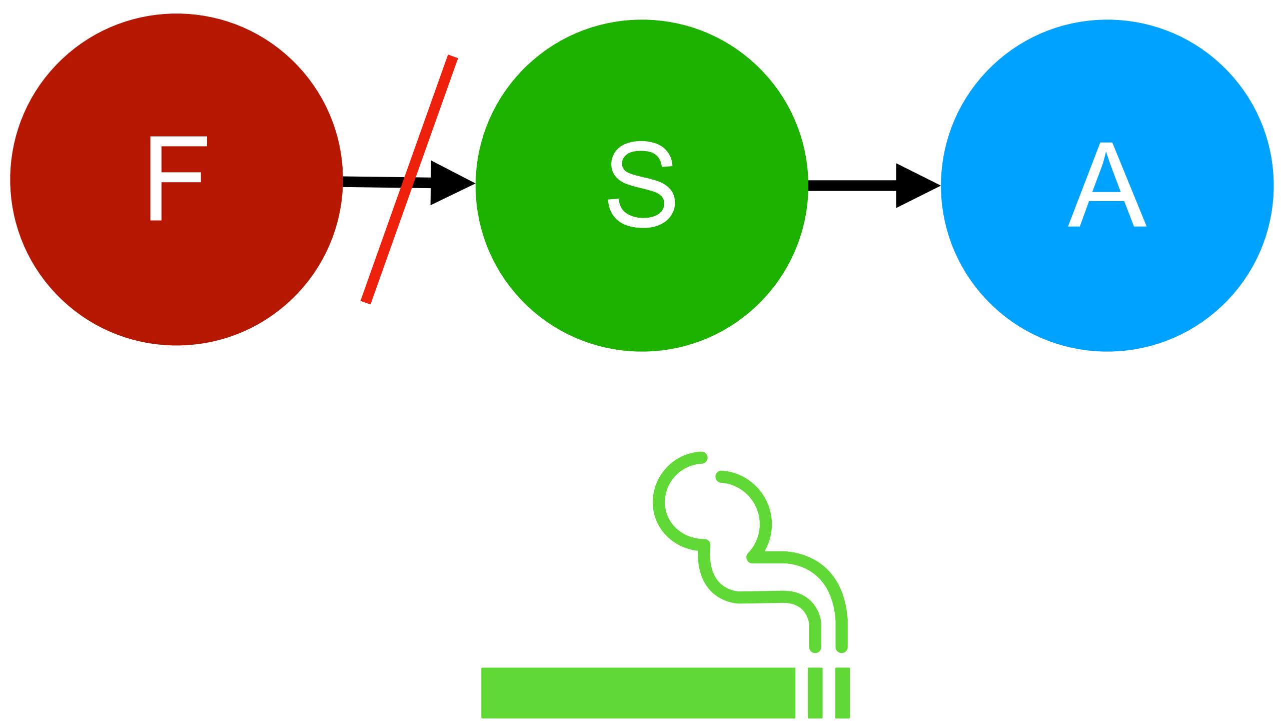


Mutilated/manipulated graphs

Graphically $p(X_{\mathbf{V}} \mid \mathrm{do}(X_{\mathbf{W}} = x_{\mathbf{W}})) = \prod_{i \in \bar{V} \setminus \bar{W}} p(X_i \mid X_{pa(i)}) \cdot \mathbb{1}(X_{\bar{W}} = x_{\bar{W}})$

can be represented as cutting the incoming edges to $X_{\mathbf{W}}$



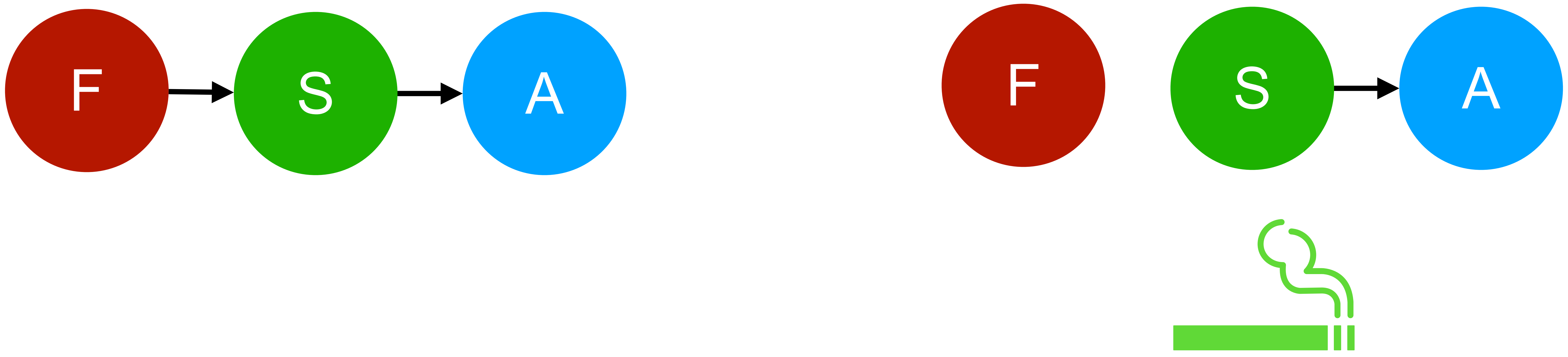


$\mathrm{do}(S = 1)$

Mutilated/manipulated graphs

Graphically $p(X_{\mathbf{V}} \mid \mathrm{do}(X_{\mathbf{W}} = x_{\mathbf{W}})) = \prod_{i \in \bar{V} \setminus \bar{W}} p(X_i \mid X_{pa(i)}) \cdot \mathbb{1}(X_{\bar{w}} = x_{\bar{w}})$

can be represented as cutting the incoming edges to $X_{\mathbf{W}}$

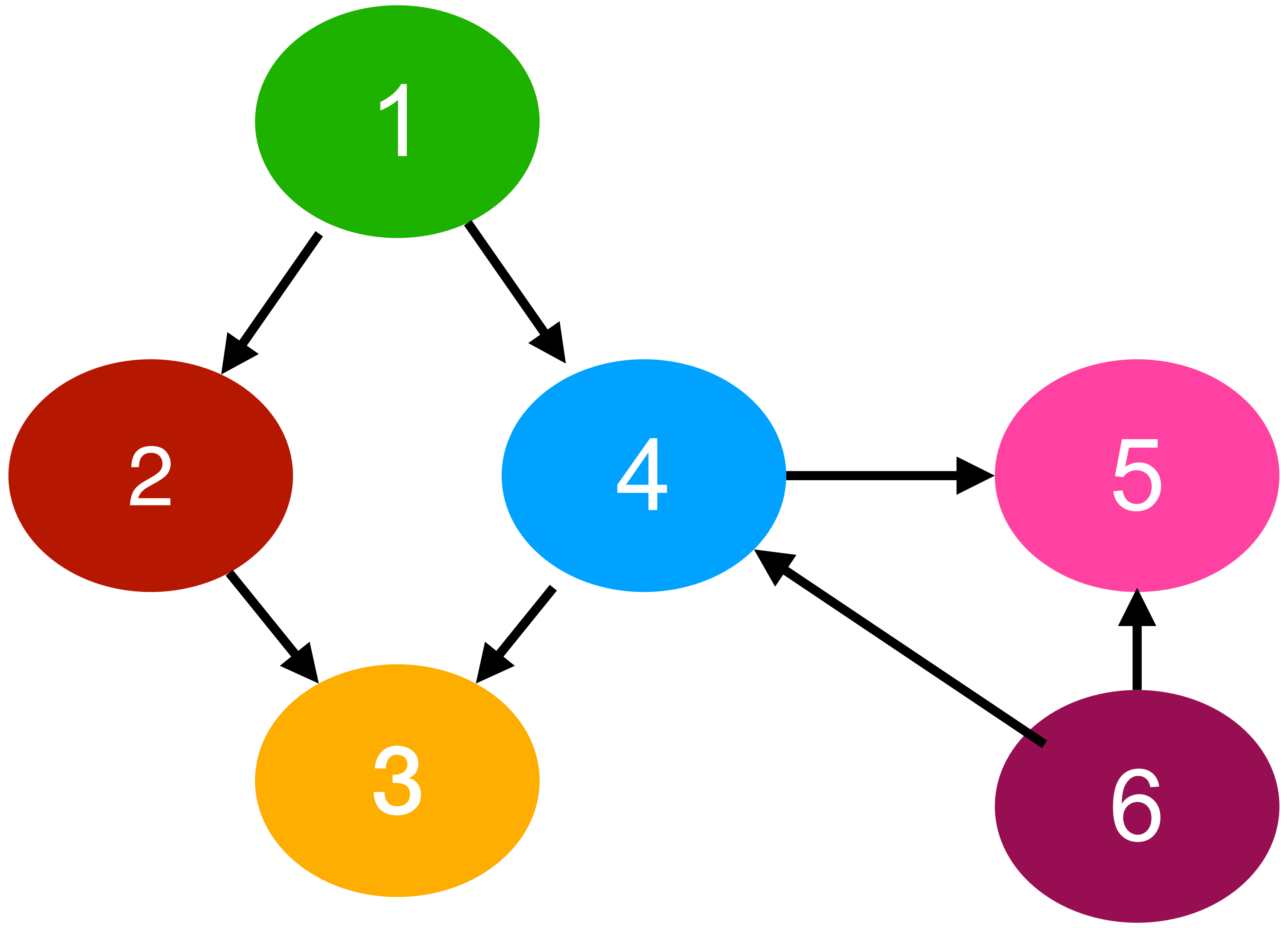


$\mathrm{do}(S = 1)$

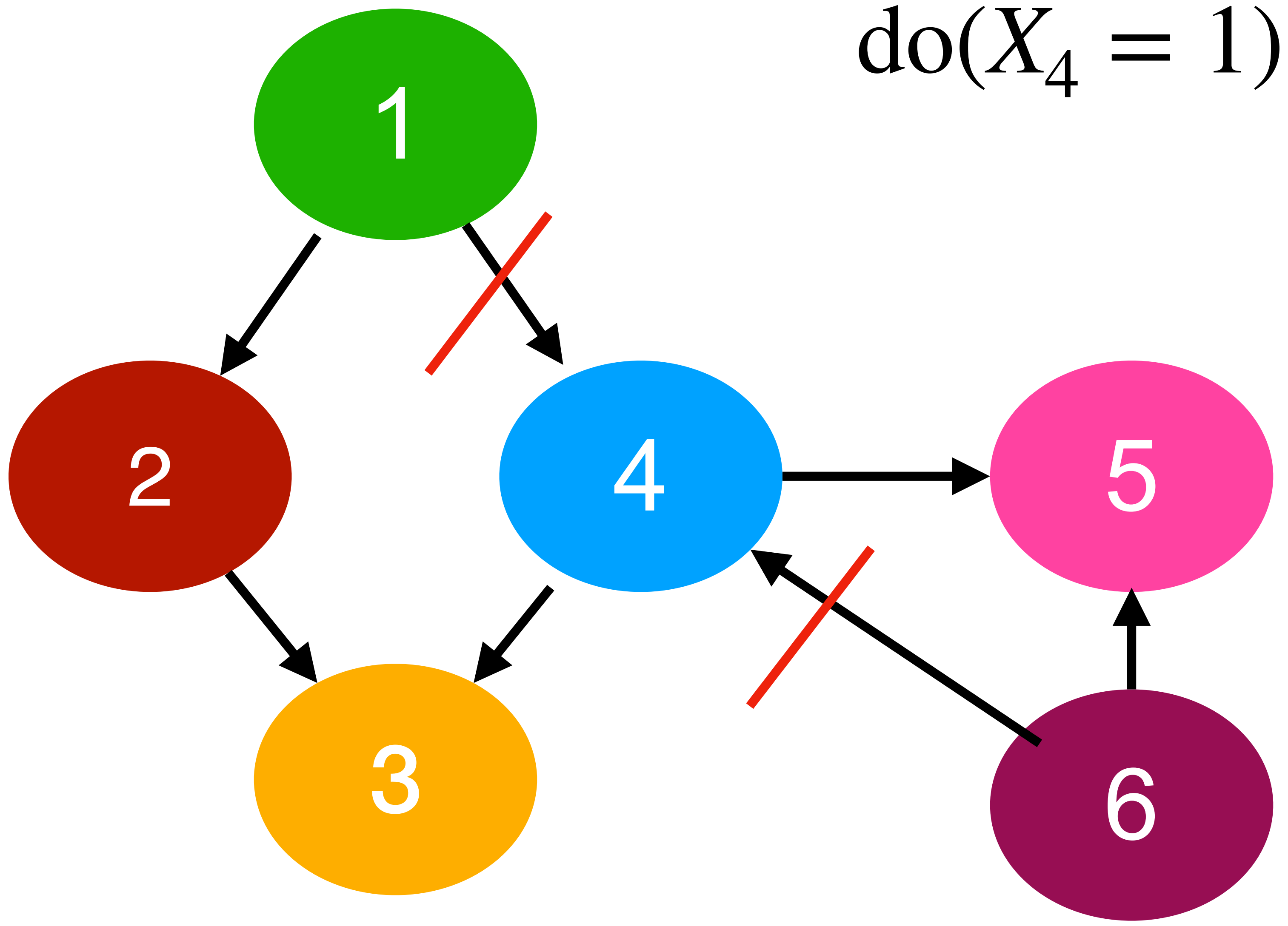
Mutilated/manipulated graphs

Graphically $p(X_{\mathbf{V}} \mid \mathrm{do}(X_{\mathbf{W}} = x_{\mathbf{W}})) = \prod_{i \in \bar{V} \setminus \bar{W}} p(X_i \mid X_{pa(i)}) \cdot \mathbb{1}(X_{\bar{w}} = x_{\bar{w}})$

