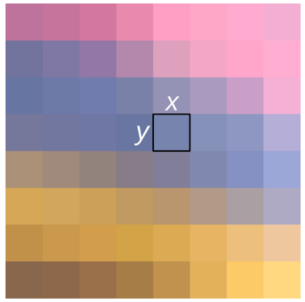
Causality for Fairness and Explainability

Karima Makhlouf
Ruta Binkyte-Sadauskiene
Carlos Pinzon
Catuscia Palamidessi
Sami Zhioua

Training data Y

Machine Learning (ML)



Input features

X

Gender
Race
Age
.....

A

Automatic decision system

Cat

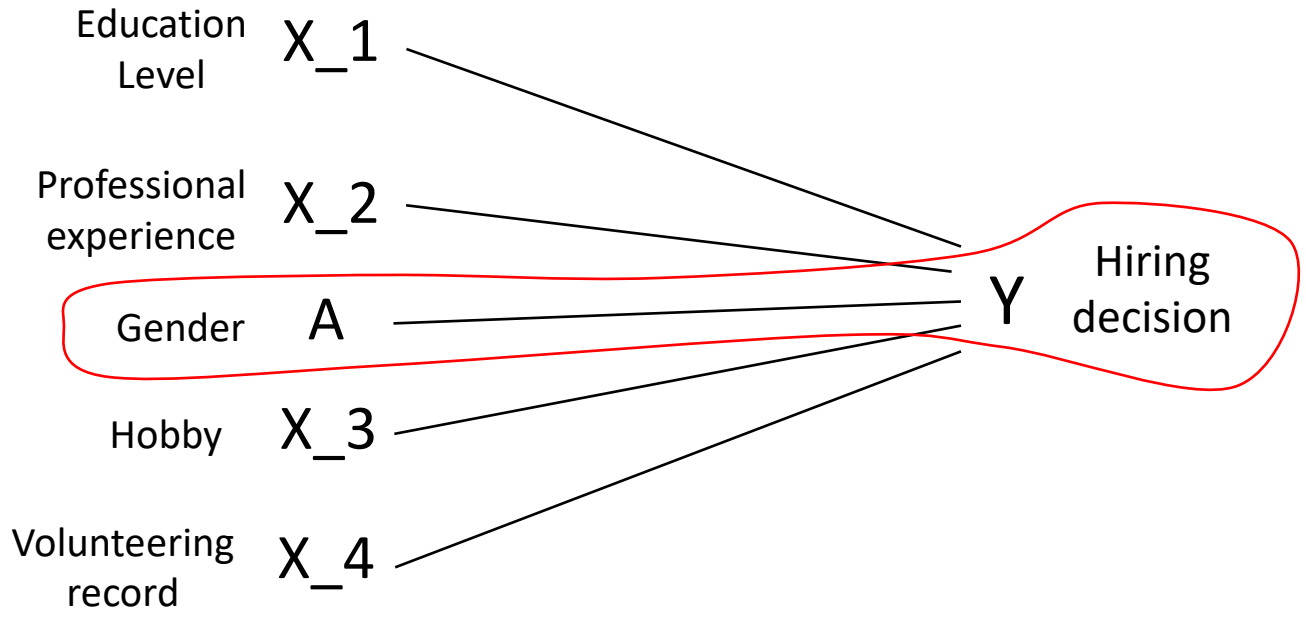
Output Y Ŷ

Job hiring
College admission
Recidivism
.....

Fairness: Is the output fair with respect to individuals or subpopulations ?

Explainability: How the output can be explained in terms of the input features ?

Statistical (Observational) notions of fairness



$$P(\hat{Y} \mid A = 0) = P(\hat{Y} \mid A = 1)$$

Statistical Parity

$$P(\hat{Y} = 1 \mid E = e, A = 0) = P(\hat{Y} = 1 \mid E = e, A = 1) \quad \forall e$$

Conditional Statistical Parity

$$P(\hat{Y} = 1 \mid Y = 1, A = 0) = P(\hat{Y} = 1 \mid Y = 1, A = 1)$$

Equal Opportunity

$$E[S \mid Y = 1, A = 0)] = E[S \mid Y = 1, A = 1]$$

Balance

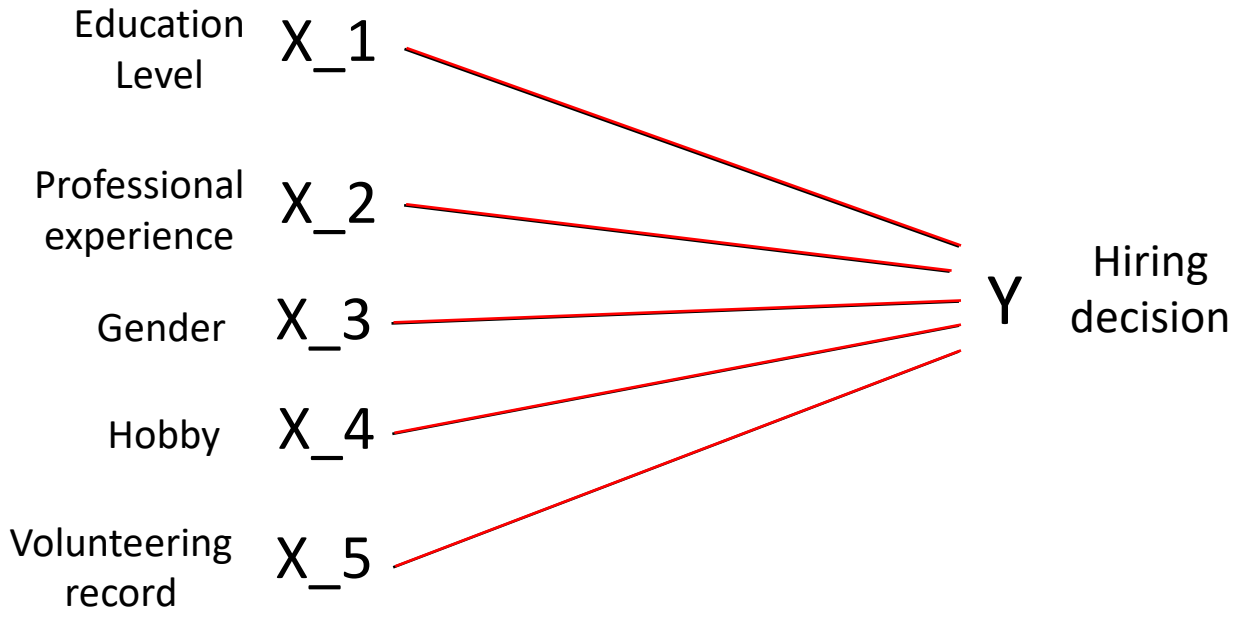
$$P(Y = 1 \mid \hat{Y} = 1, A = 0) = P(Y = 1 \mid \hat{Y} = 1, A = 1)$$

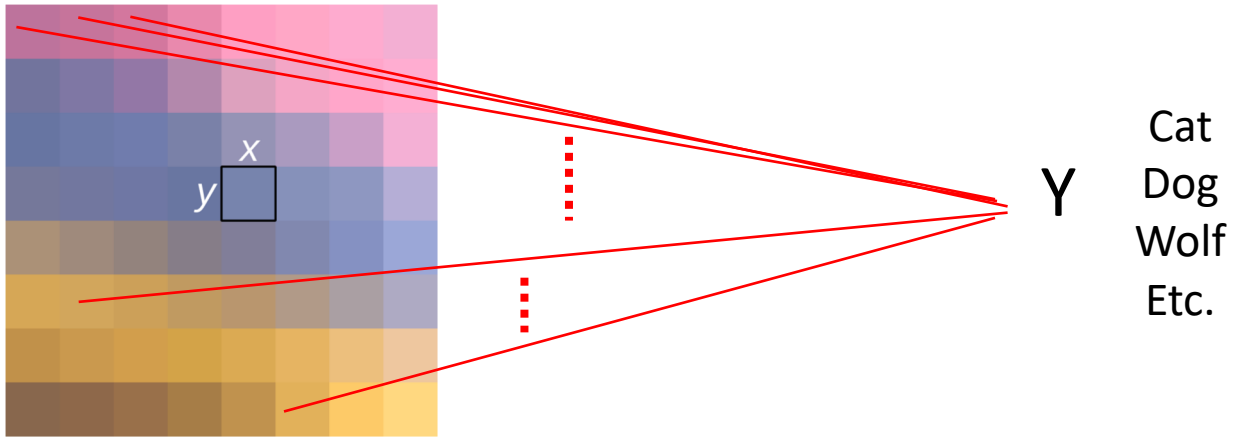
Predictive Parity

$$P(Y = 1 \mid S = s, A = 0) = P(Y = 1 \mid S = s, A = 1) \quad \forall s \in [0, 1]$$

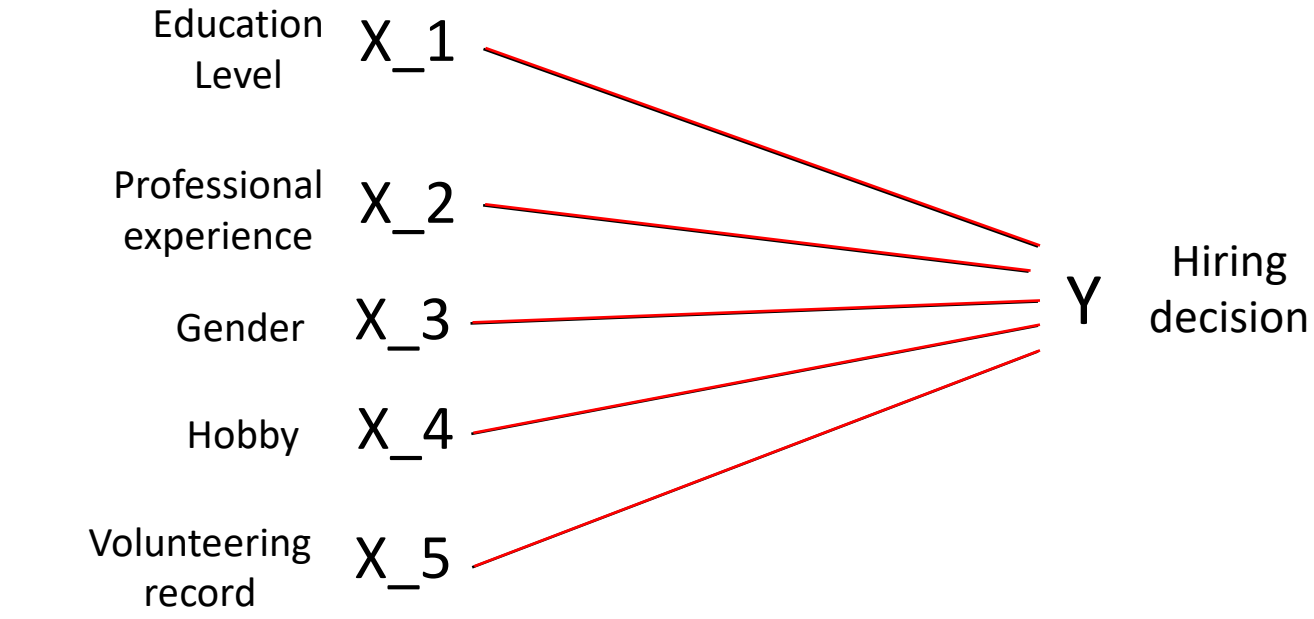
Calibration

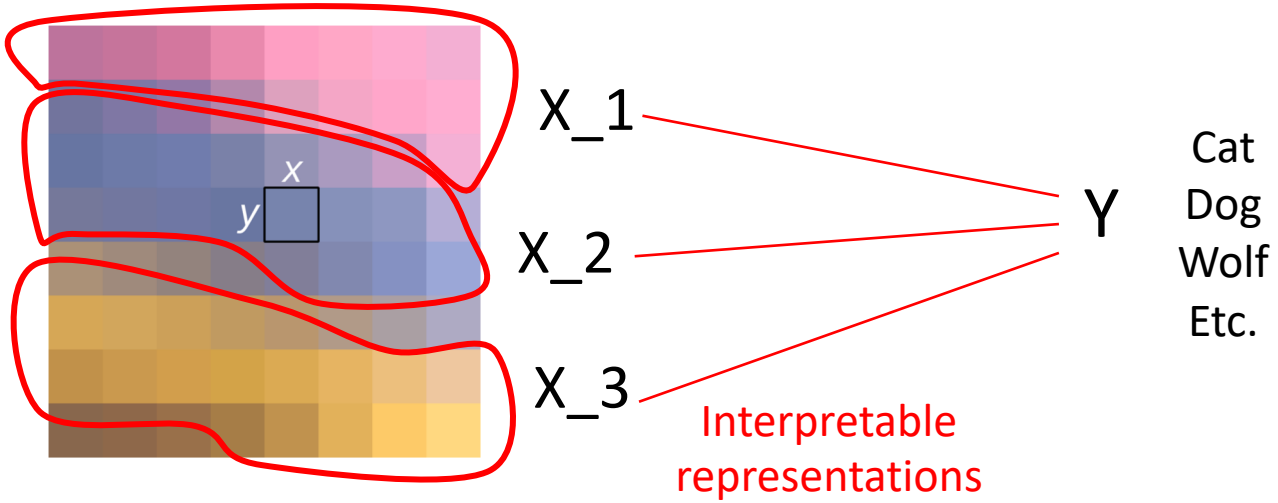
Purely associational explainability



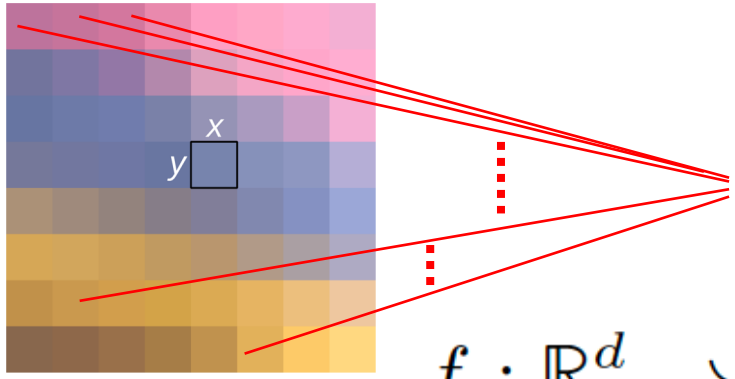


Purely associational explainability





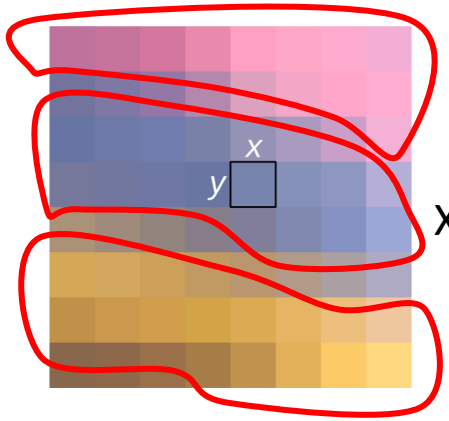
Purely associational explainability (LIME)*