can be represented as cutting the incoming edges to $X_{\mathbf{W}}$



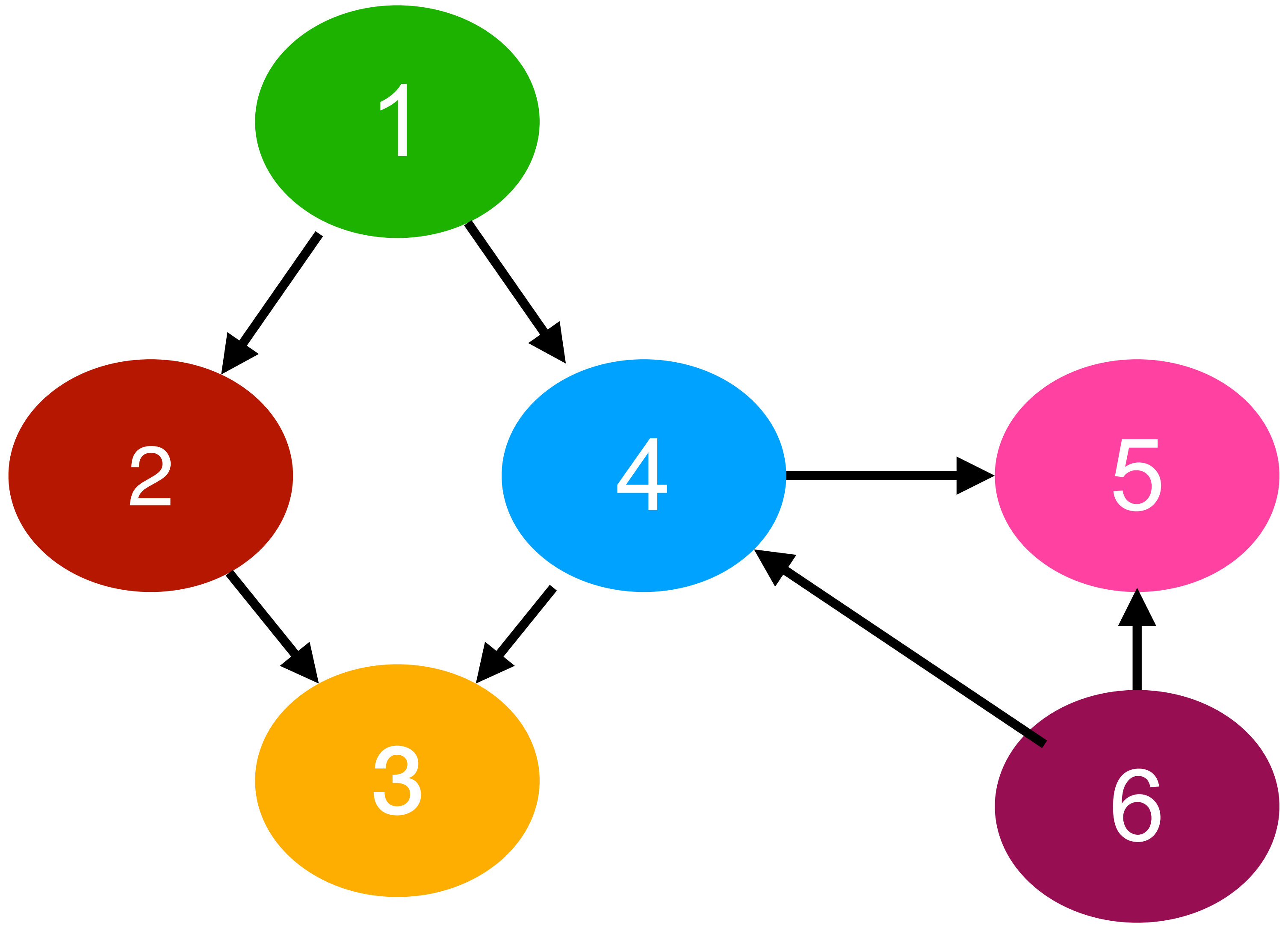
$\mathrm{do}(X_4 = 1)$



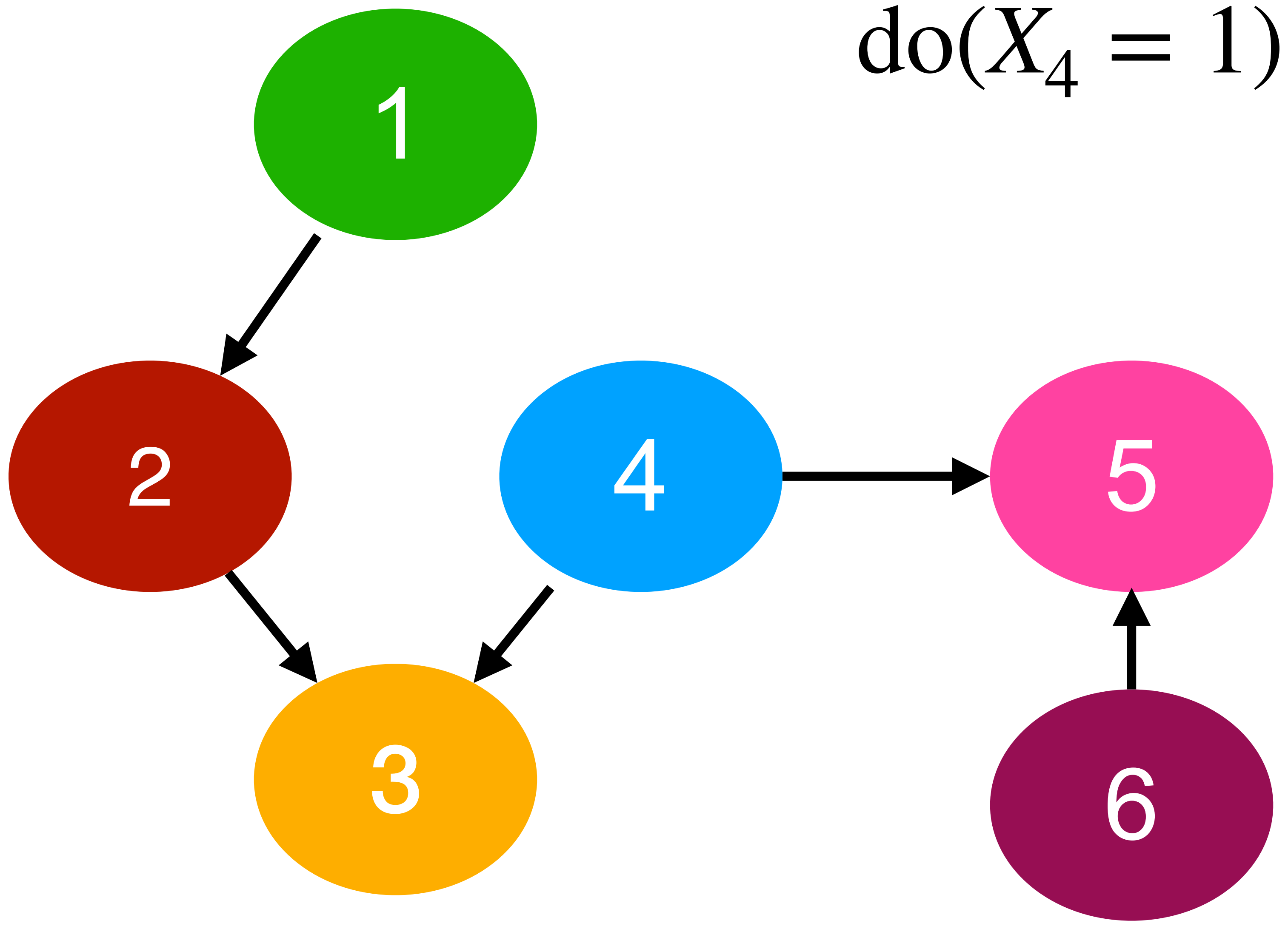
Mutilated/manipulated graphs

Graphically $p(X_\mathbf{V} \mid \mathrm{do}(X_\mathbf{W} = x_\mathbf{W})) = \prod_{i \in \bar{V} \setminus \bar{W}} p(X_i \mid X_{Pa(i)}) \cdot \mathbb{1}(X_{\bar{W}} = x_{\bar{W}})$

can be represented as cutting the incoming edges to $X_\mathbf{W}$



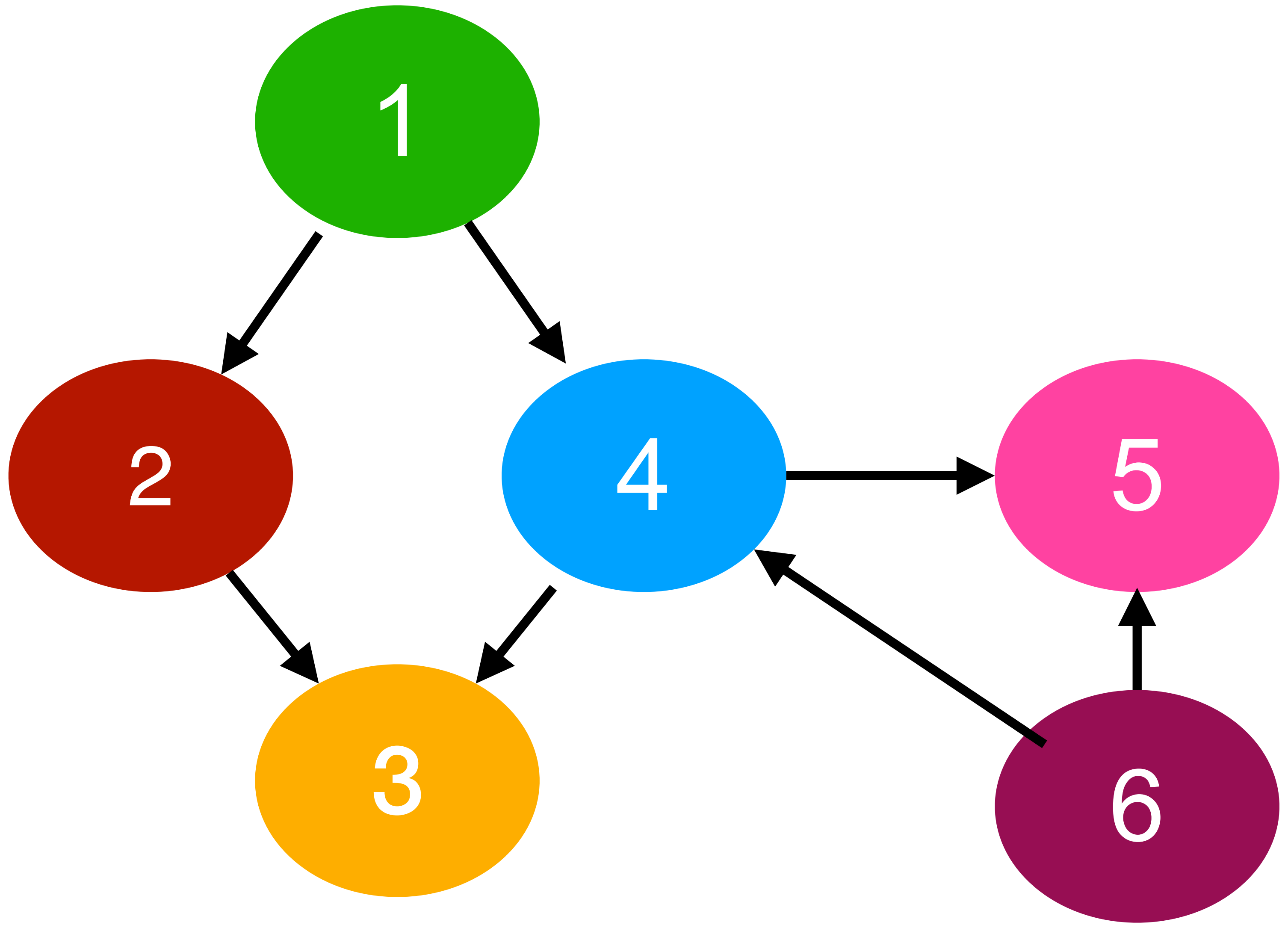
$\mathrm{do}(X_4 = 1)$



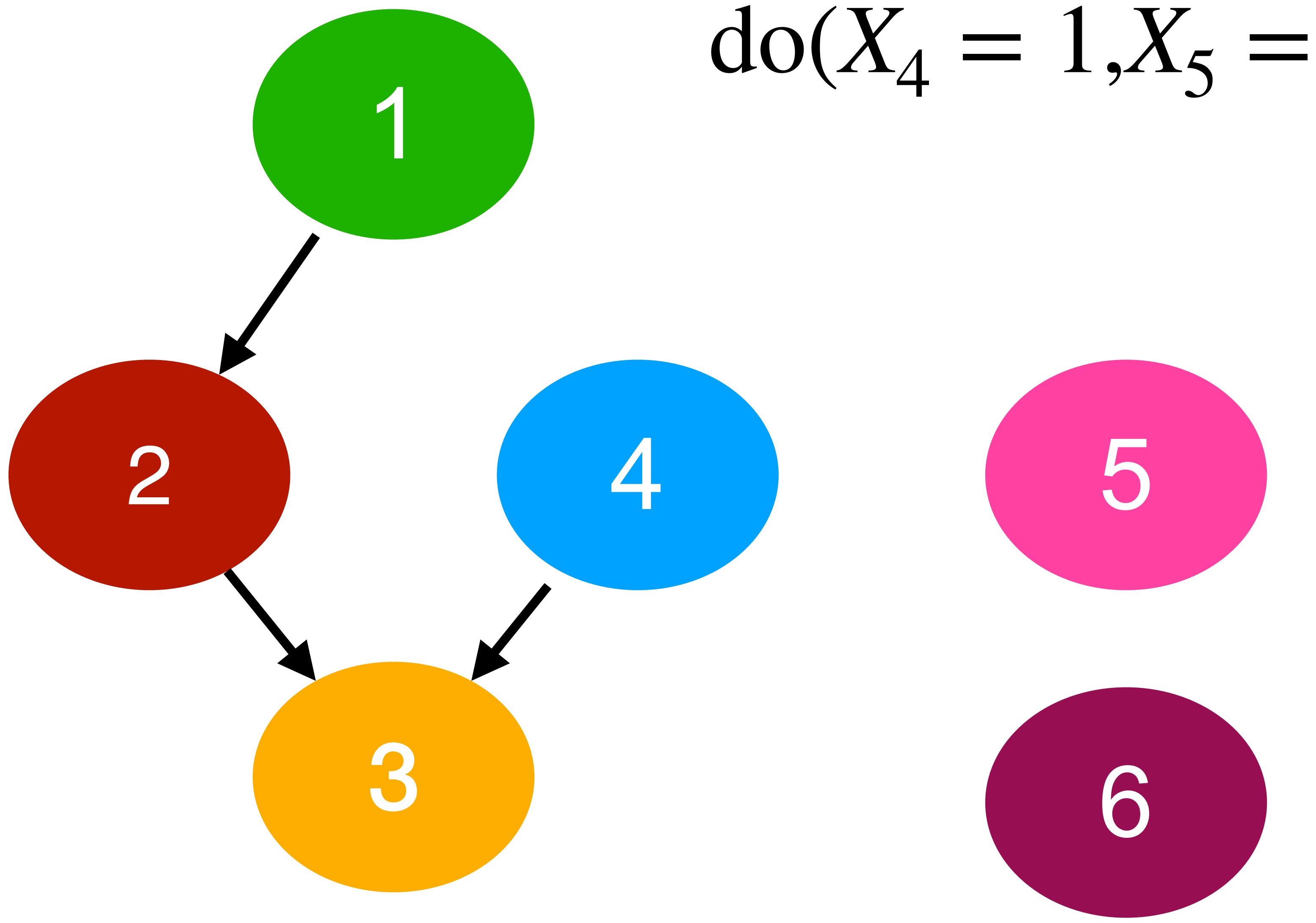
Mutilated/manipulated graphs

Graphically $p(X_{\mathbf{V}} \,|\, \mathrm{do}(X_{\mathbf{W}} = x_{\mathbf{W}})) = \prod_{i \in \bar{V} \setminus \bar{W}} p(X_i \mid X_{pa(i)}) \cdot \mathbb{1}(X_{\bar{W}} = x_{\bar{W}})$

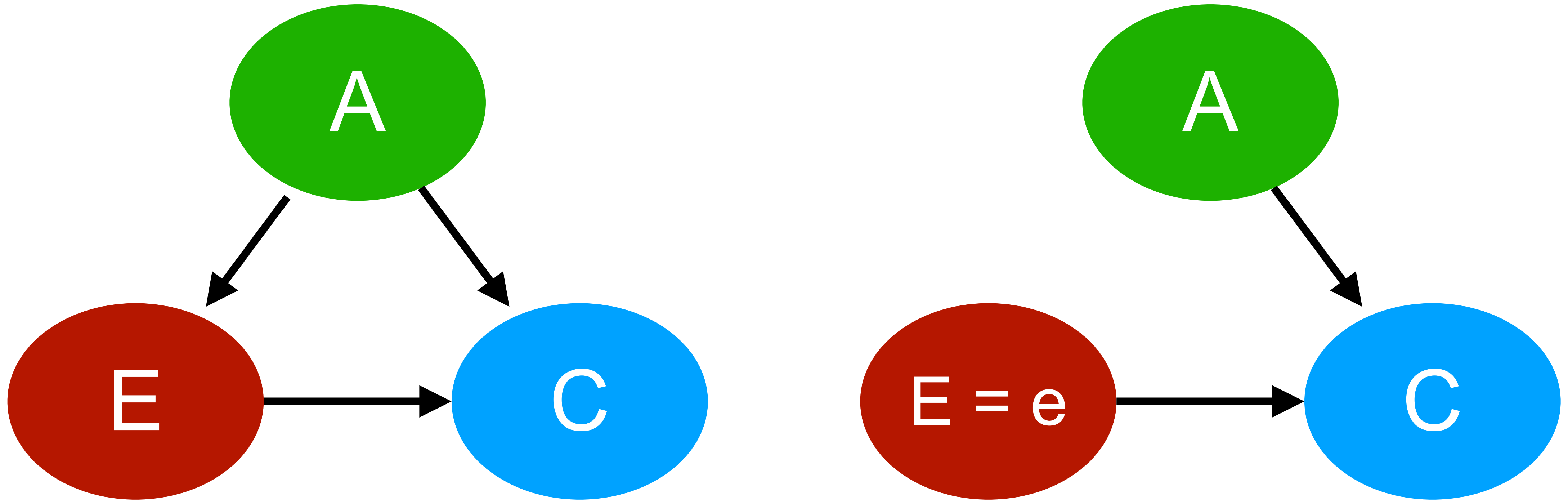
can be represented as cutting the incoming edges to $X_{\mathbf{W}}$

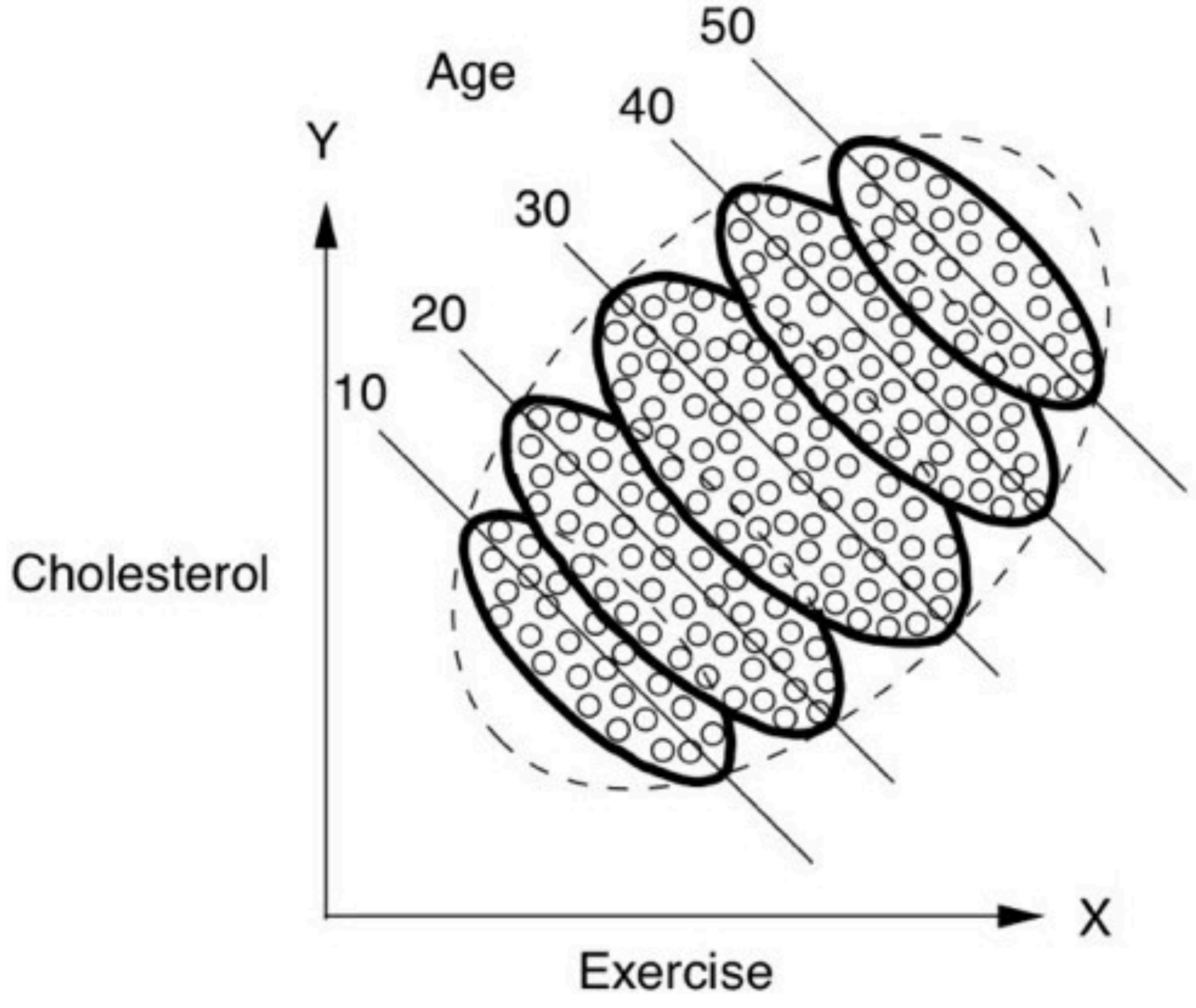


$\mathrm{do}(X_4 = 1, X_5 = 0)$



Simpson paradox examples

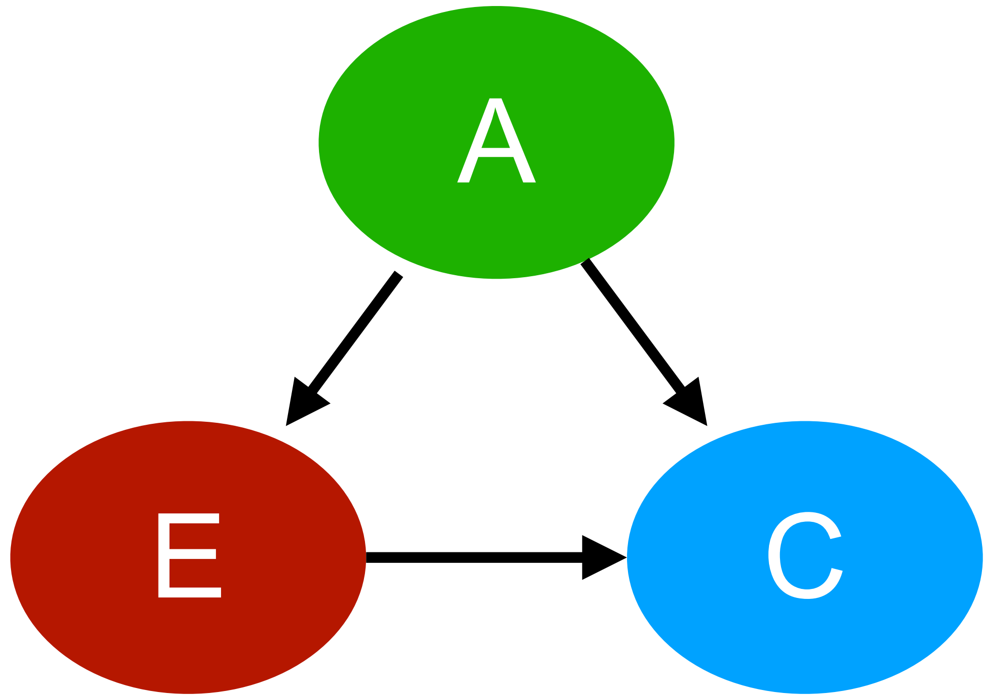


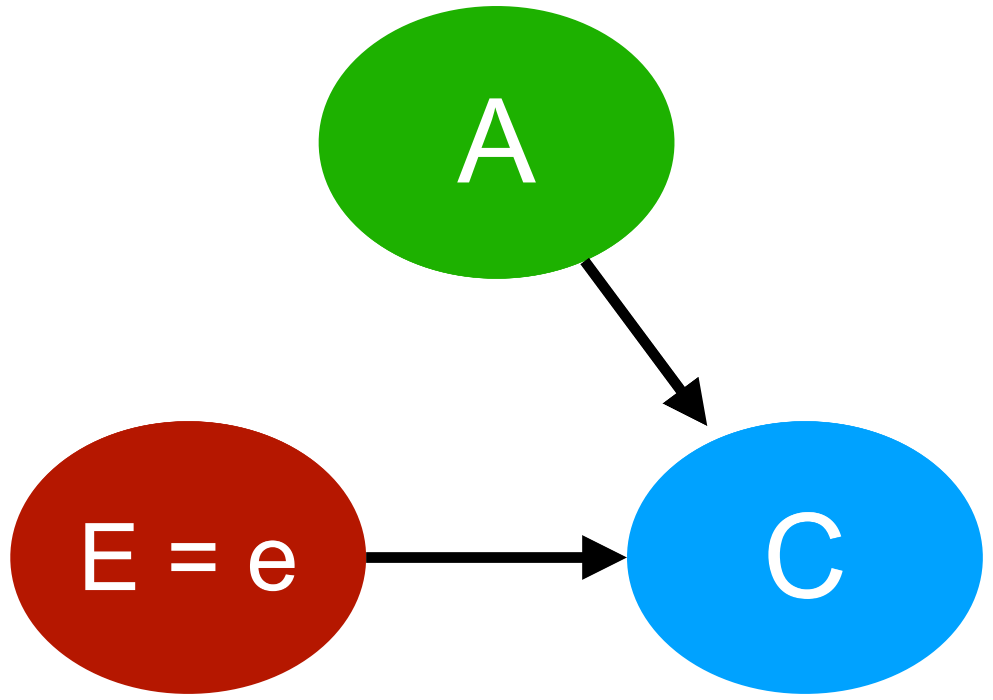


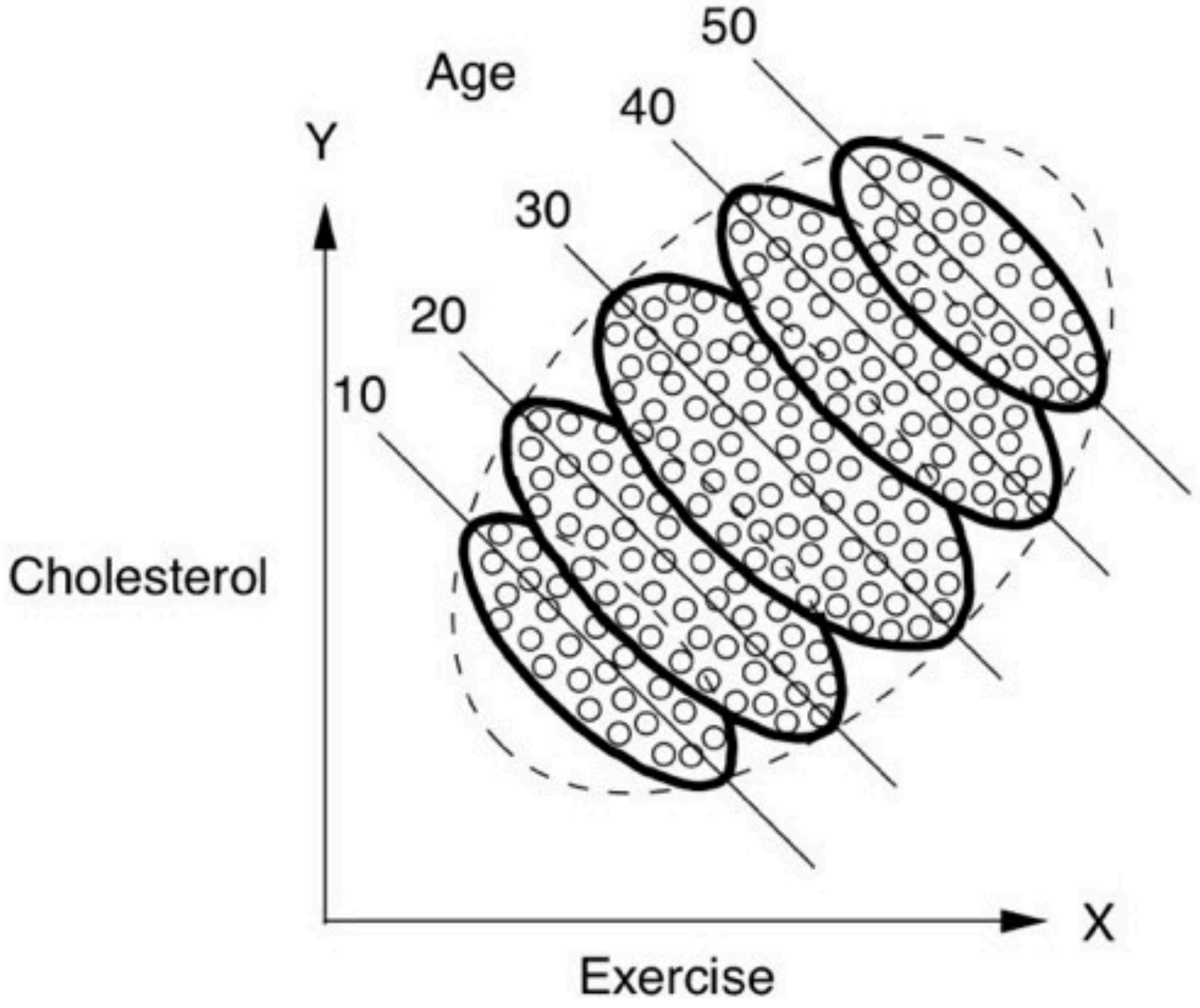
From the Book of Why [Pearl 2018]