Y
Cat
Dog
Wolf
Etc.

$$f : \mathbb{R}^d \rightarrow \mathbb{R}$$

(Model to be explained)



X_1

X_2

X_3

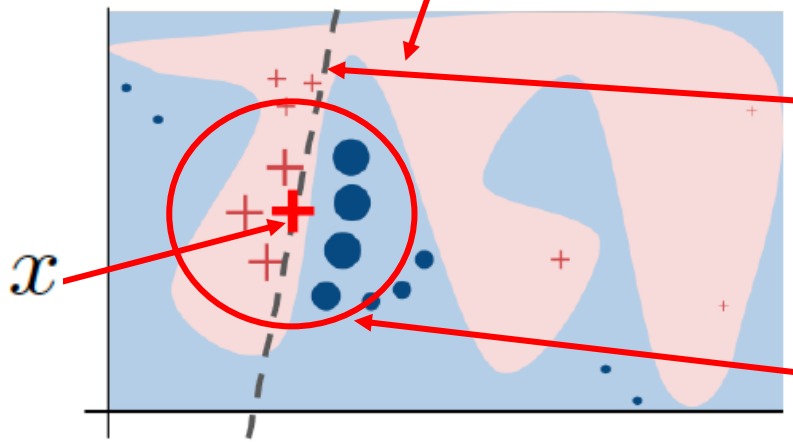
Y
Cat
Dog
Wolf
Etc.

$$g : \{0,1\}^{d'} \rightarrow \mathbb{R}$$

(Explanation)

$\Omega(g)$

(Measure of complexity)



x

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \; \mathcal{L}(f, g, \pi_x) + \Omega(g)$$

$$g(z') = w_g \cdot z'$$

π_x

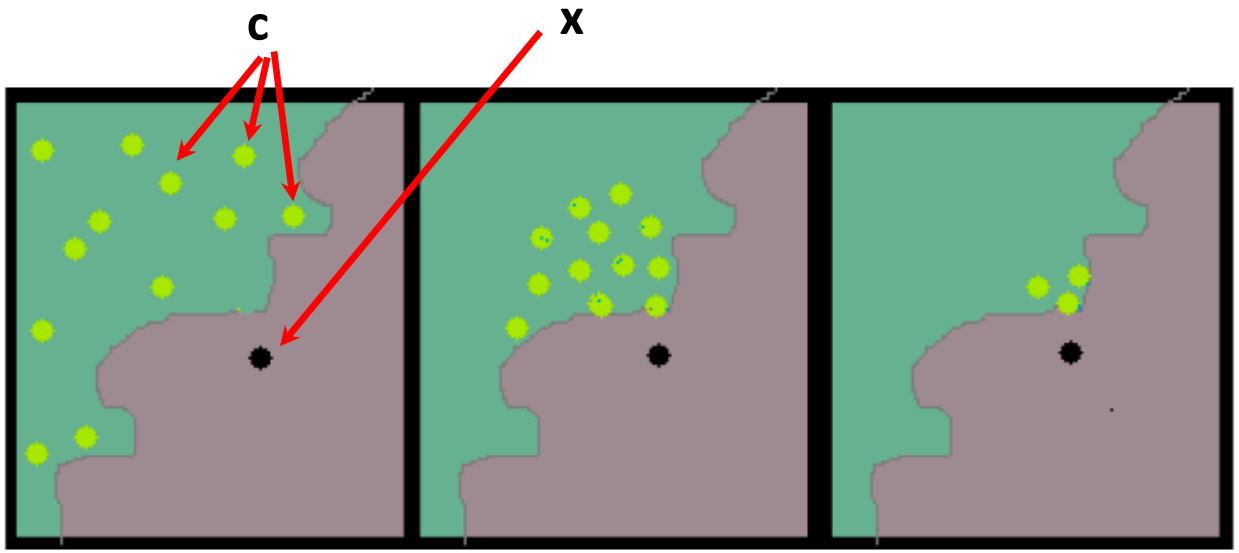
(Proximity measure)

(a measure of how unfaithful g is in approximating f)

* Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. *22nd ACM SIGKDD.*

Purely associational explainability (Counterfactual)*

A ***counterfactual*** is a generated data point that is as close to the input data point as possible for which the model gives a different outcome.

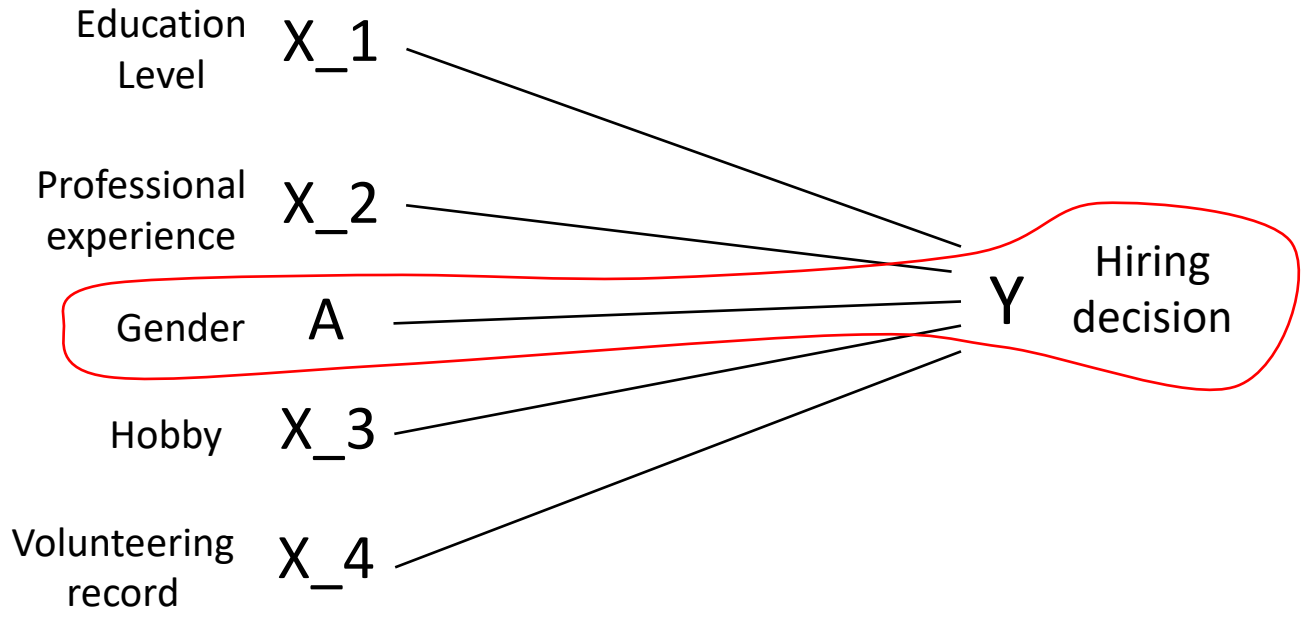


$$\min_{c} d(\mathbf{x}, \mathbf{c})$$

$$\text{s.t.} f(\mathbf{c}) \neq f(\mathbf{x})$$

* Sharma, S., Henderson, J., & Ghosh, J. (2019). Certifai: Counterfactual explanations for robustness, transparency, interpretability, and fairness of artificial intelligence models.

Statistical (Observational) notions of fairness



$$P(\hat{Y} \mid A = 0) = P(\hat{Y} \mid A = 1)$$

Statistical Parity

$$P(\hat{Y} = 1 \mid Y = 1, A = 0) = P(\hat{Y} = 1 \mid Y = 1, A = 1)$$

Equal Opportunity

$$E[S \mid Y = 1, A = 0)] = E[S \mid Y = 1, A = 1]$$

Balance

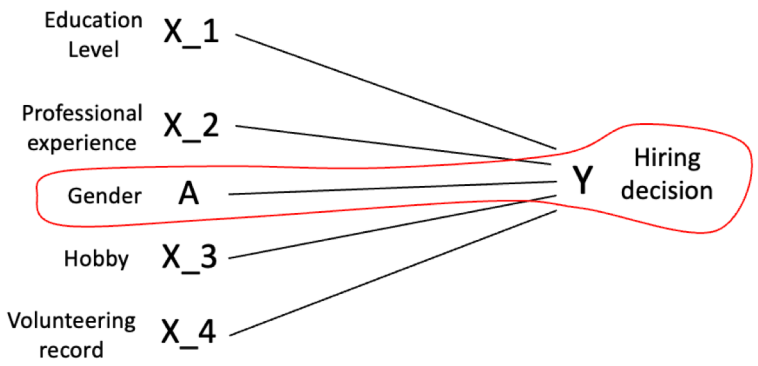
$$\boldsymbol{P(Y = 1 \mid \hat{Y} = 1, A = 0) = P(Y = 1 \mid \hat{Y} = 1, A = 1)}$$

Predictive Parity

$$P(Y = 1 \mid S = s, A = 0) = P(Y = 1 \mid S = s, A = 1) \quad \forall s \in [0, 1]$$

Calibration

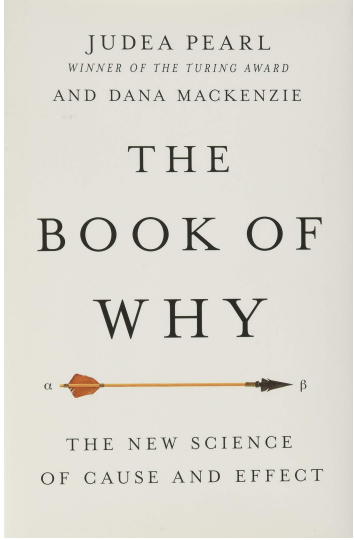
How strong is the effect of A on Y ?



Why not P(Y|A)? → Selection bias

TV = P(Y=1|A=1) - P(Y=1|A=0)

The illusion of correlation



"The correlation we observe is an illusion. An illusion we brought upon ourselves by choosing which events to include in our dataset and which to ignore."

Example 1:

Flip two coins 100 times, and write down the results only when at least one of them comes up head

Coin 1	Coin 2
Head	head
Tail	head
head	tail
Tail	head
Head	head

Notice the dependence: every time coin1 lands tail, coin2 lands head !

Example 2:

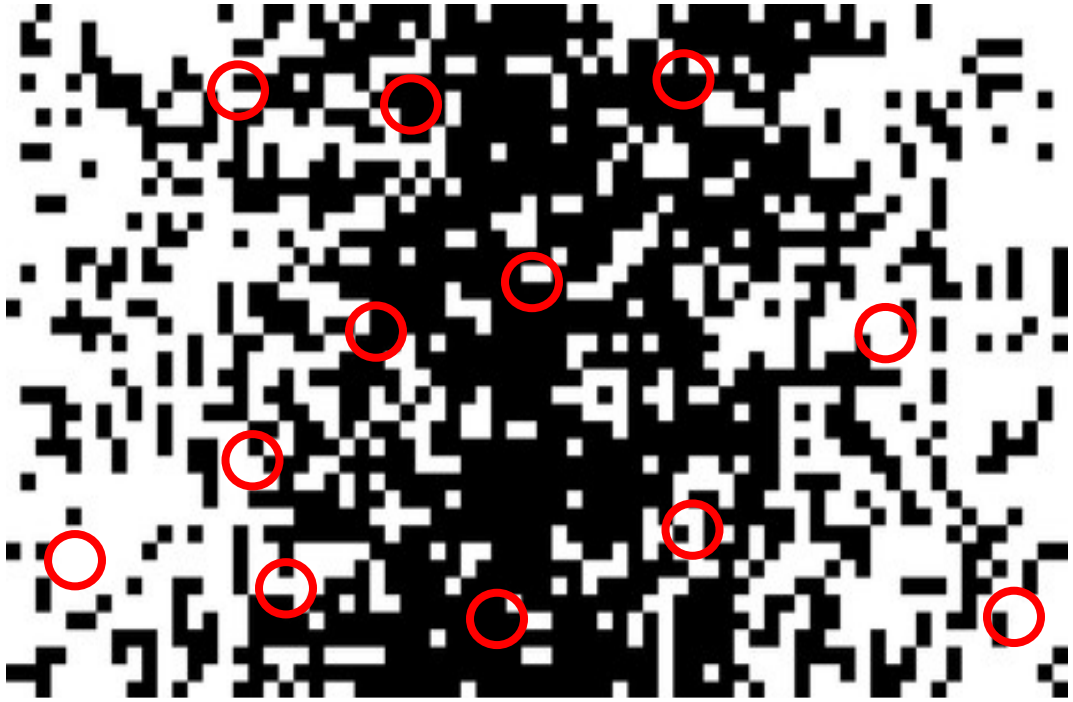
Did you notice that among the people you date, the attractive ones are more likely to be jerks ?

You are dating from these:

Attractive	Jerk
Attractive	Nice
Not attractive	Nice
Not attractive	Jerk

Why observable association is not reliable to establish the effect of a variable on another variable ?