$$P(C, A, E \mid do(E=e)) = P(A) \cdot P(C \mid A, E) \cdot \mathbb{1}(E=e)$$

Simpson paradox examples







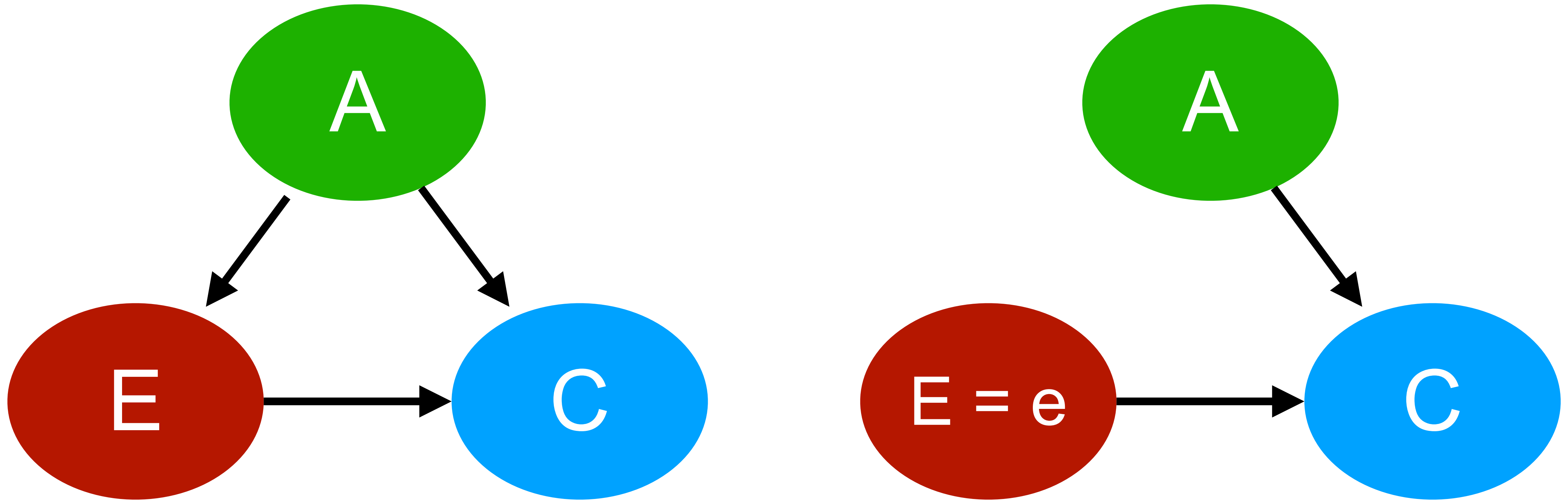
From the Book of Why [Pearl 2018]

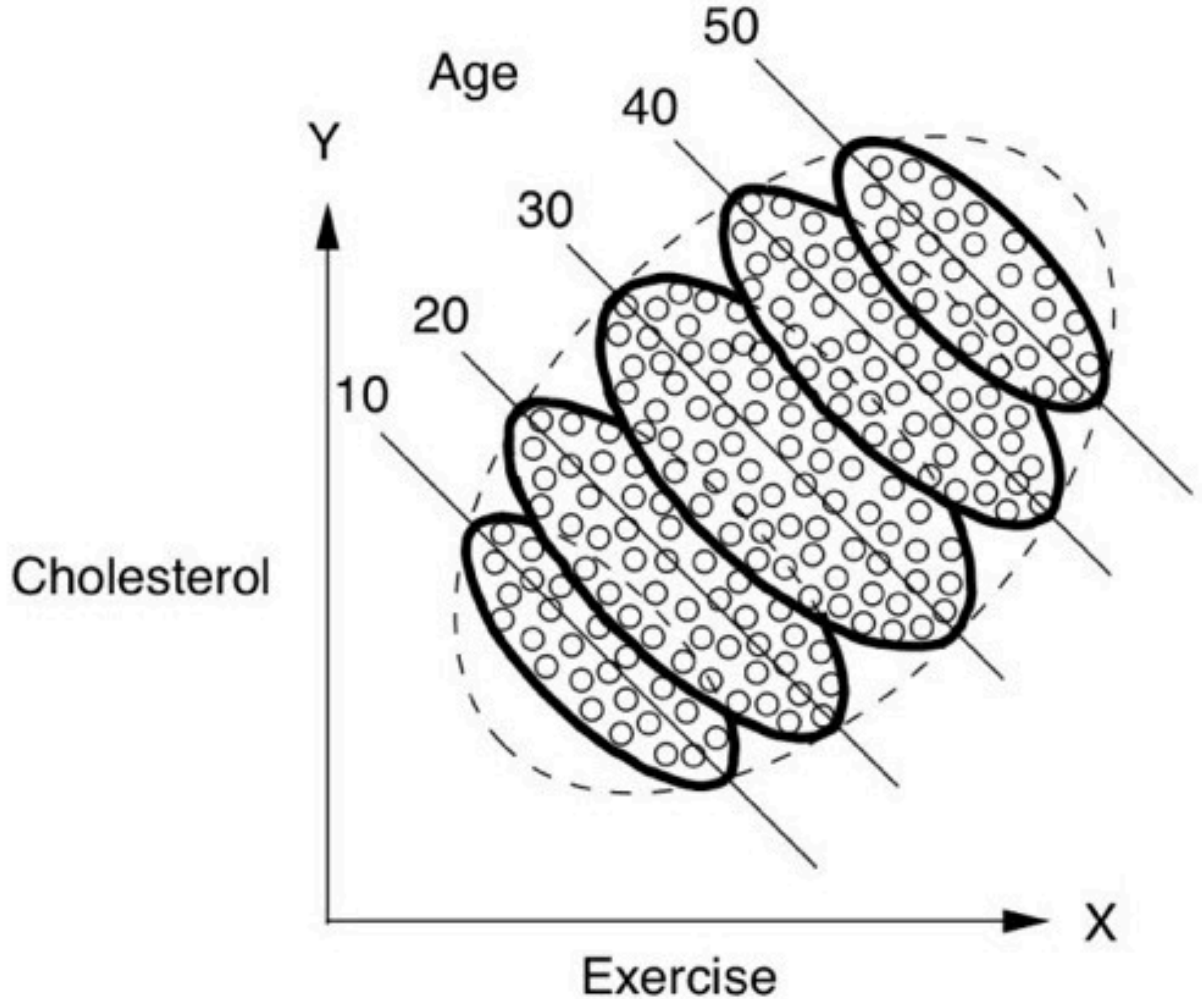
$$P(C, A, E \mid do(E=e)) = P(A) \cdot P(C \mid A, E) \cdot \mathbb{1}(E=e)$$

$$P(C, A \mid do(E=e)) = \sum_{e'} P(A) P(C \mid A, E) \cdot \mathbb{1}(E=e)$$

$$P(X = x) = \sum_{y \in \mathcal{Y}} P(X = x, Y = y)$$

Simpson paradox examples





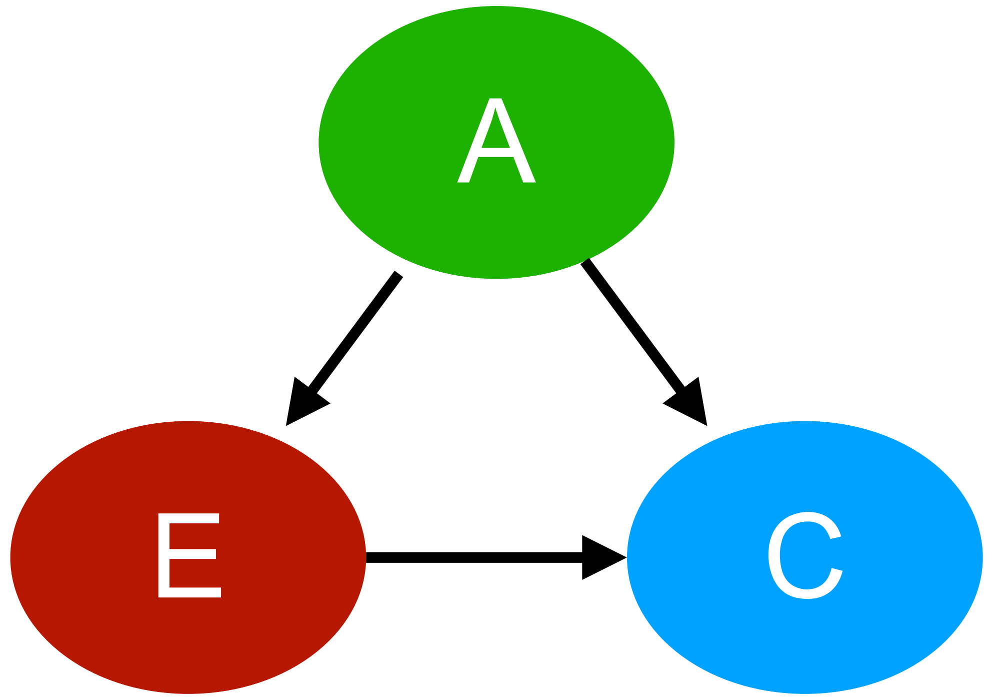
From the Book of Why [Pearl 2018]

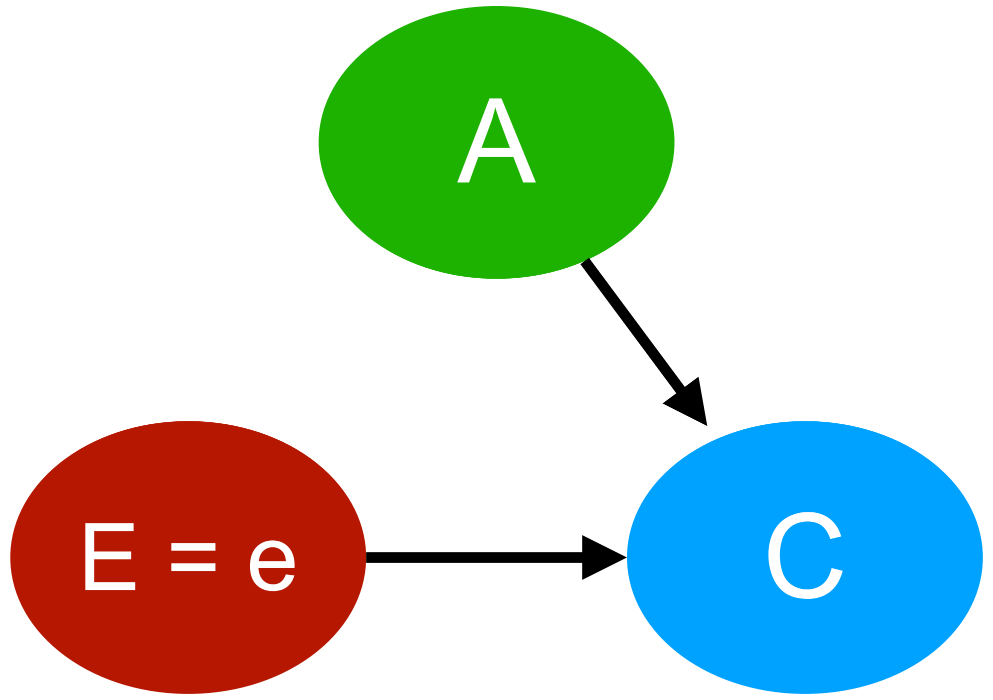
$$P(C,A,E \mid do(E=e)) = P(A) \cdot P(C \mid A,E) \cdot \mathbb{1}(E=e)$$

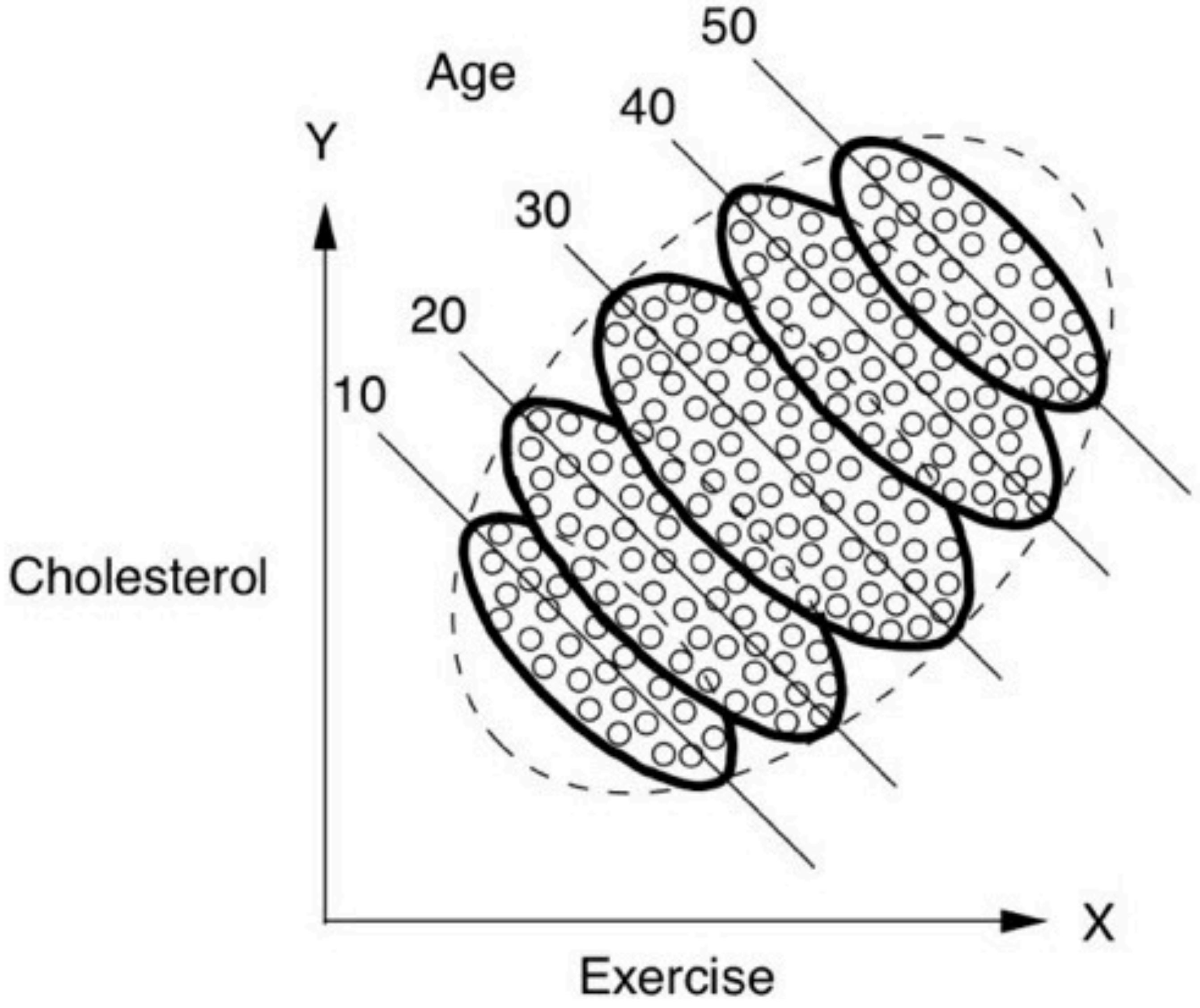
$$P(C,A \mid do(E=e)) = \sum_{e'} P(A)P(C \mid A,E) \cdot \mathbb{1}(E=e) = P(A) \cdot P(C \mid A, E=e)$$

Intuition: substitute E=e in the factors in which E appears

Simpson paradox examples







From the Book of Why [Pearl 2018]

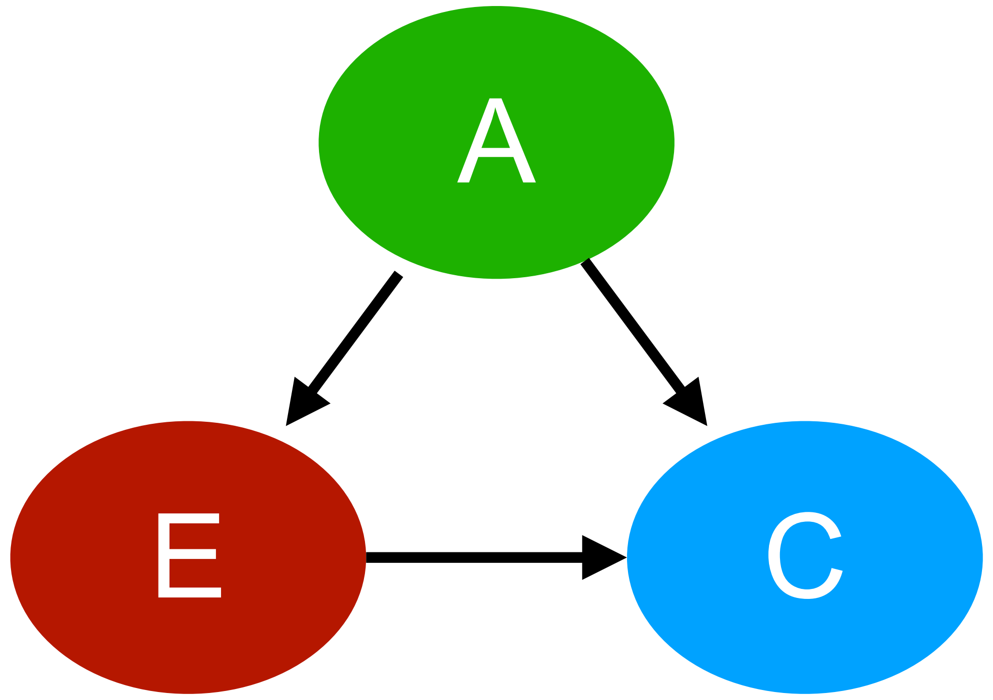
$$P(C, A, E \mid do(E=e)) = P(A) \cdot P(C \mid A, E) \cdot \mathbb{1}(E=e)$$

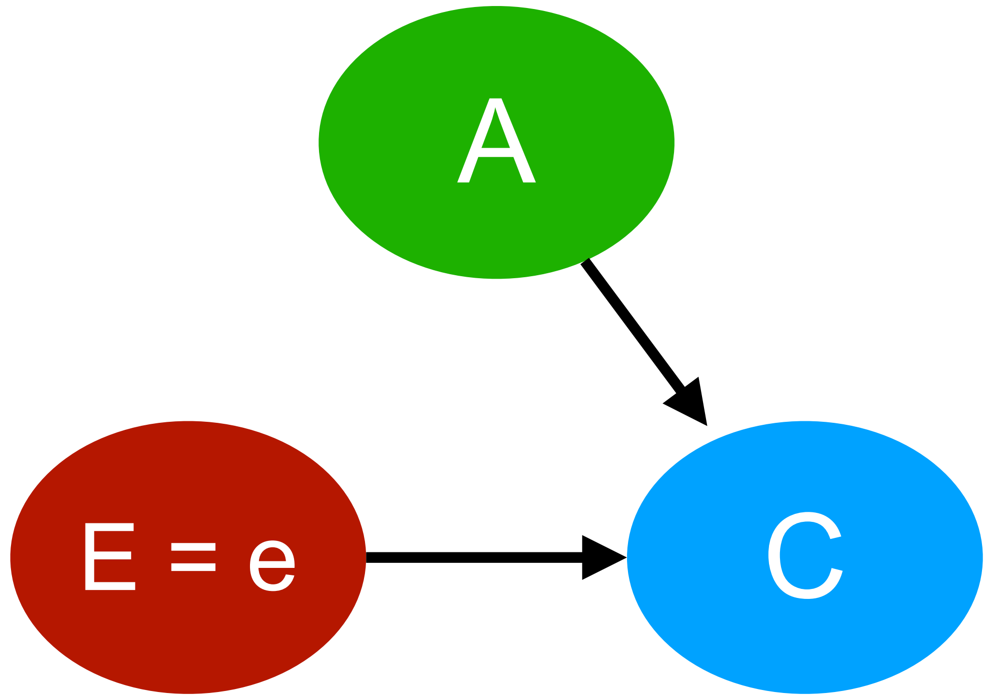
$$P(C, A \mid do(E=e)) = \sum_{e'} P(A) P(C \mid A, E) \cdot \mathbb{1}(E=e) = P(A) \cdot P(C \mid A, E=e)$$

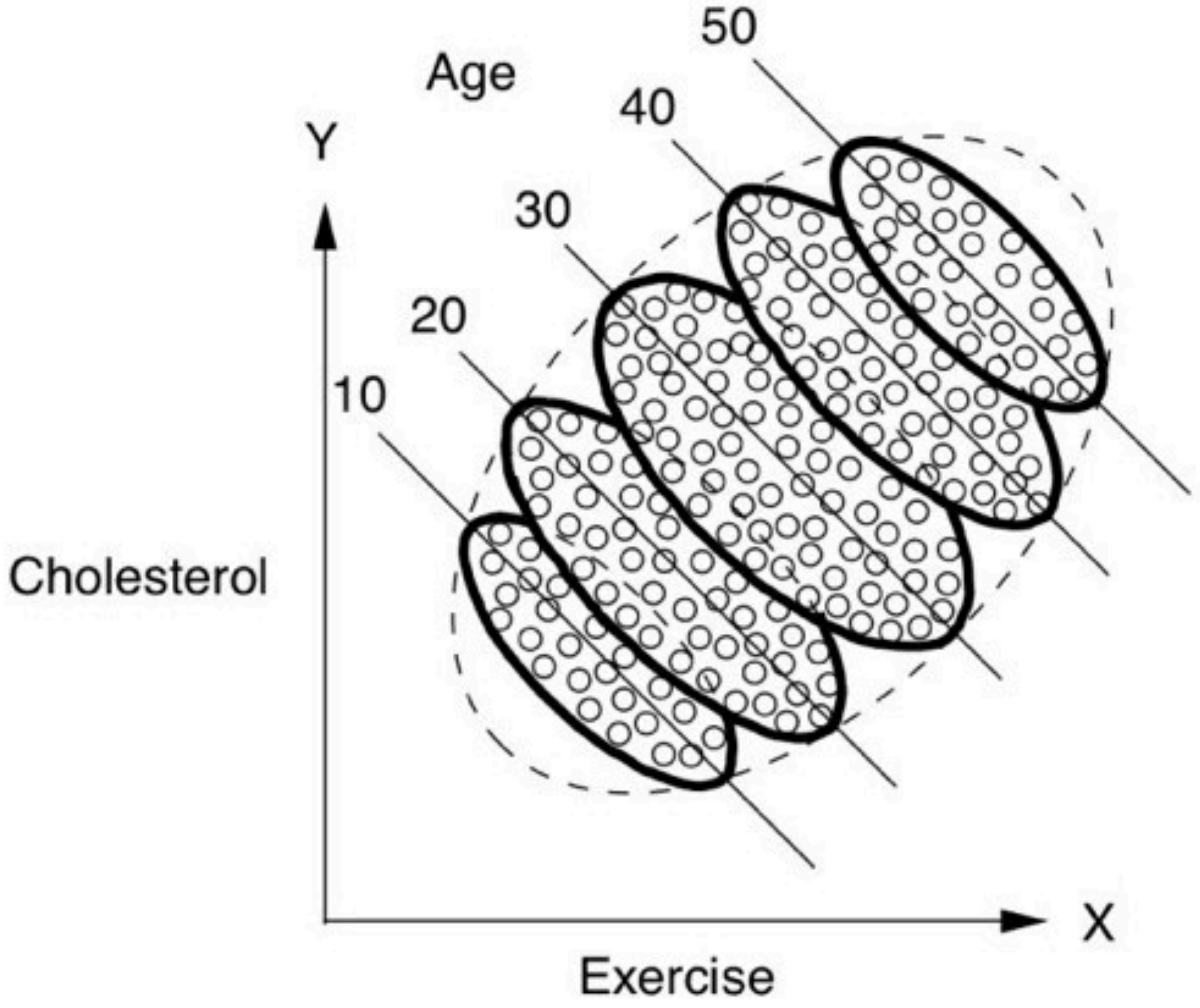
$$P(C \mid do(E=e)) = \sum_{a} P(A=a) \cdot P(C \mid A=a, E=e)$$

$$P(X = x) = \sum_{y \in \mathcal{Y}} P(X = x, Y = y)$$

Simpson paradox examples







From the Book of Why [Pearl 2018]

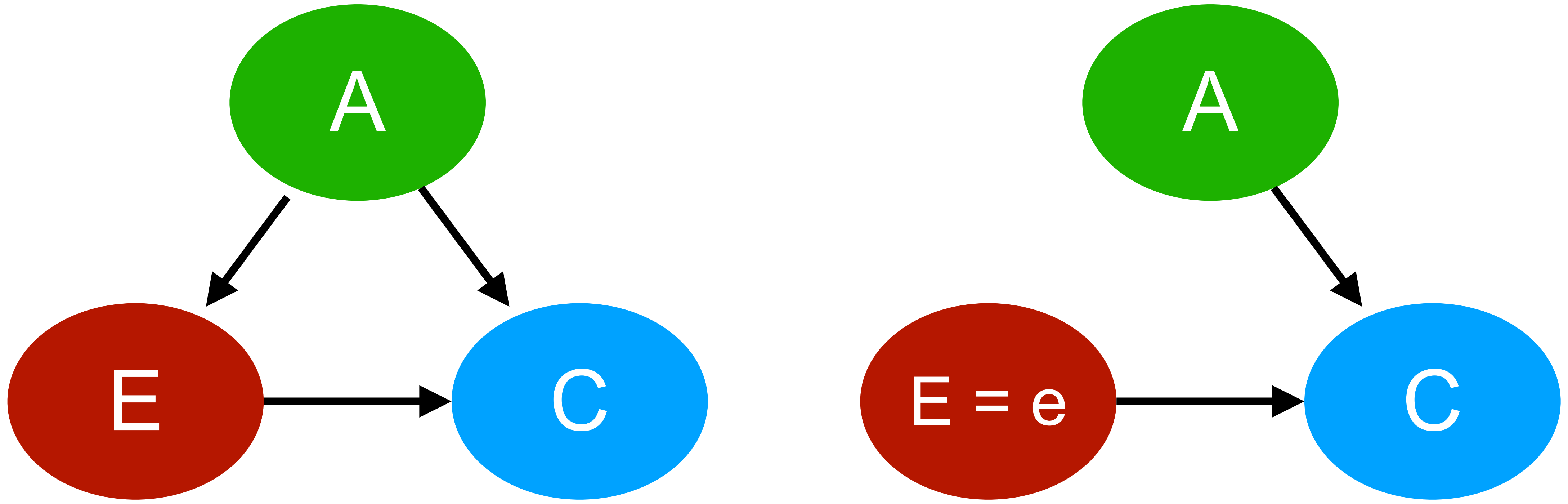
$$P(C,A,E \mid do(E=e)) = P(A)\cdot P(C \mid A,E)\cdot \mathbb{1}(E=e)$$