Y



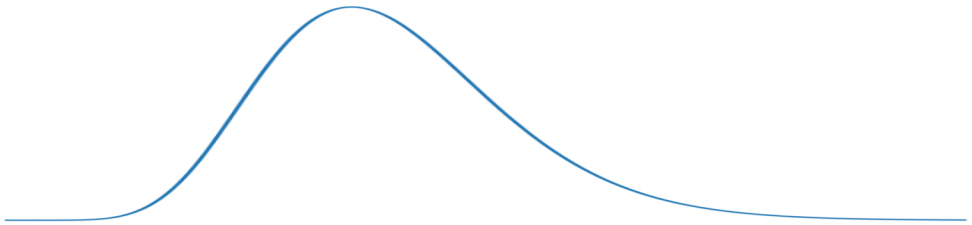
X

Data Collection

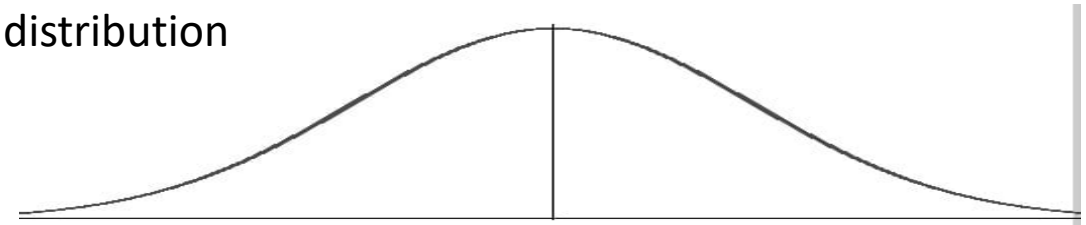
Sampling bias

Selection bias

Sample distribution



Population distribution



Simpson's Paradox

Statistical parity = 7/15 – 8/15 = **-1/15**

Discrimination against women

Discrimation in favor of women

	A	T	Y
	Gender	Job Type	Hiring
A=1 (Women)	1	0	1
	1	0	1
	1	0	1
	1	0	0
	1	0	0
	1	0	0
	1	0	0
	1	0	0
	1	0	0
	1	1	1
	1	1	1
	1	1	1
	1	1	1
	1	1	0

Hiring rate (T = 0) 3/10 = 0.3

Hiring rate (T = 1) 4/5 = 0.8

Total hiring rate 7/15

	A	T	Y
	Gender	Job Type	Hiring
A=0 (Men)	0	0	1
	0	0	0
	0	0	0
	0	0	0
	0	0	0
	0	1	1
	0	1	1
	0	1	1
	0	1	1
	0	1	1
	0	1	1
	0	1	1
	0	1	0
	0	1	0
	0	1	0

Hiring rate (T=0) 1/5 = 0.2

Hiring rate (T=1) 7/10 = 0.7

Total hiring rate 8/15

A = 0	Man
A = 1	Woman

T = 0	Flexible time job
T = 1	Non-flexible time job

Y=0	Not hired
Y=1	Hired

Causality

“The ability to learn causality is considered as a significant component of human-level intelligence and can serve as the foundation of AI” [Pearl 2018]

“The discovery of causal relationships from purely observational data is a fundamental problem in science” [Mooij 2016]

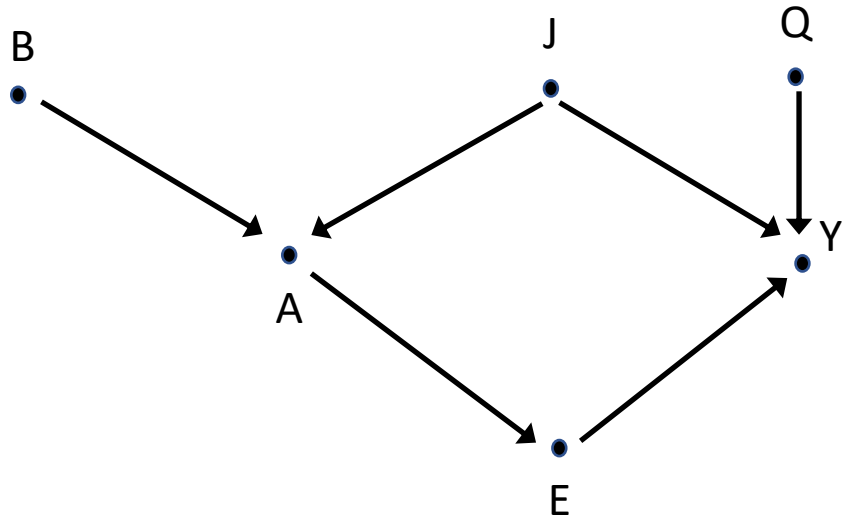
“Almost all of science is about identifying causal relations and the laws or regularities that govern them” [Glymour 2019]

[Pearl 2018] Judea Pearl. 2018. Theoretical impediments to machine learning with seven sparks from the causal revolution

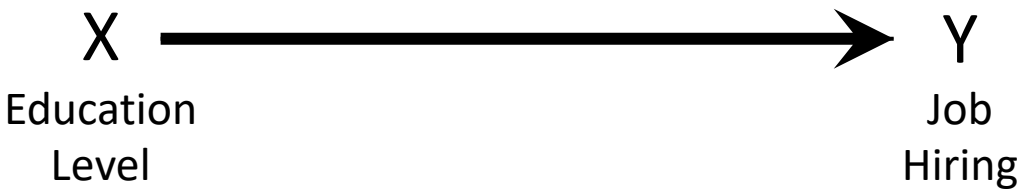
[Mooij 2016] Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., & Schölkopf, B. (2016). Distinguishing cause from effect using observational data: methods and benchmarks..

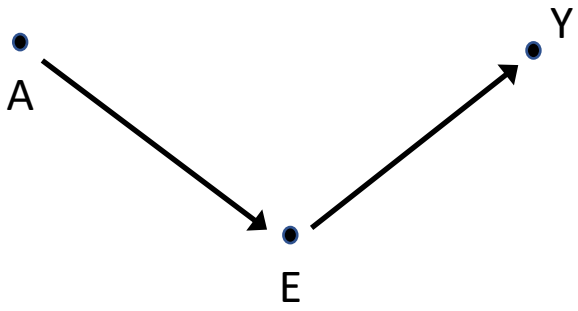
[Glymour 2019] Glymour, Clark, Kun Zhang, and Peter Spirtes. "Review of causal discovery methods based on graphical models." (2019).

Causal Relation



Causal Relation



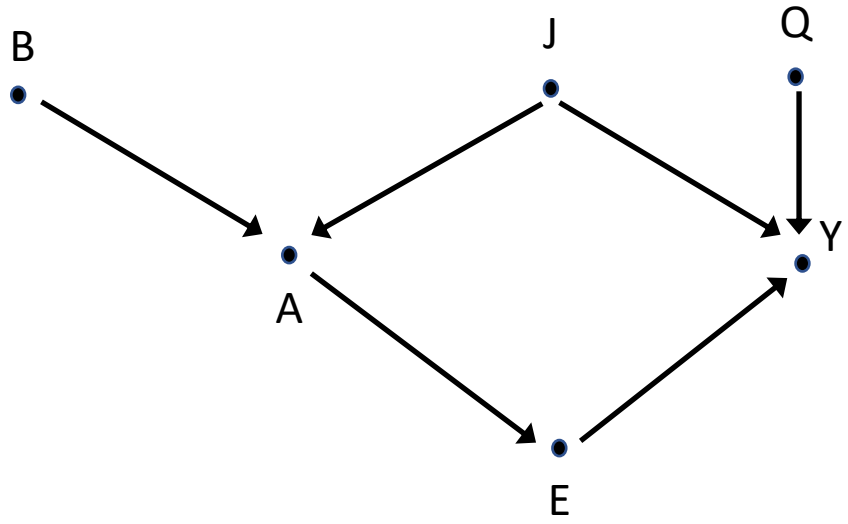


Mediator

$A \not\perp Y$

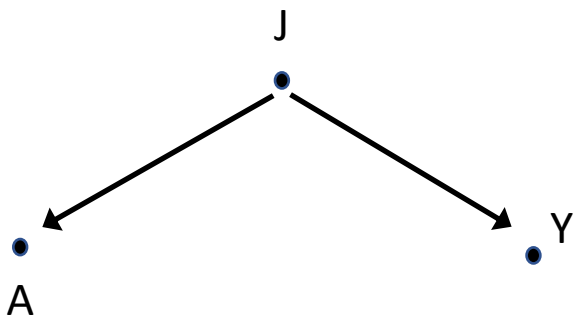
$A \perp Y \mid E$

Causal Relation



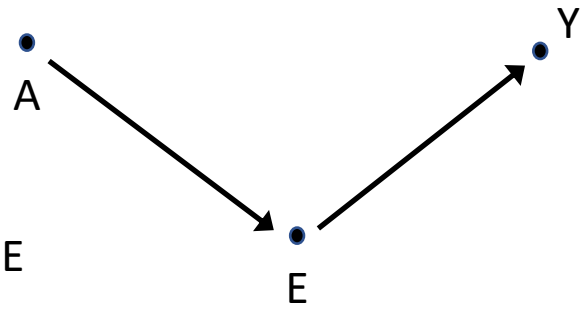
Causal Relation

Confounder (The correlation is not causal)



$A \not\perp Y$

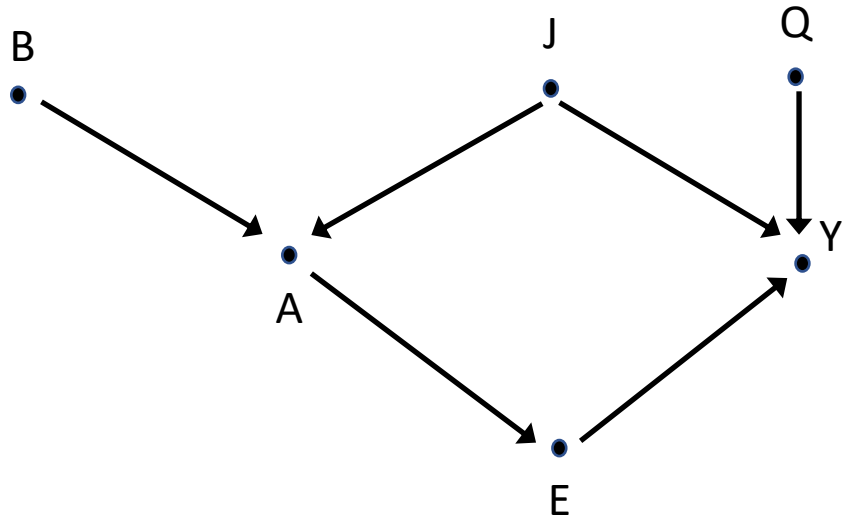
$A \perp Y \mid J$



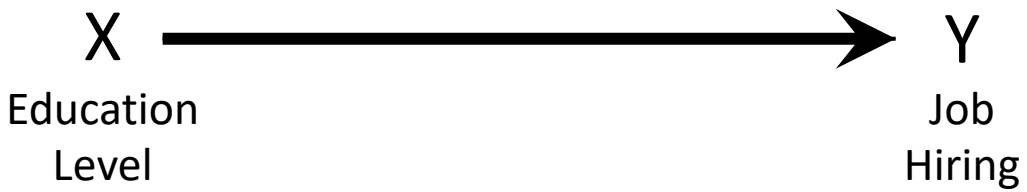
$A \perp Y \mid E$

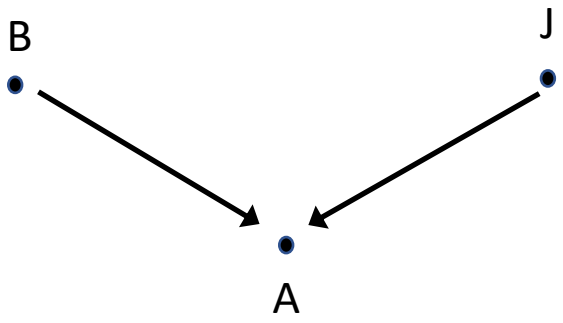
Mediator (The correlation is causal)

Causal Relation



Causal Relation



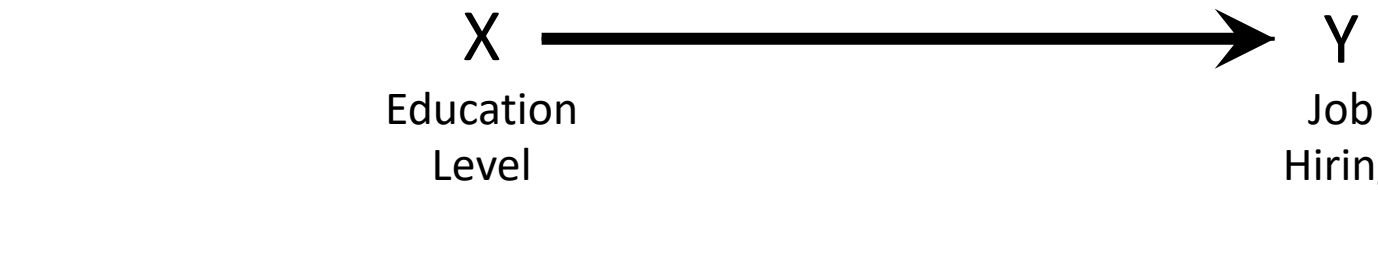


Collider

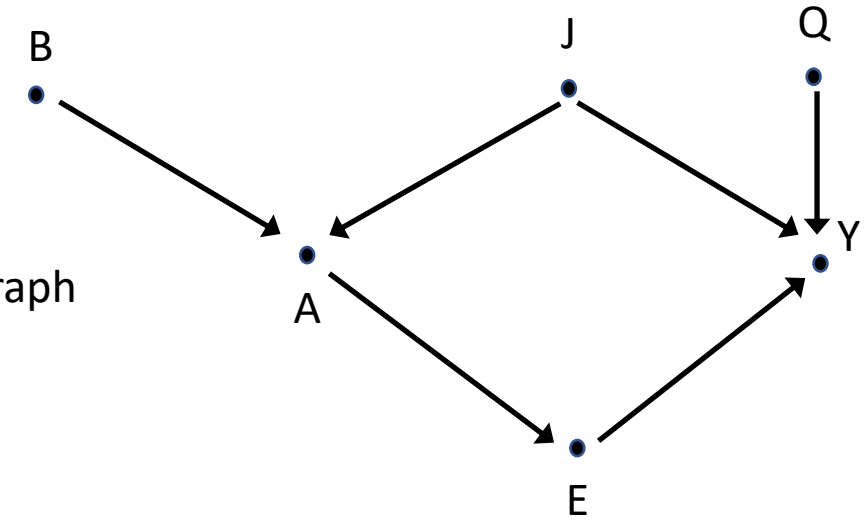
$B \perp J$

$B \not\perp J \mid A$

Causal Relation



Causal Graph



$A \not\perp Y$

$A \perp Y \mid J, E$

$B \not\perp E \mid$

$B \perp E \mid A$

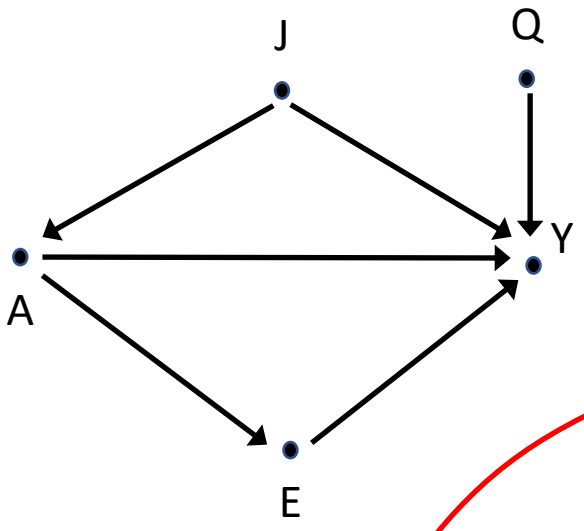
$J \perp Q$

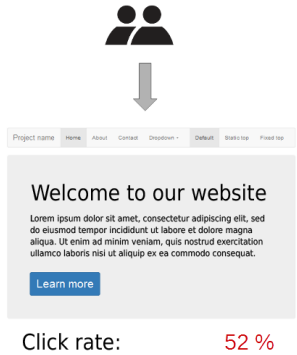
$J \not\perp Q \mid Y$

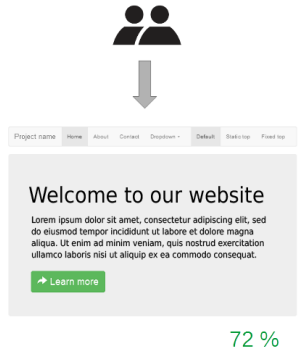
Statistical tests

Data

How to measure the causal effect reliably ?