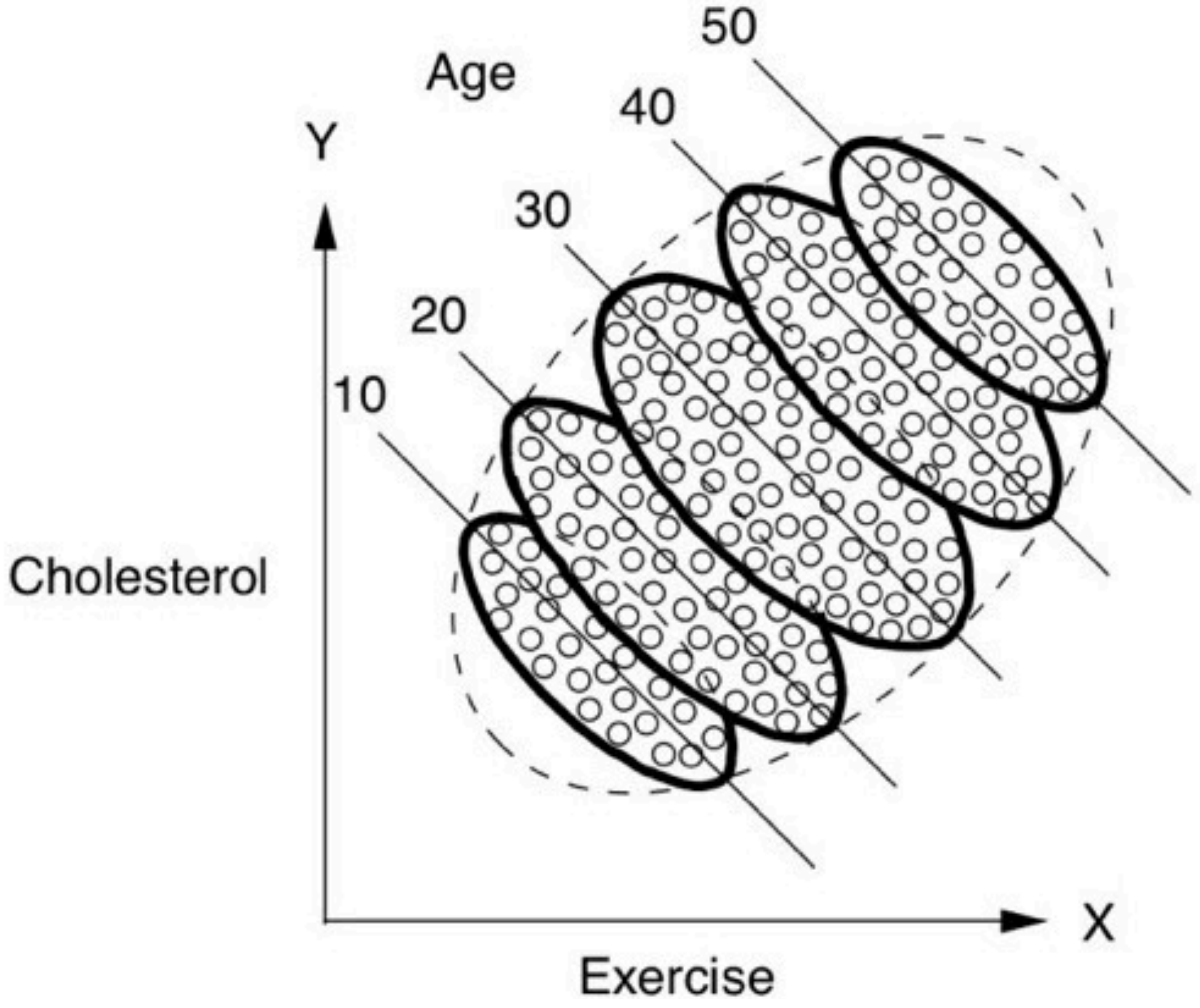
$$P(C,A \mid do(E=e)) = \sum_{e'} P(A)P(C \mid A,E)\cdot \mathbb{1}(E=e) = P(A)\cdot P(C \mid A,E=e)$$

$$P(C \mid do(E=e)) = \sum_{a} P(A=a)\cdot P(C \mid A=a,E=e)$$

all observational distributions!

Simpson paradox examples



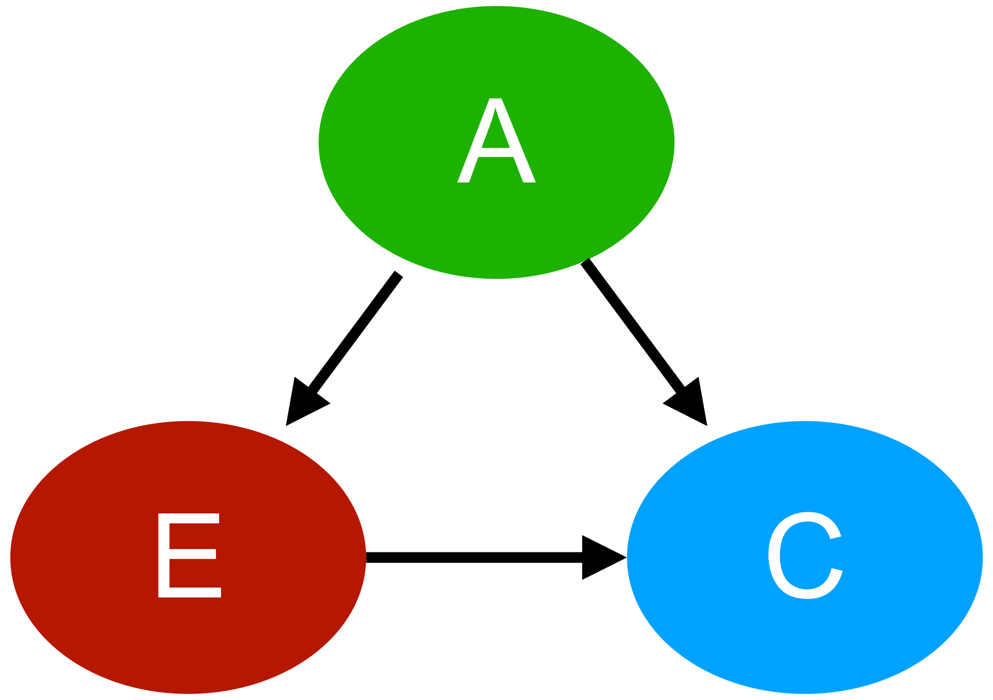


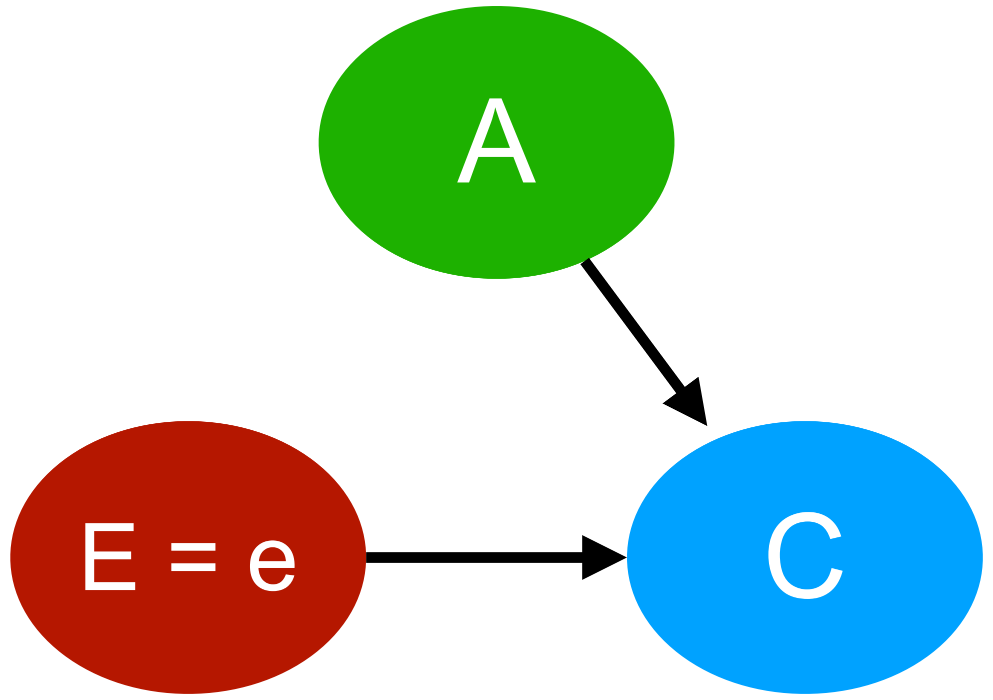
From the Book of Why [Pearl 2018]

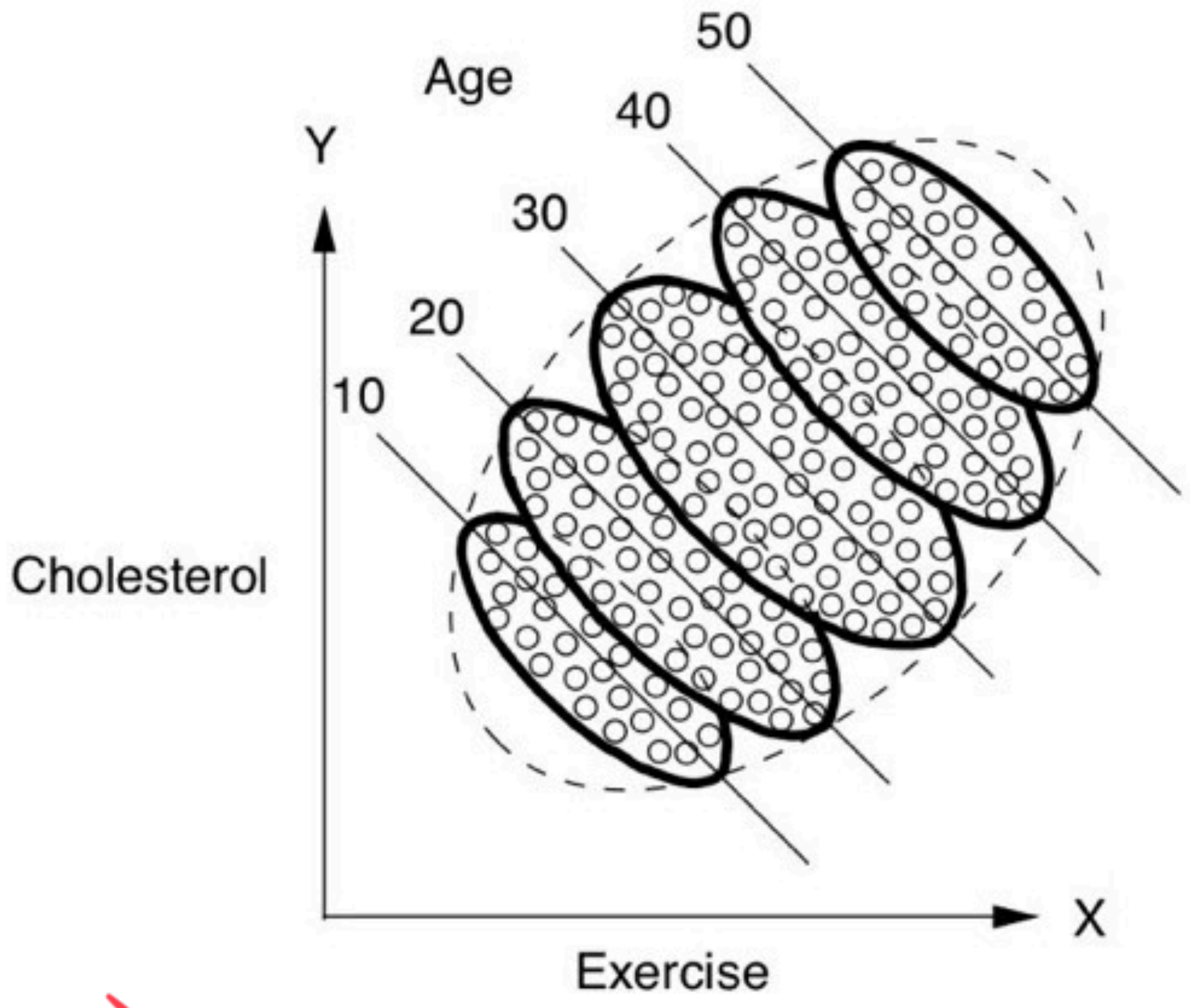
$$P(C, A \mid E) = \frac{P(A) \cdot P(E \mid A) \cdot P(C \mid A, E)}{P(E)}$$