A/B testing

The golden standard to measure causal effects is:

Randomized Controlled Trials (RCT)

Randomly allocating subjects to two or more groups

Treatment ← Comparison → Control

(Receives the intervention) (No-intervention, Placebo, etc.)

- It is the experimenter that does the allocation (not the subjects that choose)
- The experiment should be properly randomized:

All factors that influence the outcome variable are either static, or vary at random, except one

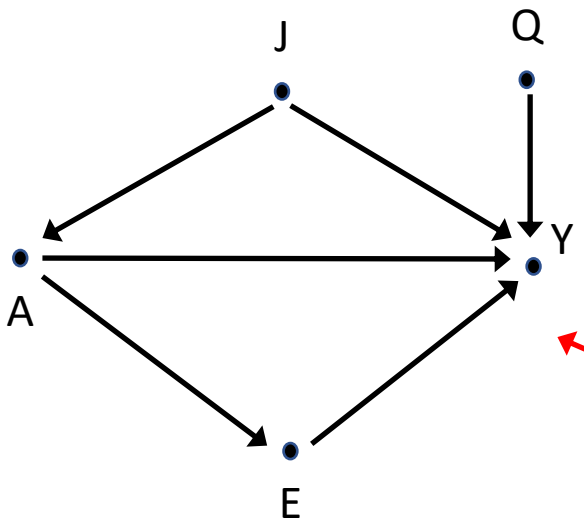
⇒ So any change in the outcome variable must be due to that one input variable.

An experiment involves an action (not mere observation)

In medical studies: select half of individuals randomly, and give them the treatment

In fairness problems: select half of candidates and ***set*** their gender to protected group (female).

How to measure the causal effect reliably ?

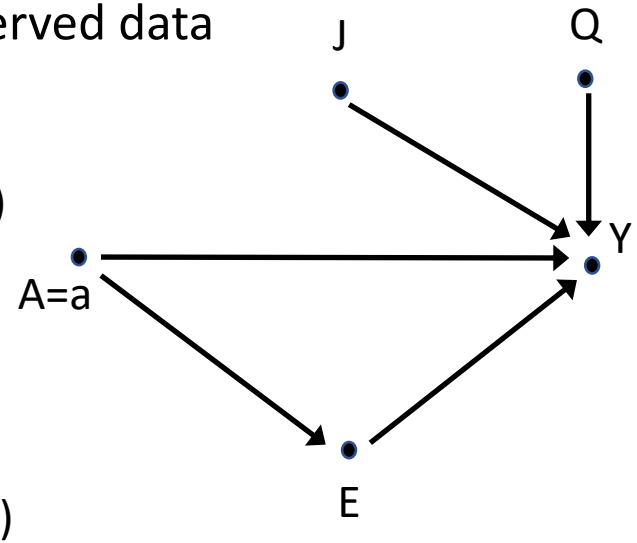


Causal Inference:

Estimating the effect of the intervention from observed data

Intervention: setting the value of a variable do(A = a)

Two different causal models



$P(Y=y|A=a)$

The population distribution of Y among individuals whose A value is a

$P(Y=y|do(A=a))$

The population distribution of Y if everyone in the population had their A value fixed at a.

Total (causal) Effect:

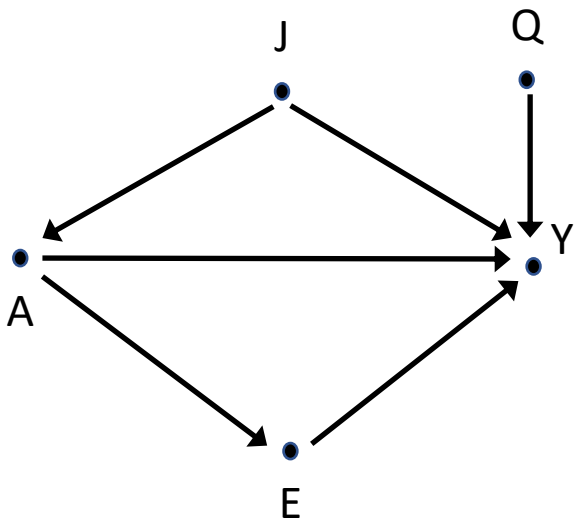
$TE = ACE = P(Y=1|do(A=1)) - P(Y=1|do(A=0))$

$P(y_{A=a})$

$P(y_{A \leftarrow a})$

$P(y_a)$

How to measure the causal effect reliably ?



Estimating P(Y|do(A=a)) from observed data

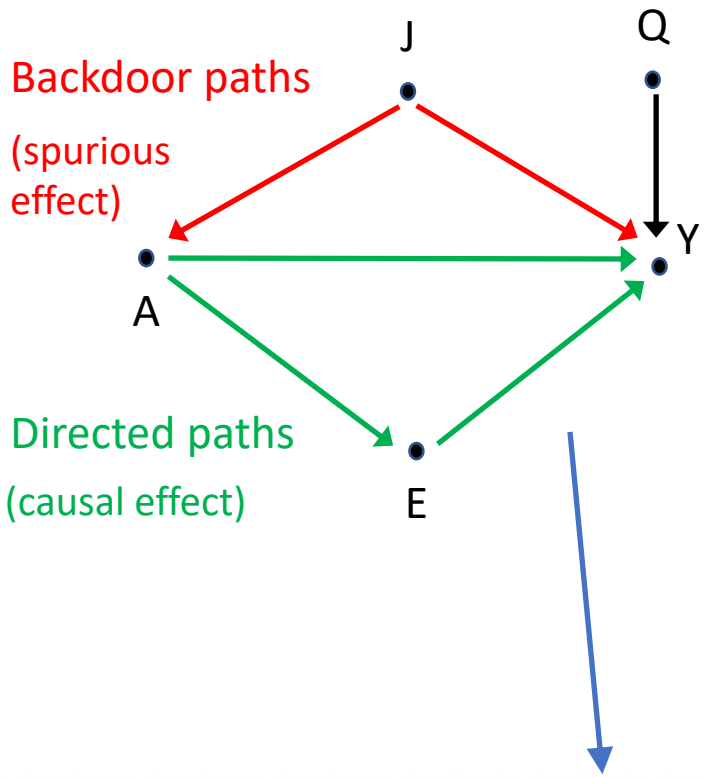
Definition 3.3.1 (The Backdoor Criterion) *Given an ordered pair of variables (X, Y) in a directed acyclic graph G, a set of variables Z satisfies the* backdoor criterion *relative to (X, Y) if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X.*

If a set of variables Z satisfies the backdoor criterion for X and Y, then the causal effect of X on Y is given by the formula

$$P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, Z = z)P(Z = z)$$

* Glymour, M., Pearl, J., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

How to measure the causal effect reliably ?



Estimating P(Y|do(A=a)) from observed data

Definition 3.3.1 (The Backdoor Criterion) *Given an ordered pair of variables (X, Y) in a directed acyclic graph G, a set of variables Z satisfies the* backdoor criterion *relative to (X, Y) if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X.*

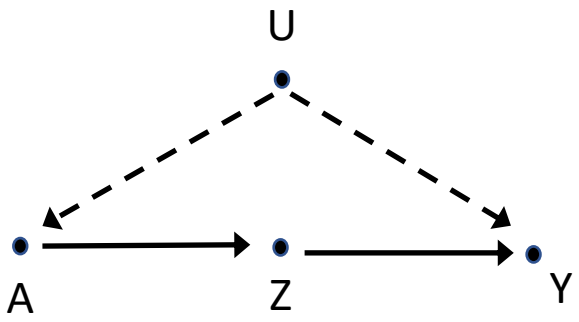
If a set of variables Z satisfies the backdoor criterion for X and Y, then the causal effect of X on Y is given by the formula

$$P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, Z = z)P(Z = z)$$

$$P(Y = y|do(A = a)) = \sum_j P(Y = y|A = a, J = j)\, P(J = j)$$

* Glymour, M., Pearl, J., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

How to measure the causal effect reliably ?



Estimating P(Y|do(A=a)) from observed data

Definition 3.4.1 (Front-Door) *A set of variables Z is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) if*

1. *Z intercepts all directed paths from X to Y.*
2. *There is no unblocked path from X to Z.*
3. *All backdoor paths from Z to Y are blocked by X.*

Theorem 3.4.1 (Front-Door Adjustment) *If Z satisfies the front-door criterion relative to (X, Y) and if P(x, z) > 0, then the causal effect of X on Y is identifiable and is given by the formula*

$$P(y|do(x)) = \sum_z P(z|x) \sum_{x'} P(y|x', z)P(x') \qquad (3.16)$$

* Glymour, M., Pearl, J., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

How to measure the causal effect reliably ?

Estimating P(Y|do(A=a)) from observed data

Back-door adjustment

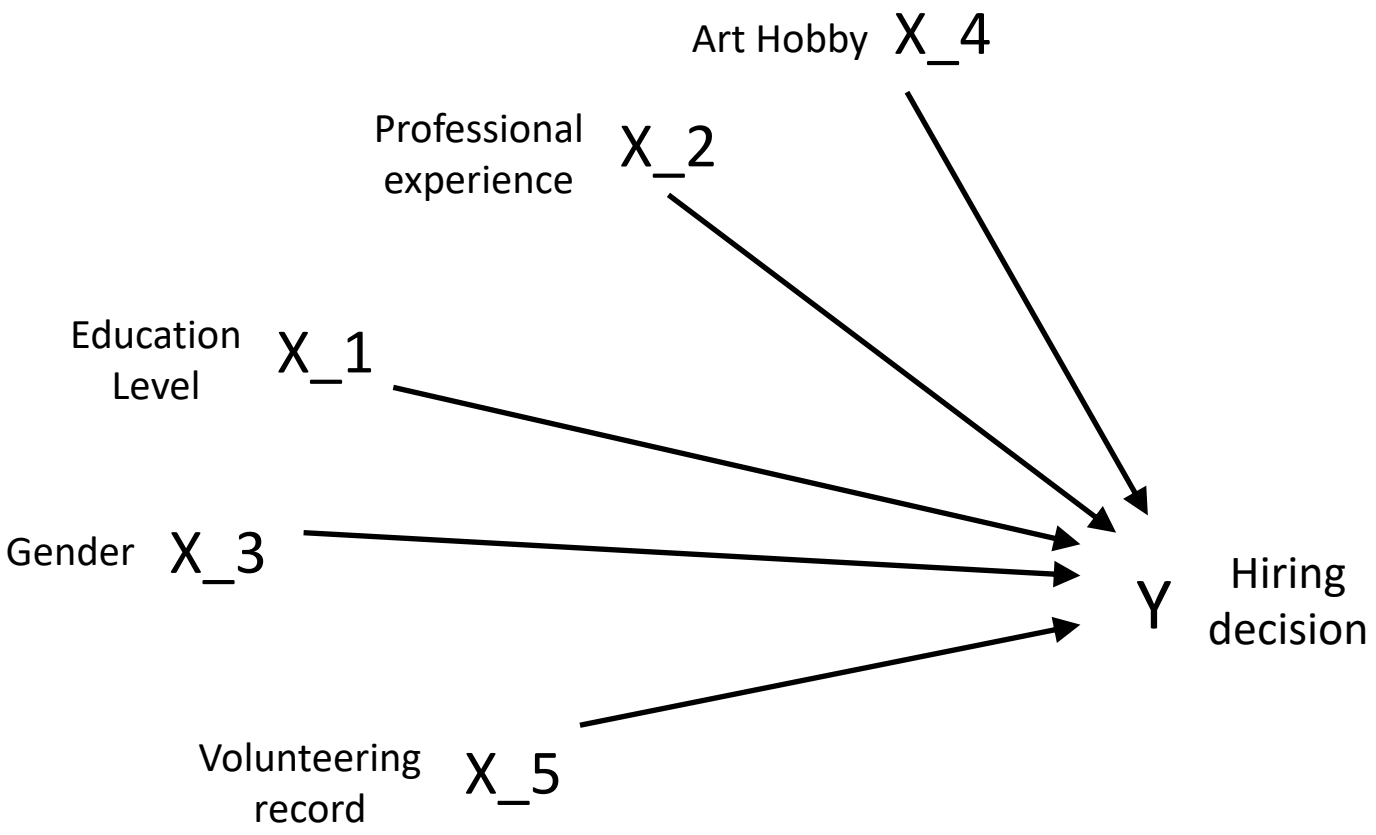
Front-door adjustment

Proposition 1 (Rules of *do*-Calculus). *Let $\mathcal{G}$ be the DAG of a CBN. For any disjoint subsets of nodes* $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, $\mathbf{W}$, *there are following rules.*

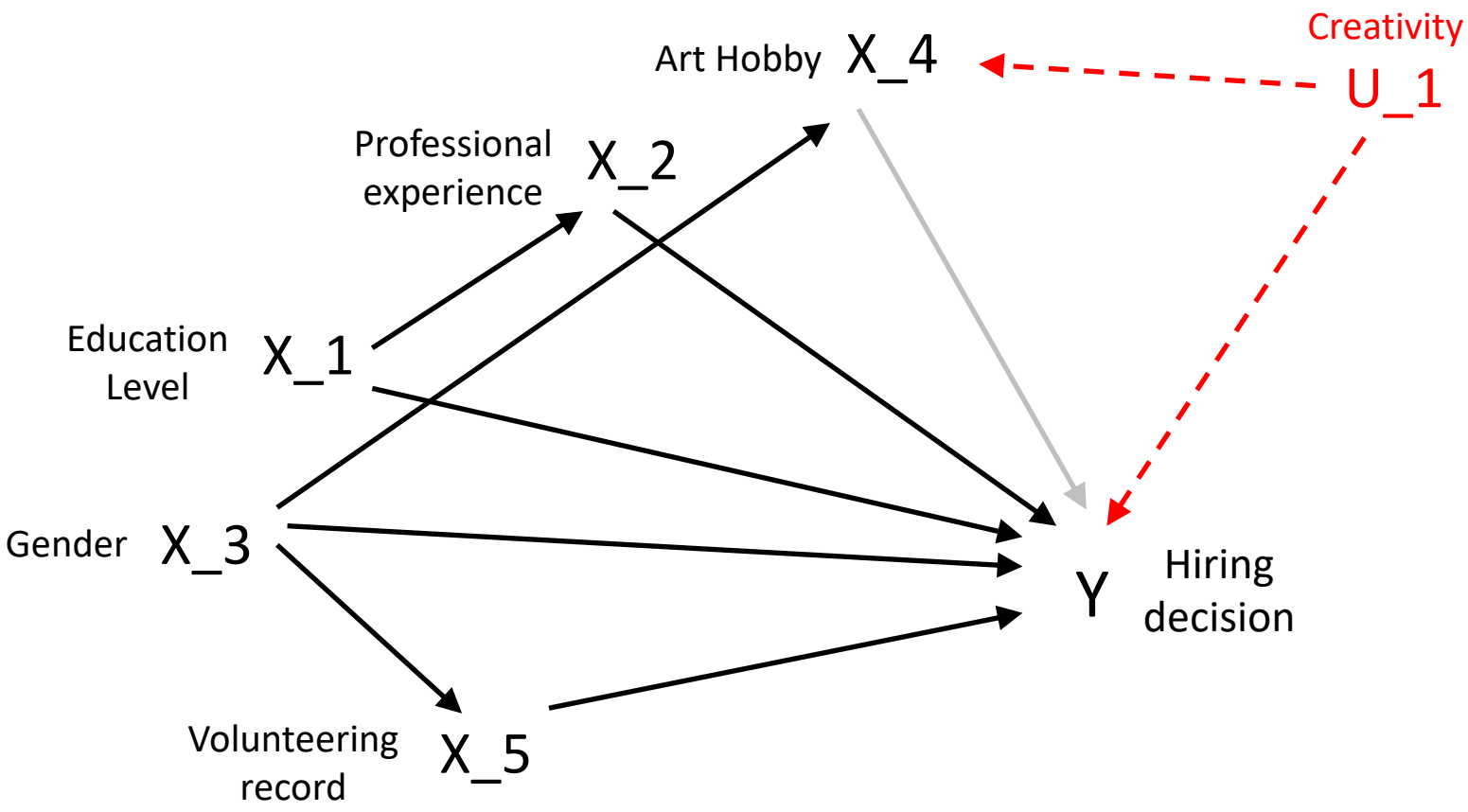
1. $P(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, \mathbf{w}) = P(\mathbf{y}|do(\mathbf{x}), \mathbf{w})$, *if* $\mathbf{Y}$ *and* $\mathbf{Z}$ *are d-separated by* $\mathbf{X} \cup \mathbf{W}$ *in* $\mathcal{G}_{\overline{\mathbf{X}}}$.
2. $P(\mathbf{y}|do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = P(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, \mathbf{w})$, *if* $\mathbf{Y}$ *and* $\mathbf{Z}$ *are d-separated by* $\mathbf{X} \cup \mathbf{W}$ *in* $\mathcal{G}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}$.
3. $P(\mathbf{y}|do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = P(\mathbf{y}|do(\mathbf{x}), \mathbf{w})$, *if* $\mathbf{Y}$ *and* $\mathbf{Z}$ *are d-separated by* $\mathbf{X} \cup \mathbf{W}$ *in* $\mathcal{G}_{\overline{\mathbf{X}}\underline{\mathbf{Z}'}}$, *where* $\mathbf{Z}'$ *is the nodes in* $\mathbf{Z}$ *that are not ancestors of any nodes in* $\mathbf{W}$ *in* $\mathcal{G}_{\overline{\mathbf{X}}}$.