$$P(C \mid do(E=e)) = \sum_a P(A=a) \cdot P(C \mid A=a, E=e)$$

Simpson paradox examples





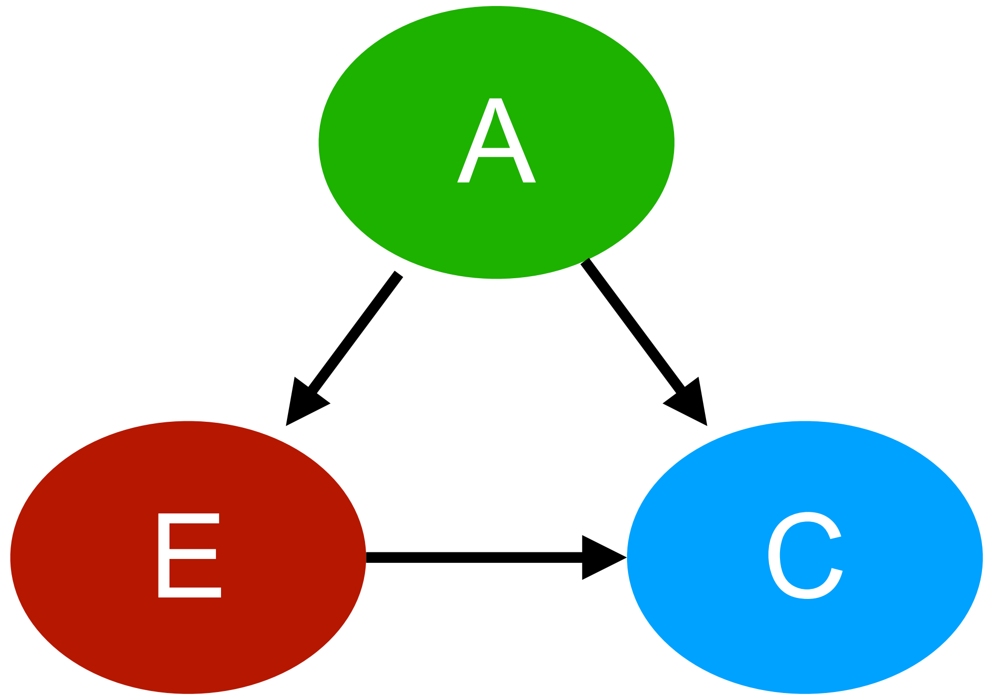


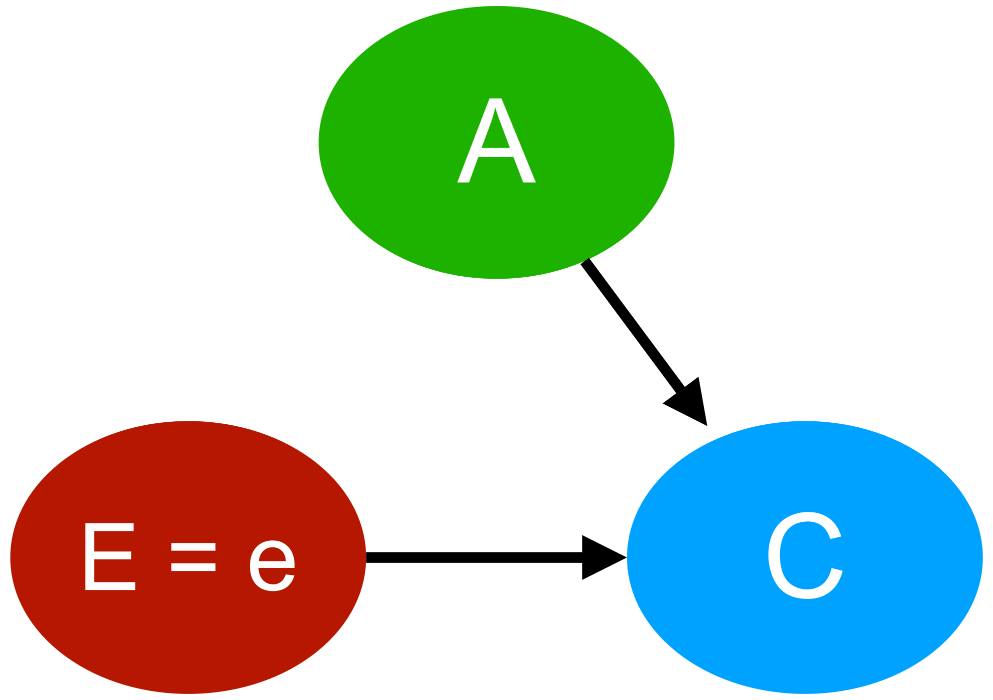
From the Book of Why [Pearl 2018]

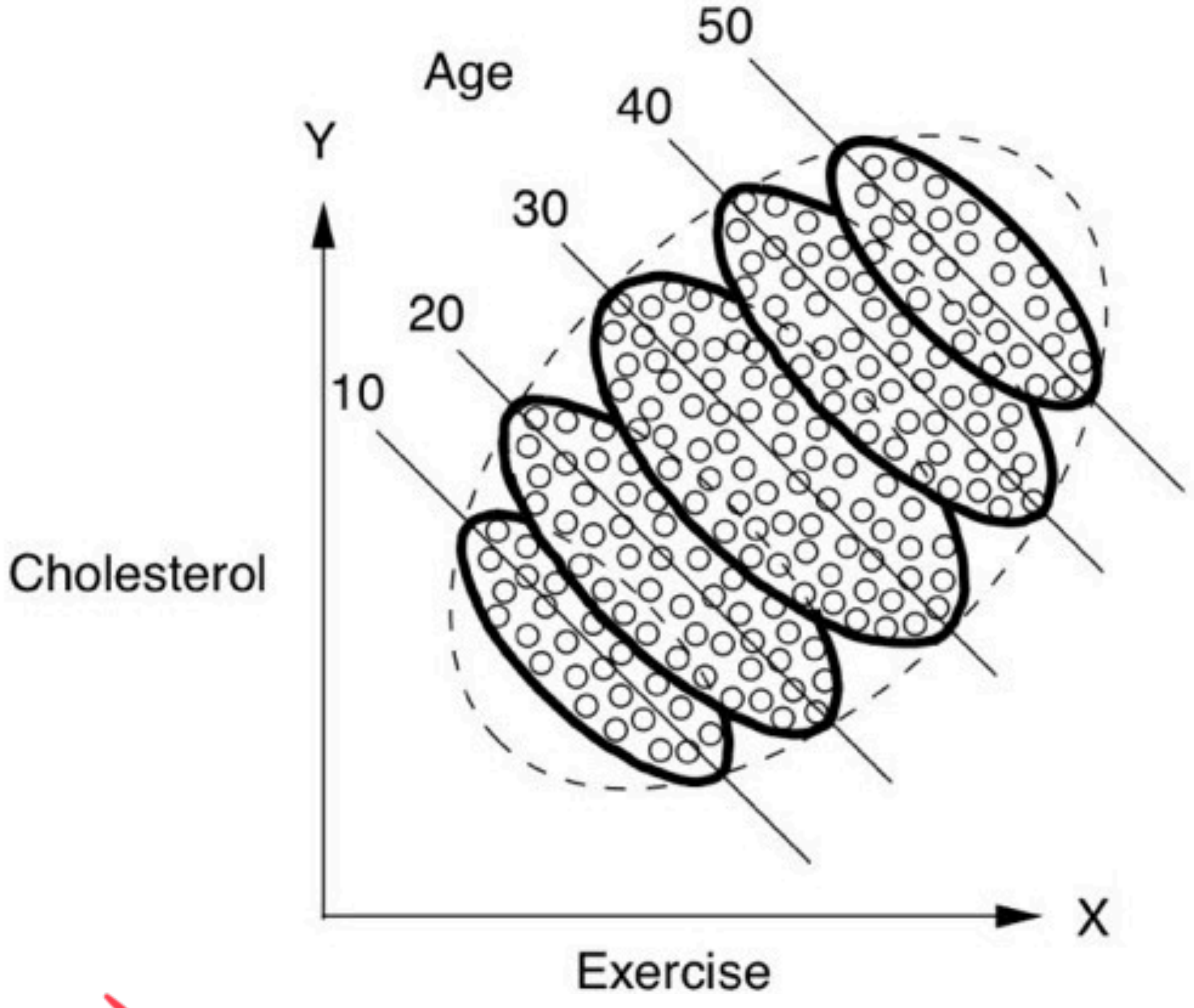
$$P(C, A \mid E) = \frac{P(A) \cdot P(E \mid A) \cdot P(C \mid A, E)}{P(E)} = \frac{\cancel{P(E)} \cdot P(A \mid E) \cdot P(C \mid A, E)}{\cancel{P(E)}}$$

$$P(C \mid do(E=e)) = \sum_a P(A=a) \cdot P(C \mid A=a, E=e)$$

Simpson paradox examples







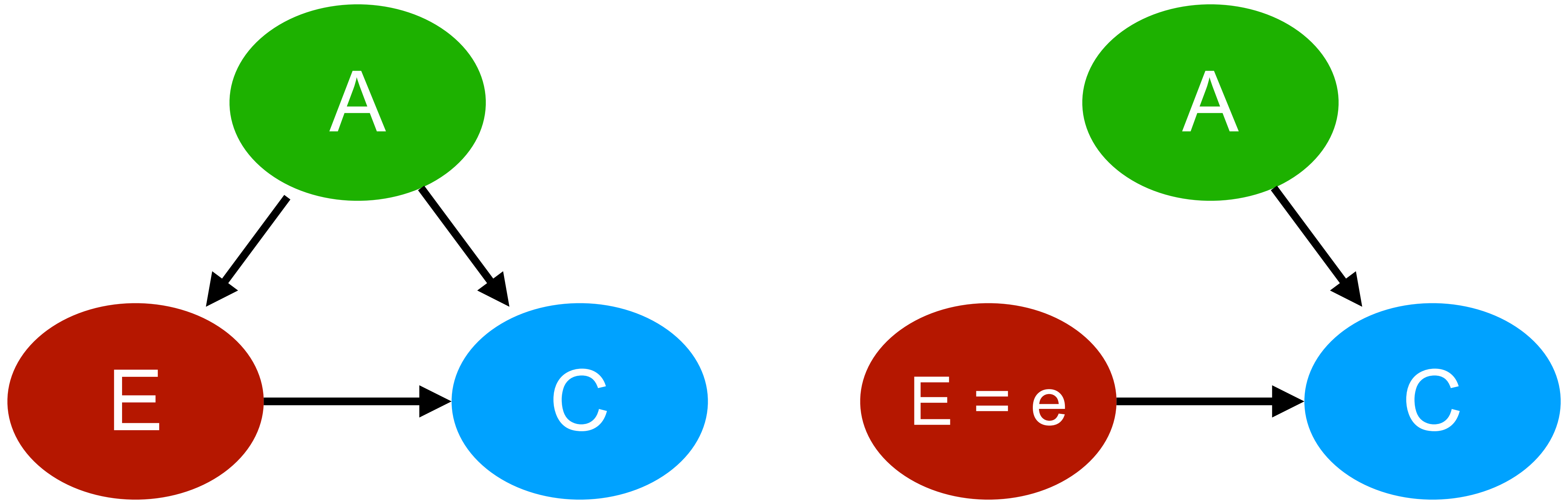
From the Book of Why [Pearl 2018]

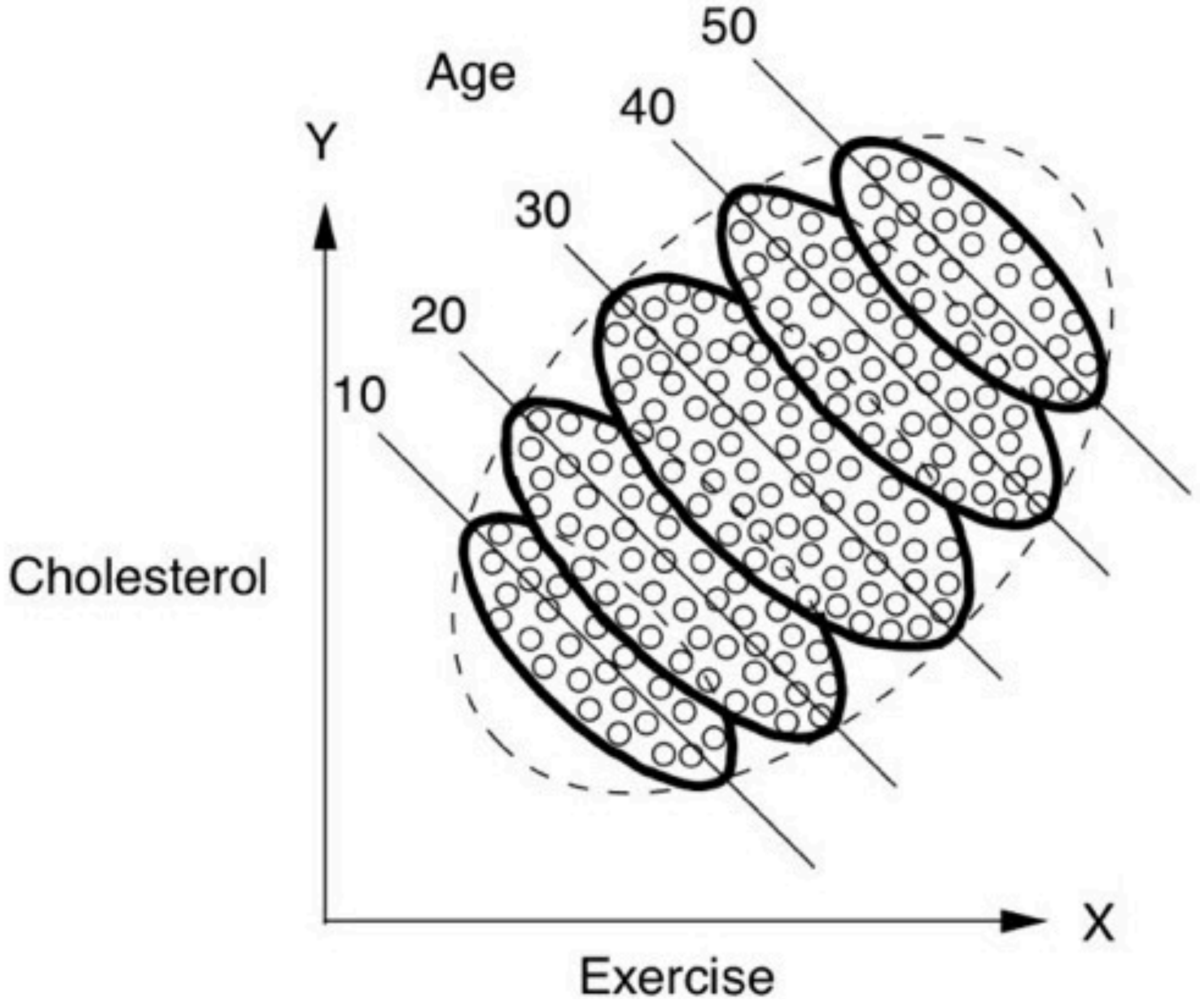
$$P(C, A \mid E) = \frac{P(A) \cdot P(E \mid A) \cdot P(C \mid A, E)}{P(E)} = \frac{\cancel{P(E)} \cdot P(A \mid E) \cdot P(C \mid A, E)}{\cancel{P(E)}}$$