* Glymour, M., Pearl, J., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

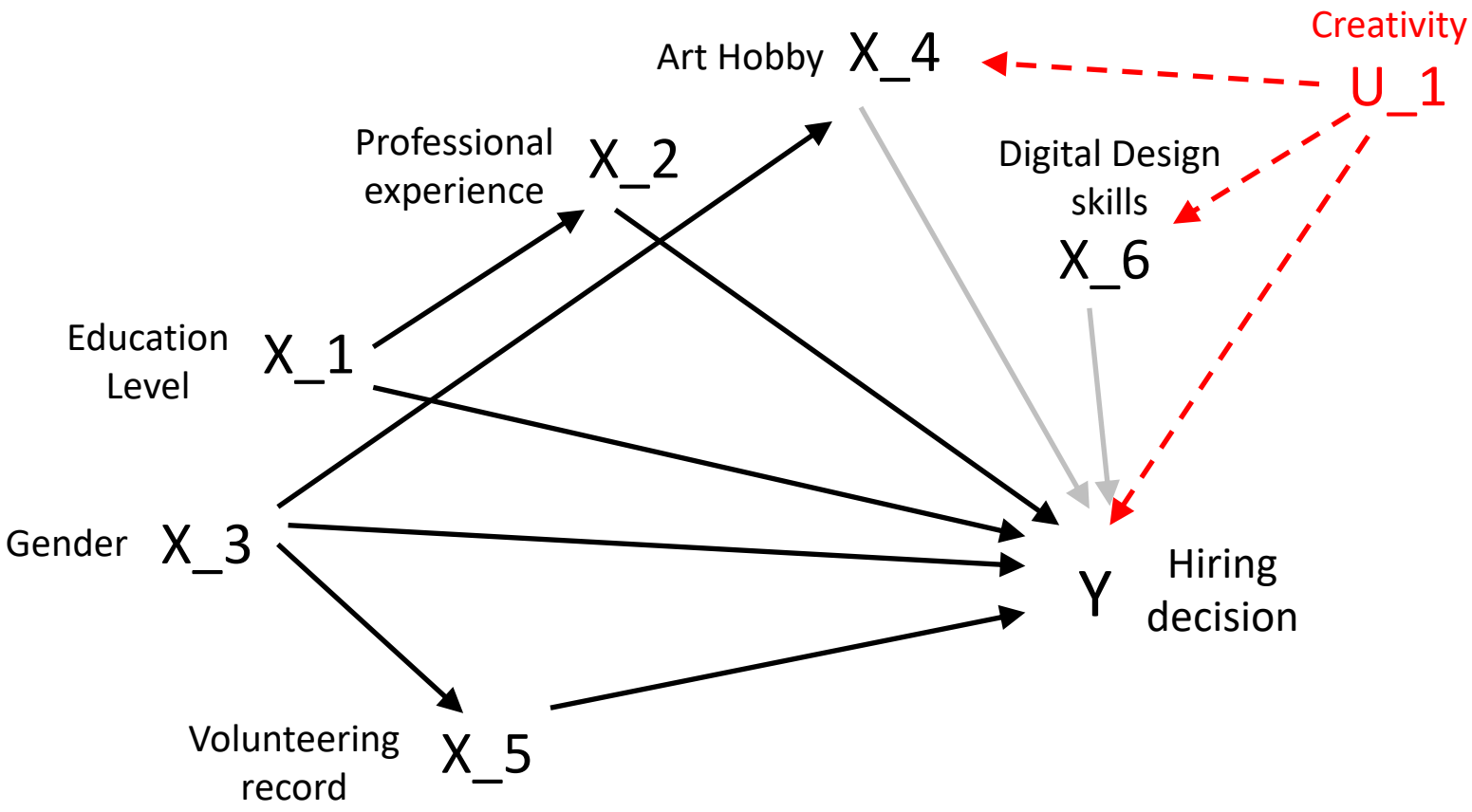
Causal explainability



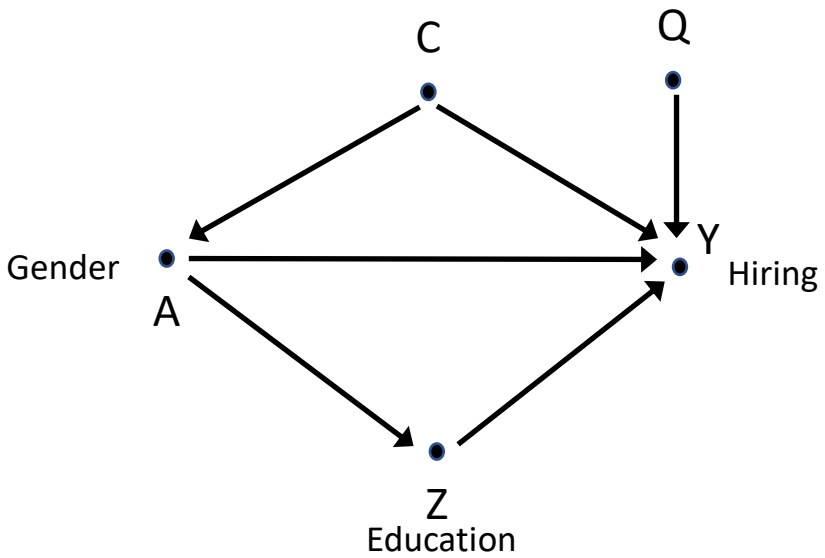
Causal explainability



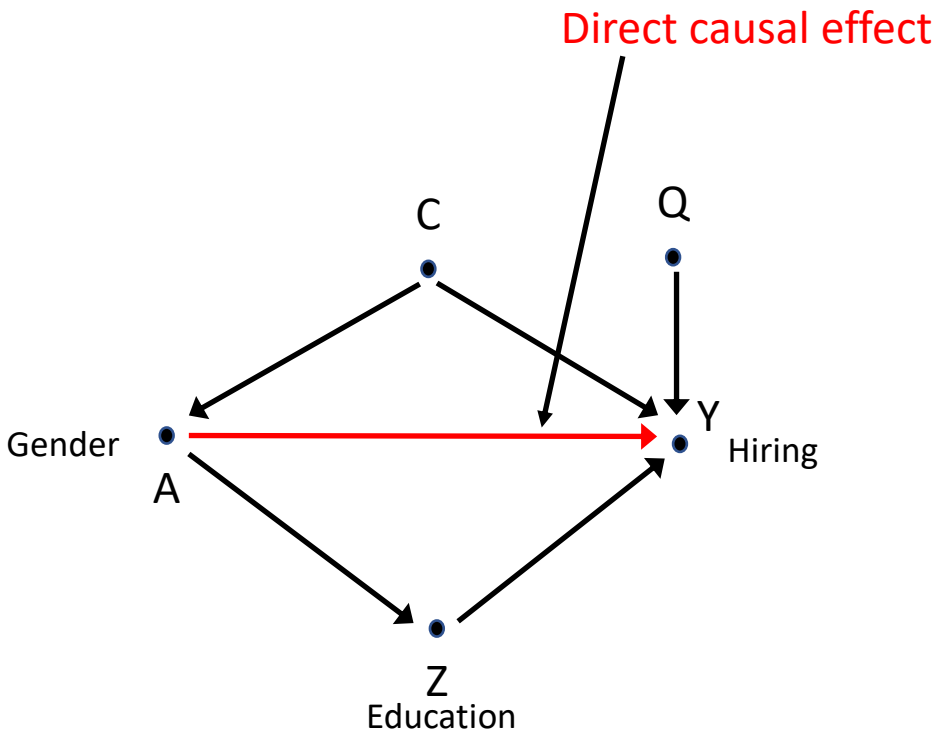
Causal explainability



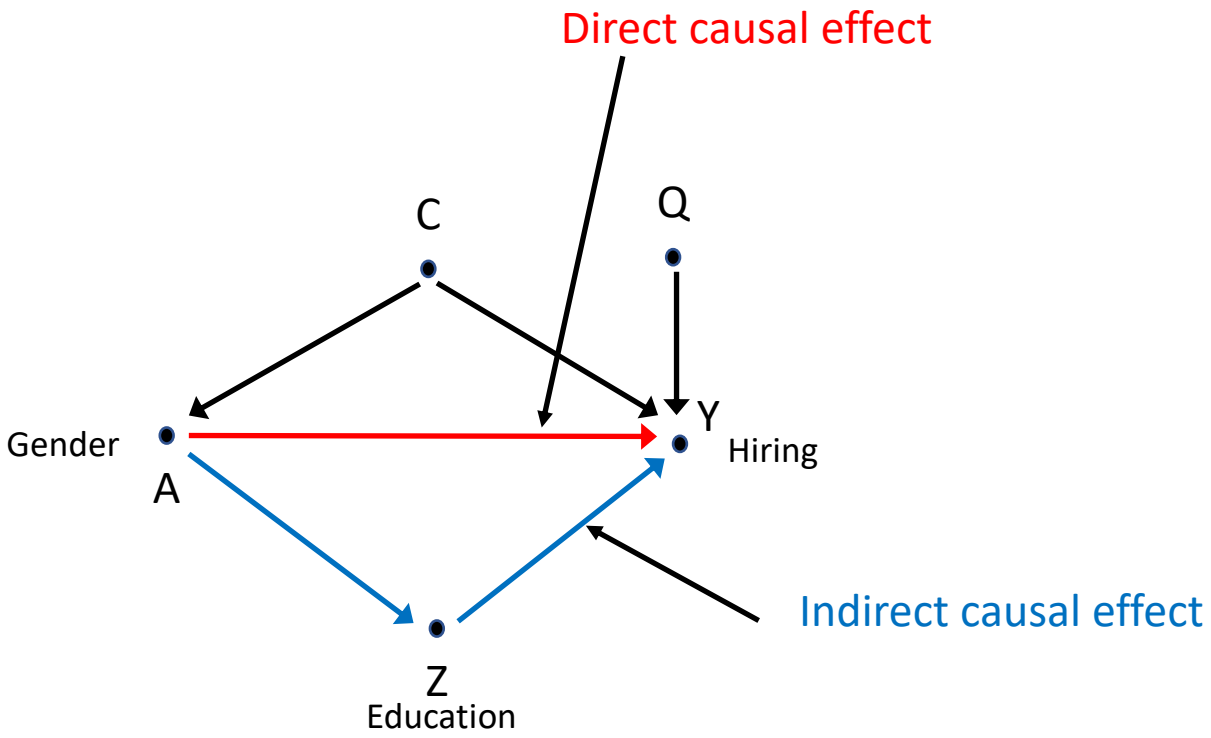
Mediation Analysis



Mediation Analysis

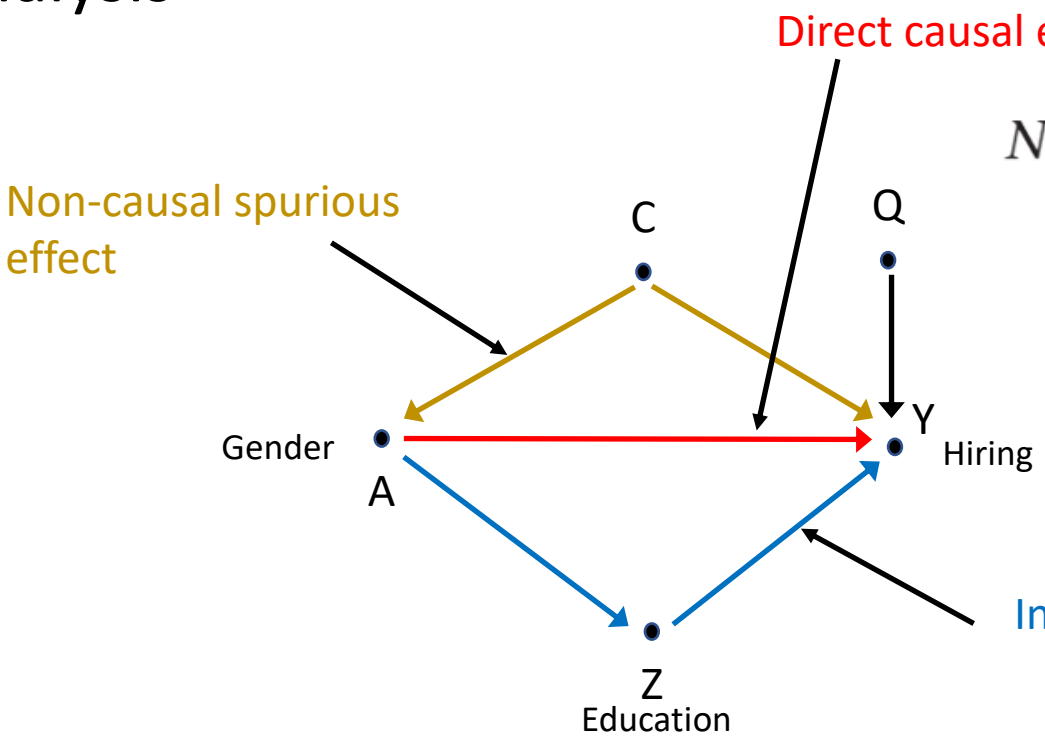


Mediation Analysis



$P(Y=y|do(A=a)) = P(y_a)$

Mediation Analysis



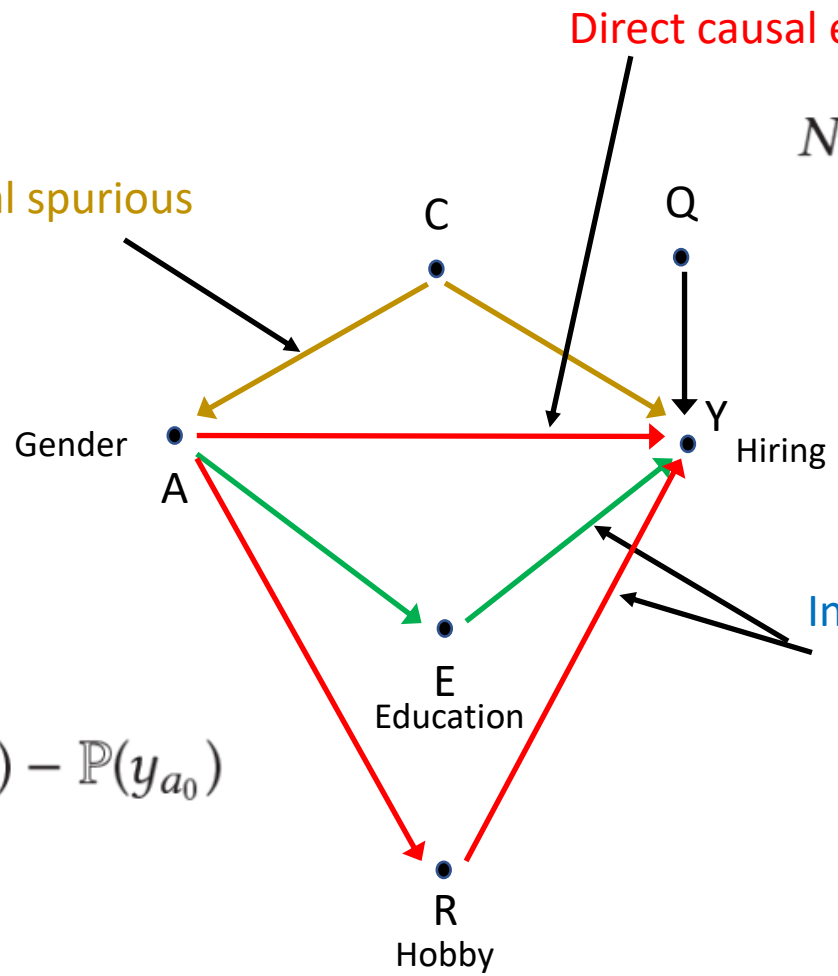
$$NDE_{a_1,a_0}(y) = \mathbb{P}(y_{a_1,Z_{a_0}}) - \mathbb{P}(y_{a_0})$$

discrimination

$$NIE_{a_1,a_0}(y) = \mathbb{P}(y_{a_0,Z_{a_1}}) - \mathbb{P}(y_{a_0})$$

Discrimination ? It depends on Z

Mediation Analysis



$$NDE_{a_1,a_0}(y) = \mathbb{P}(y_{a_1,Z_{a_0}}) - \mathbb{P}(y_{a_0})$$

discrimination

Indirect causal effect

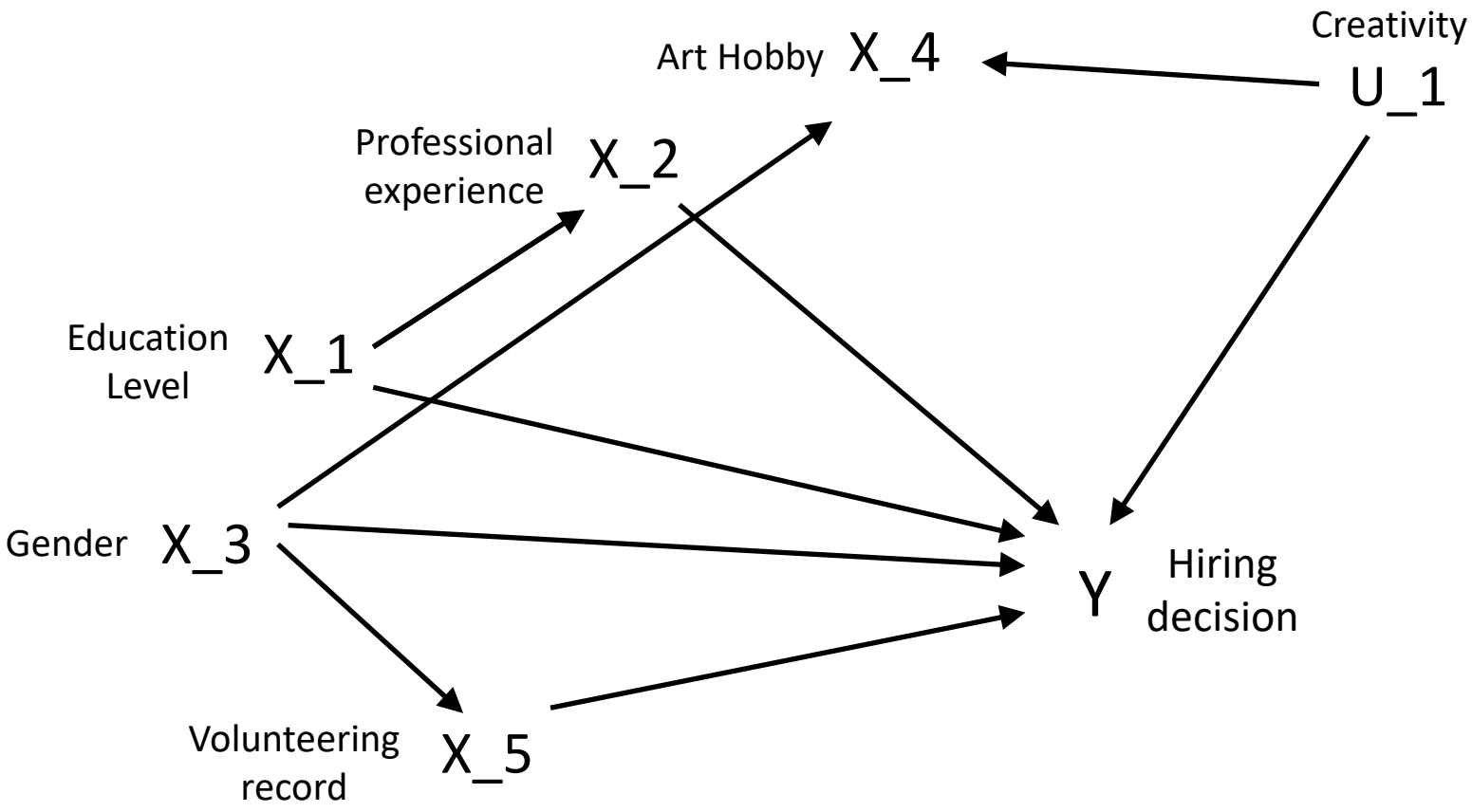
$$NIE_{a_1,a_0}(y) = \mathbb{P}(y_{a_0,Z_{a_1}}) - \mathbb{P}(y_{a_0})$$

Discrimination ? It depends on Z

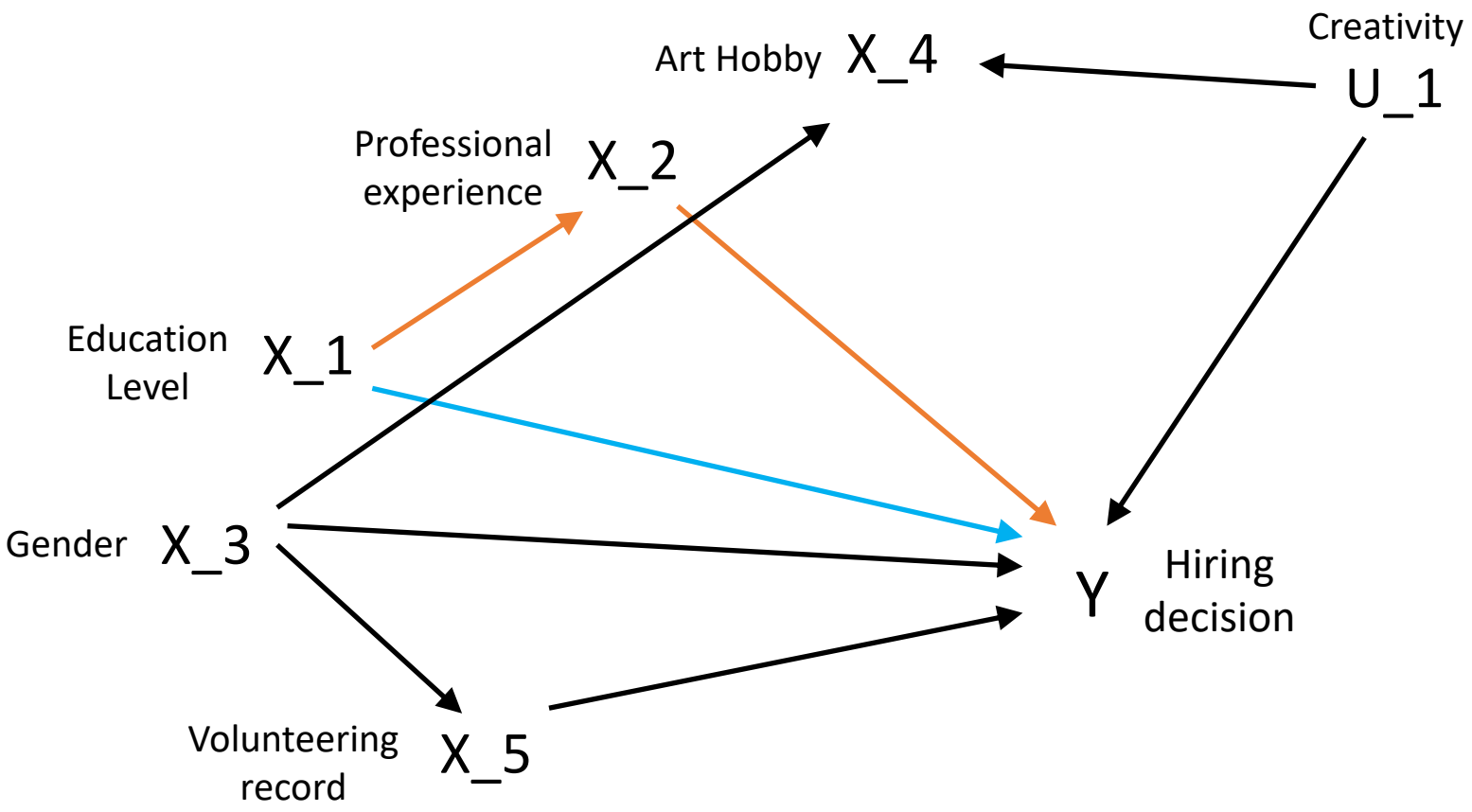
Path-Specific effect

$$PSE^{\pi}_{a_1,a_0}(y) = \mathbb{P}(y_{a_1|_{\pi},a_0|_{\overline{\pi}}}) - \mathbb{P}(y_{a_0})$$

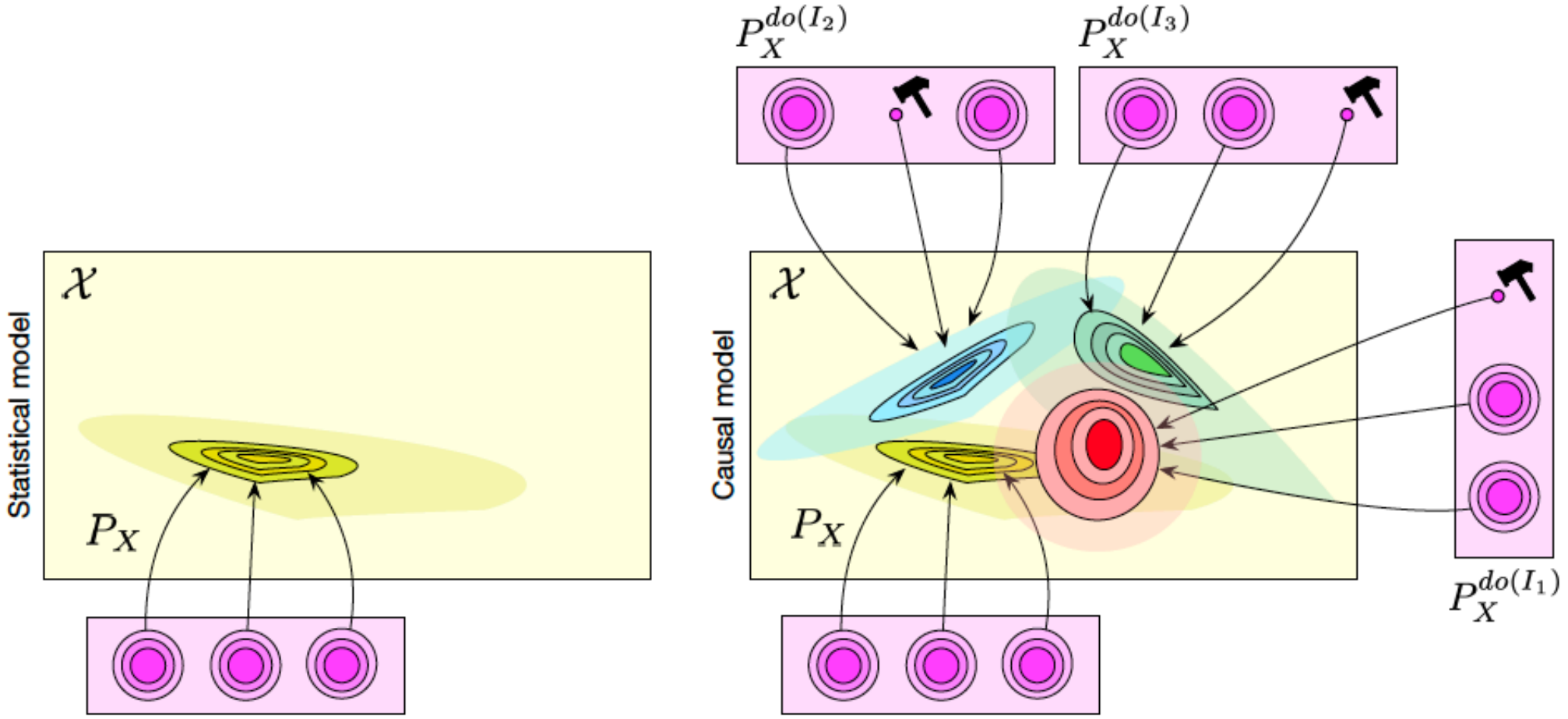
Causal explainability



Causal explainability



Causality and out-of-distribution (OOD) Learning



* Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., & Bengio, Y. (2021). Towards causal representation learning.

Current and future work

Causal discovery (structure learning) algorithms

- Does the shape of data (type, binarization, pre-processing, etc.) has an impact on the generated causal model/graph ?

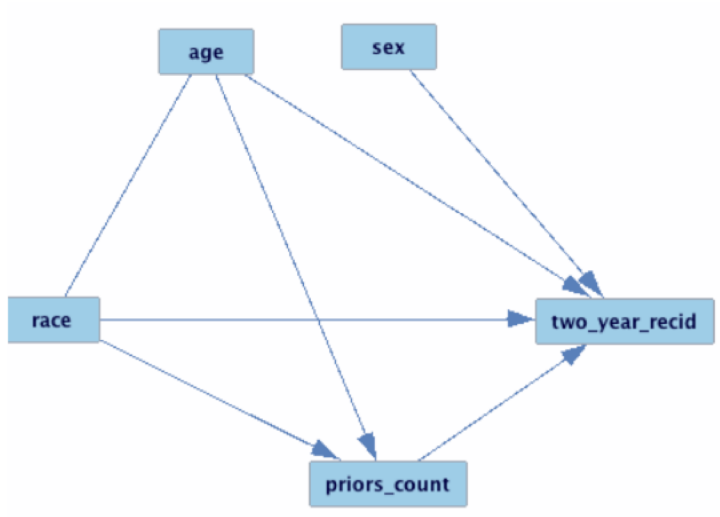
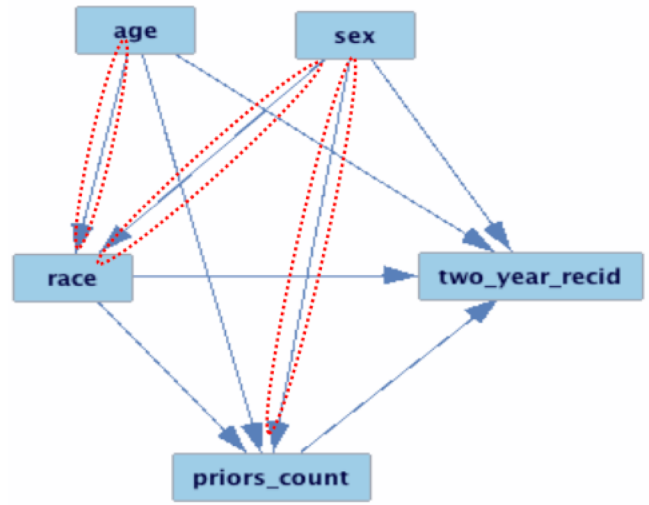
- Do causal discovery algorithms provide confidence levels on the edges?

- What is the impact of confidence levels on the causal effects ?

Mediation analysis for explainability

Experiments on Causal graph generation Using Tetrad

1- Compas dataset

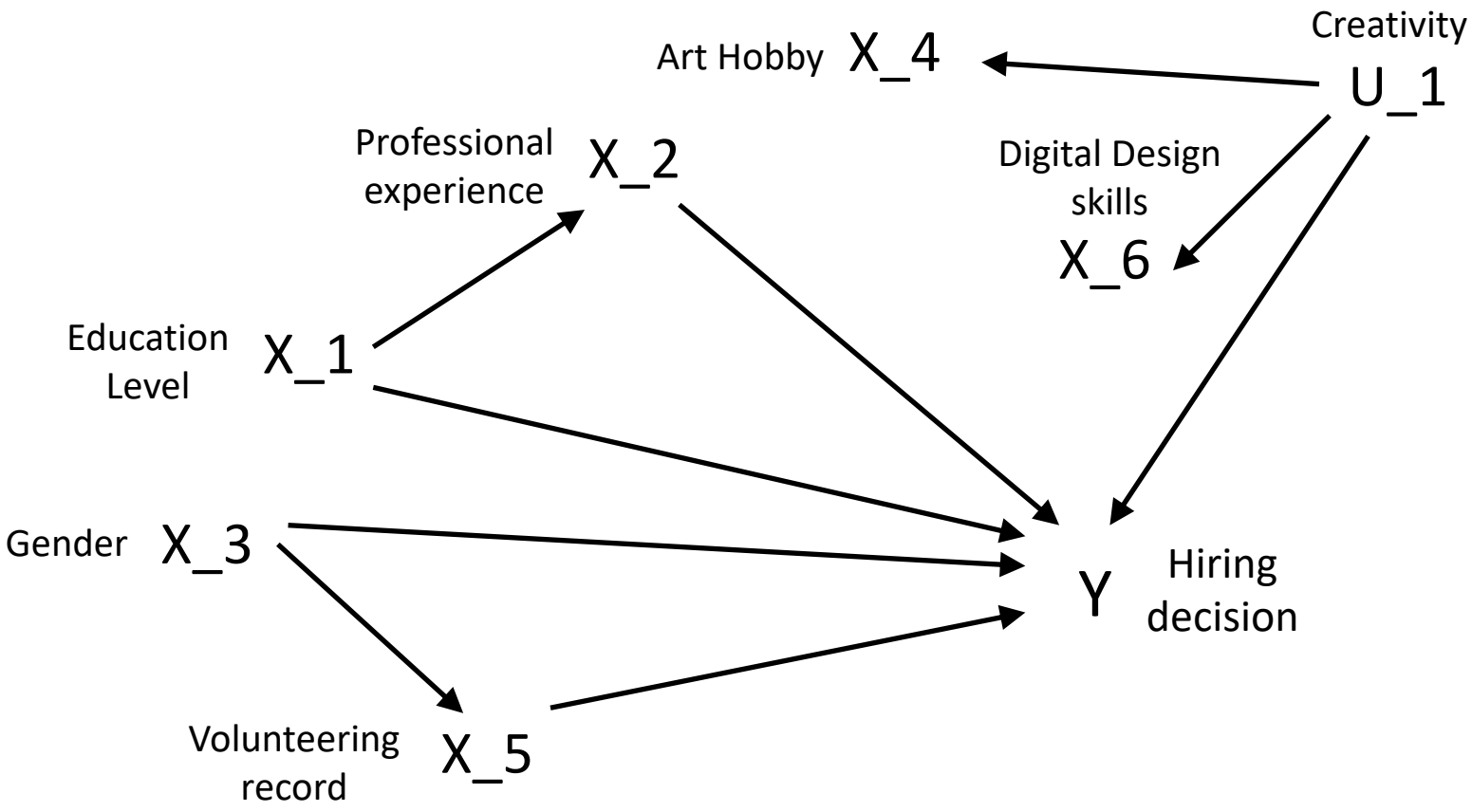
Manipulation of the data	Generated graphs	
age & priors_counts	not binarized	binarized
1- PC algorithm		

Compas

Before binarization (age, prior_counts)

After binarization (age, prior_counts)

Purely associational explainability



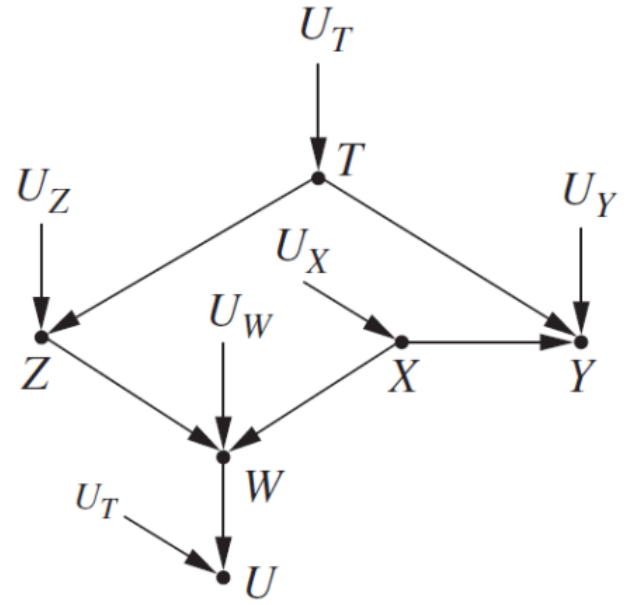
Based only on the CBN, how to tell if two variables are dependent/independent/conditionally-independent ?

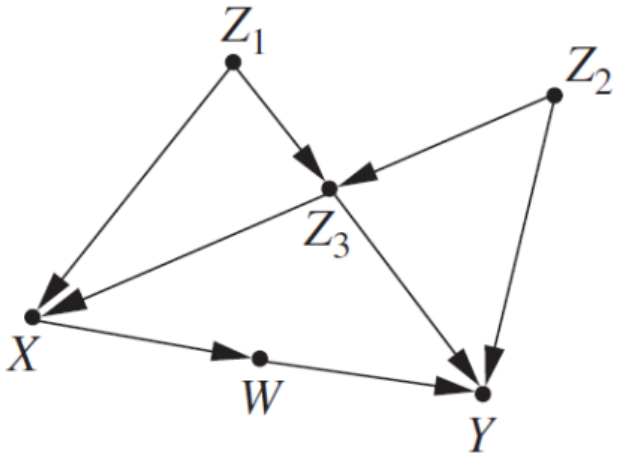
If they are conditionally independent, on which variables we should condition on ?

Definition 2.4.1 (*d*-separation) *A path p is blocked by a set of nodes Z if and only if*

1. *p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., B is conditioned on), or*
2. *p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.*

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.