$$P(C \mid E=e) = \sum_a P(A=a \mid E=e) \cdot P(C \mid A=a, E=e)$$

$$P(C \mid do(E=e)) = \sum_a P(A=a) \cdot P(C \mid A=a, E=e)$$

Simpson paradox examples





From the Book of Why [Pearl 2018]

$$P(C|E=e) = \sum_a P(A=a|E=e) \cdot P(C|A=a,E=e)$$

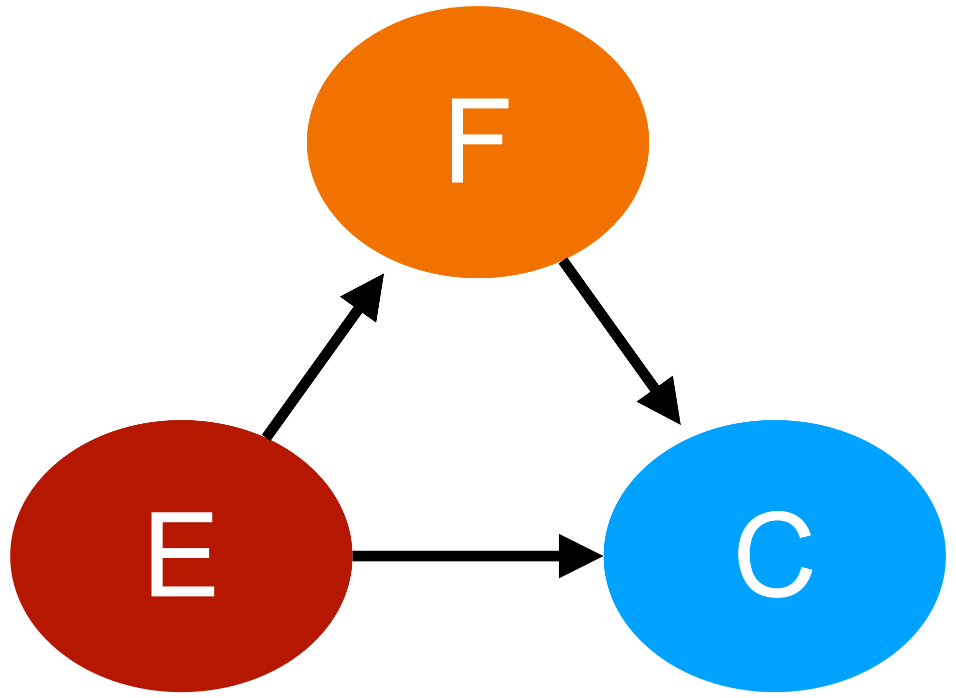
$$P(C|do(E=e)) = \sum_a P(A=a) \cdot P(C|A=a,E=e)$$

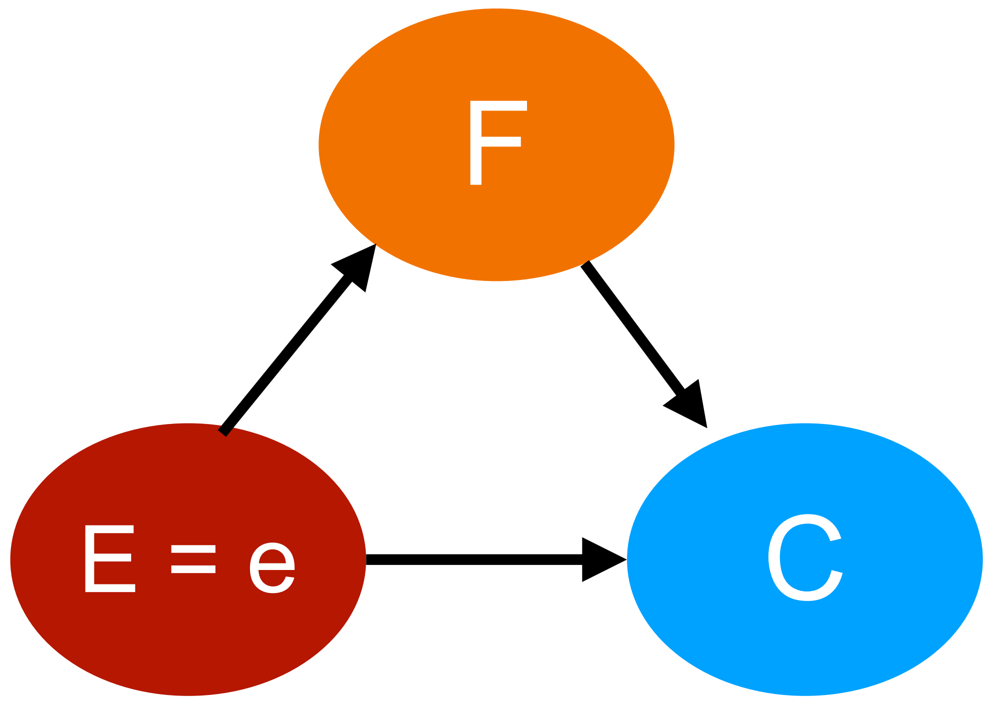
$P(A=a|E=e)$

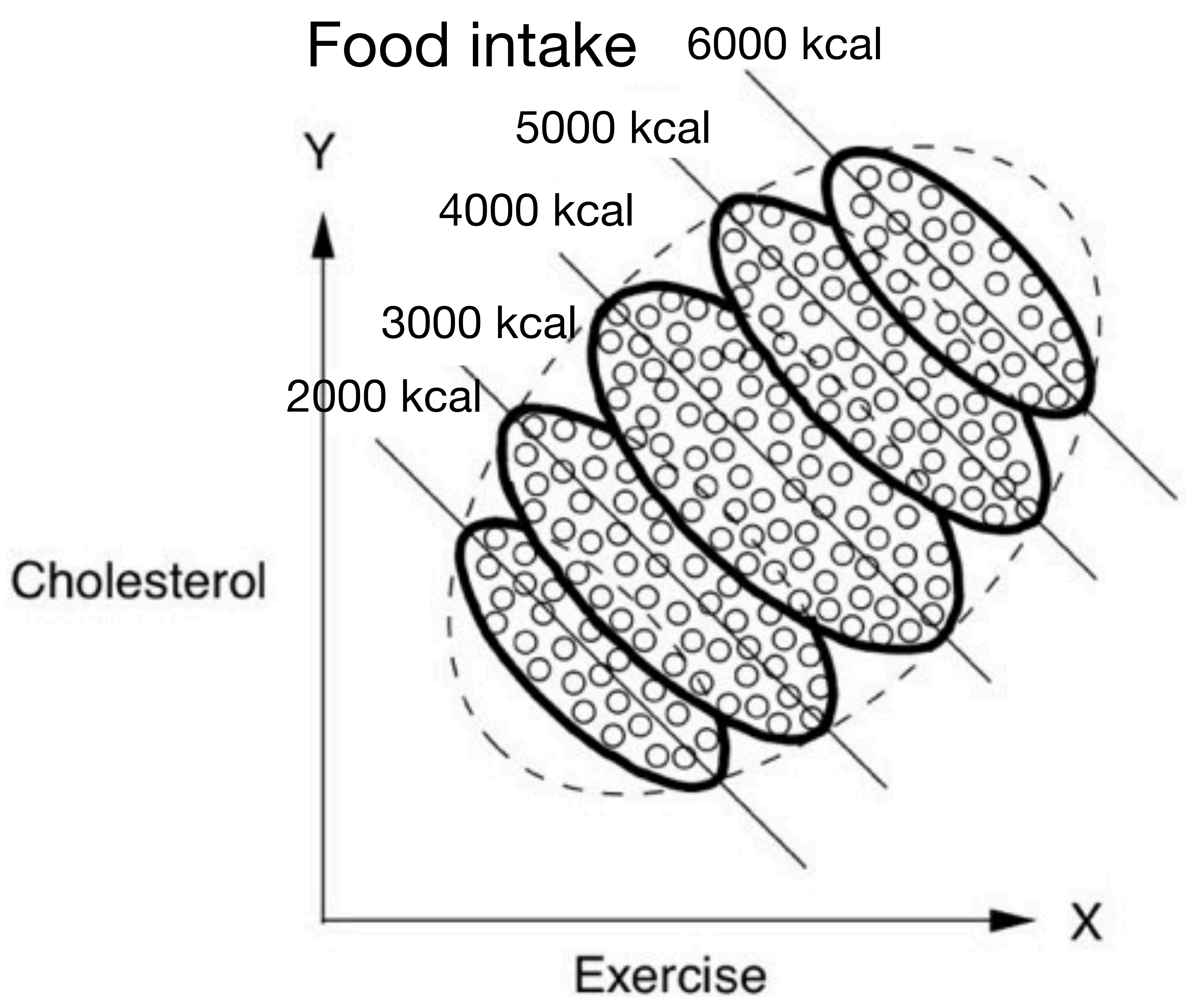
vs.

$P(A=a)$

Simpson paradox examples





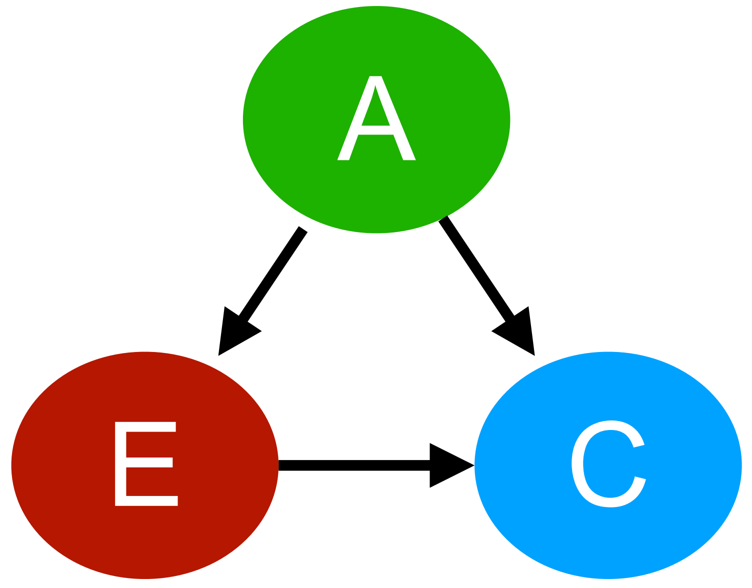


$$P(C \mid do(E=e)) = \sum_f P(F=f \mid E=e) \cdot P(C \mid F=f, E=e)$$

$$=$$

$$P(C \mid E=e) = \sum_f P(F=f \mid E=e)\, P(C \mid F=f, E=e)$$

Confounders and mediators

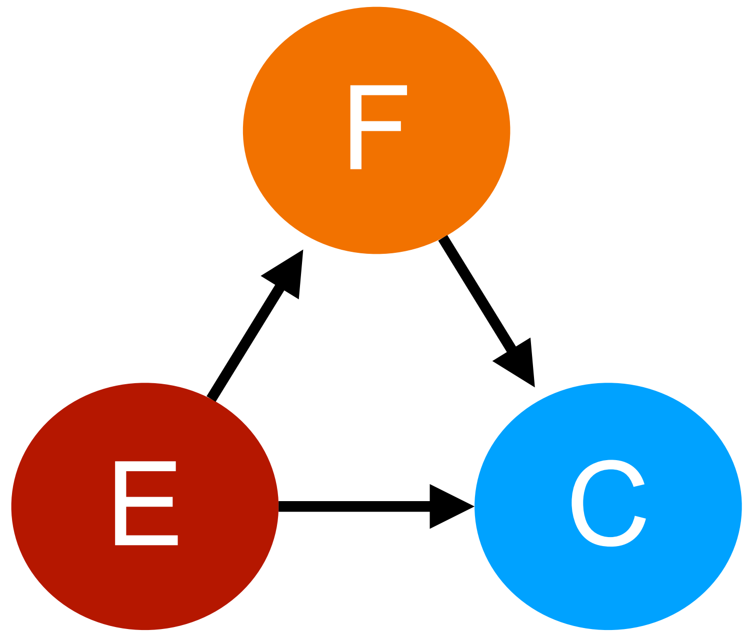


- The effect of X on Y is **confounded** whenever

$$P(Y \mid \mathrm{do}(X = x)) \neq P(Y \mid X = x)$$

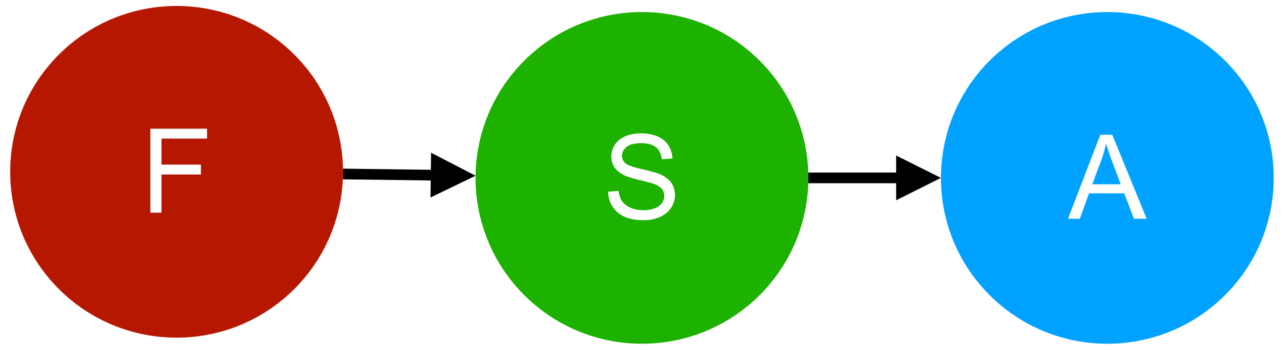
- A **confounder** is a variable k that has a **directed path to both** X **and** Y

- **A mediator** is a variable that is:
  - a descendant of X and
  - an ancestor of Y



Exercise in Canvas

Observational vs (post) interventional distributions



$P(F=1) = 0.01$

$P(S=1 \mid F=1) = 0.9$

$P(S=1 \mid F=0) = 0.2$

$P(A=1 \mid S=1) = 0.8$

$P(A=1 \mid S=0) = 0.1$