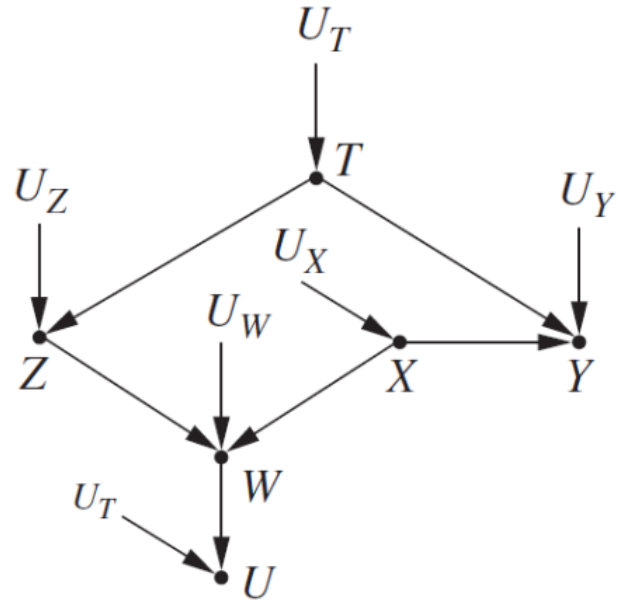


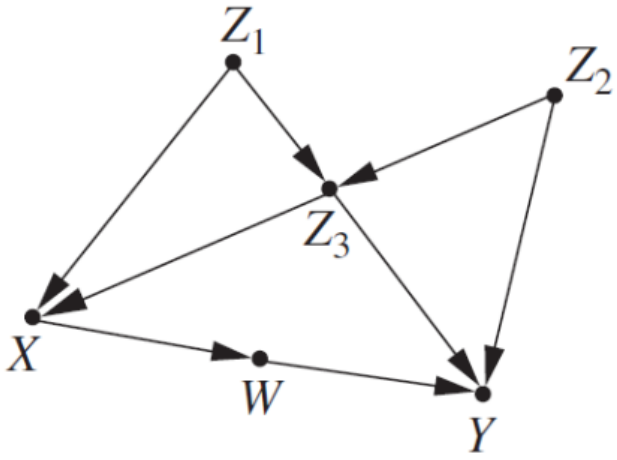
d-separation and variables independence

Definition 2.4.1 (*d*-separation)* *A path p is blocked by a set of nodes Z if and only if*

1. *p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., B is conditioned on), or*
2. *p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.*

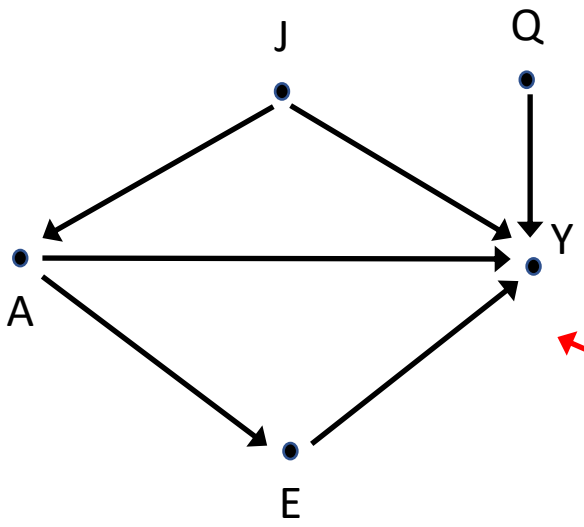
If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.





* Glymour, M., Pearl, J., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

How to measure the causal effect reliably ?



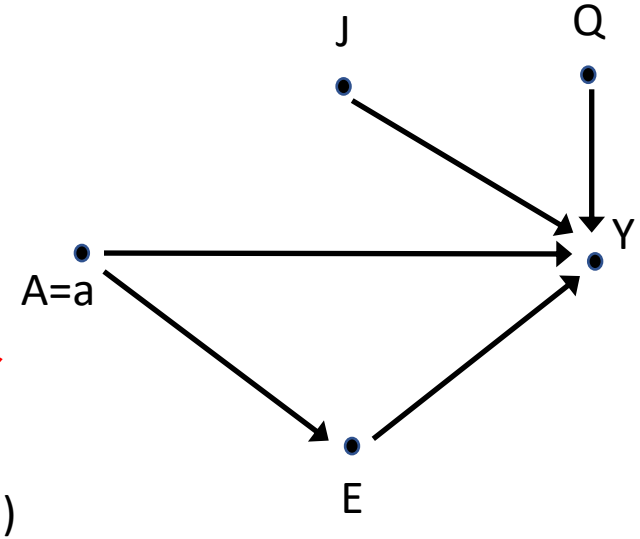
Causal Inference:

Estimating the effect of the intervention from observed data

Intervention: setting the value of a variable (A = a)

$A = F_A (c.J + u_A)$ ⟶ $A = a$

$do(A=a)$



Two different causal models

$P(Y=y|A=a)$

The population distribution of Y among individuals whose A value is a

$P(Y=y|do(A=a))$

The population distribution of Y if everyone in the population had their A value fixed at a.

Total (causal) Effect:

$TE = ACE = P(Y=1|do(A=1)) – P(Y=1|do(A=0))$

$P(y_{A=a})$

$P(y_{A\leftarrow a})$

$P(y_a)$

- Two random variables X and Y are called independent, if for each values of X and Y, x and y,
  - $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$ or
  - $P(X = x|Y = y) = P(X = x)$ or $P(Y = y|X = x) = P(Y = y)$
  - Denoted by $X \perp Y$
- Two random variables X and Y are called conditionally independent given Z, if for each values of (X, Y, Z), (x, y, z),
  - $P(X = x, Y = y|Z = z) = P(X = x|Z = z) \cdot P(Y = y|Z = z)$ or
  - $P(X = x|Y = y, Z = z) = P(X = x|Z = z)$ or
  - $P(Y = y|X = x, Z = z) = P(Y = y|Z = z)$
  - Denoted by $X \perp Y|Z$

- Note: conditional independence neither implies nor is implied by independence.

Random variables A and Y are independent

$$P(Y|A) = P(Y)$$

Our belief Y remains unchanged upon learning A.

Random variables A and Y are conditionally independent given C

$$P(Y|A,C) = P(Y|C)$$

Once we know C, our belief Y remains unchanged upon learning A.

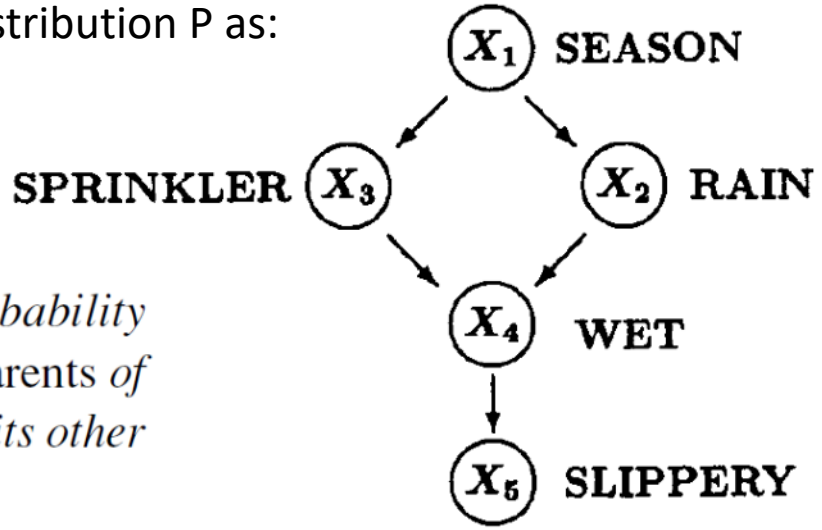
A and Y are independent in the new dataset created by filtering on C.

Causal Bayesian Network (CBN)

represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG).

Suppose we have a distribution P defined on n discrete variables which we may order arbitrarily as X_1, X_2, .. X_n. The chain rule allows to decompose the joint distribution P as:

$$P(x_1, \ldots, x_n) = \prod_j P(x_j \mid x_1, \ldots, x_{j-1}). = \prod_i P(x_i \mid pa_i).$$



Definition 1.2.1 (Markovian Parents)

Let $V = \{X_1, \ldots, X_n\}$ be an ordered set of variables, and let $P(v)$ be the joint probability distribution on these variables. A set of variables PA_j is said to be Markovian parents *of X_j if PA_j is a minimal set of predecessors of X_j that renders X_j independent of all its other predecessors. In other words, PA_j is any subset of $\{X_1, \ldots, X_{j-1}\}$ satisfying*

$$P(x_j \mid pa_j) = P(x_j \mid x_1, \ldots, x_{j-1})$$

$$P(x_1, x_2, x_3, x_4, x_5) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1)P(x_4 \mid x_2, x_3)P(x_5 \mid x_4).$$

and such that no proper subset of PA_j satisfies (1.32).[5]

Definition 1.2.2 (Markov Compatibility)

If a probability function P admits the factorization of (1.33) *relative to DAG G, we say that G* represents *P, that G and P are* compatible, *or that P is* Markov relative *to G.*[6]

Based only on the CBN, how to tell if two variables are dependent/independent/conditionally-independent ?

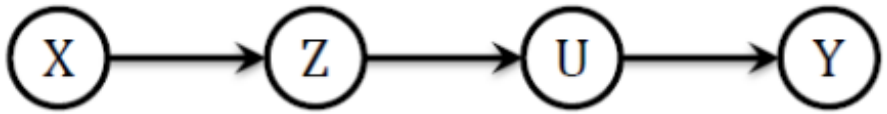
If they are conditionally independent, on which variables we should condition on ?

Definition 2.4.1 (*d*-separation) *A path p is blocked by a set of nodes Z if and only if*

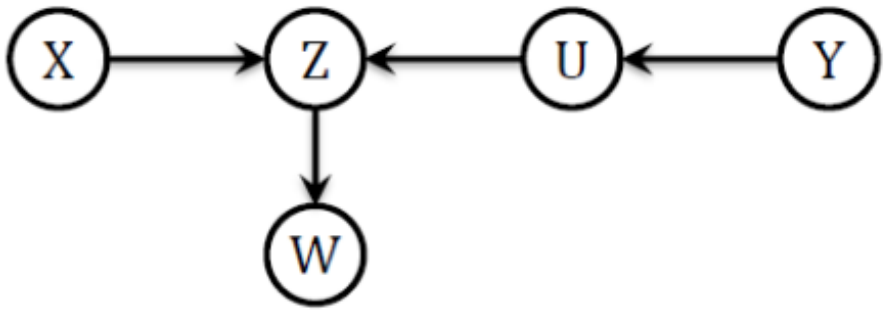
1. *p contains a chain of nodes* $A \rightarrow B \rightarrow C$ *or a fork* $A \leftarrow B \rightarrow C$ *such that the middle node B is in Z (i.e., B is conditioned on), or*
2. *p contains a collider* $A \rightarrow B \leftarrow C$ *such that the collision node B is not in Z, and no descendant of B is in Z.*

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

- Example (blocking of paths)

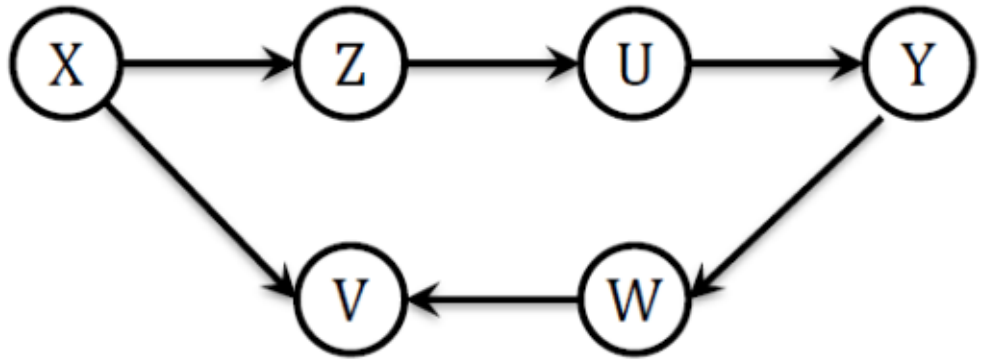


  - Path from X to Y is blocked by conditioning on $\{U\}$ or $\{Z\}$ or both $\{U, Z\}$
- Example (unblocking of paths)



  - Path from X to Y is blocked by $\emptyset$ or $\{U\}$
  - Unblocked by conditioning on $\{Z\}$ or $\{W\}$ or both $\{Z, W\}$

- Example (*d*-separation)



- We have following *d*-separation relations
  - $(X \perp Y|Z)_G, (X \perp Y|U)_G, (X \perp Y|ZU)_G$
  - $(X \perp Y|ZW)_G, (X \perp Y|UW)_G, (X \perp Y|ZUW)_G$
  - $(X \perp Y|VZUW)_G$
- However we do NOT have
  - $(X \perp Y|VZU)_G$

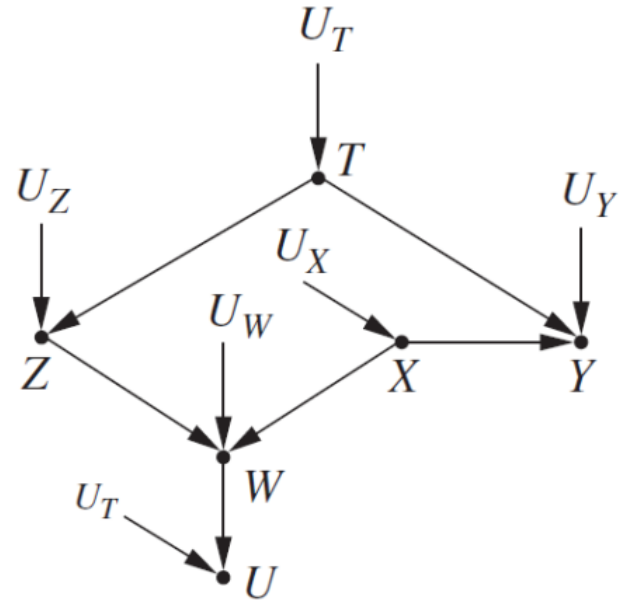
Based only on the CBN, how to tell if two variables are dependent/independent/conditionally-independent ?

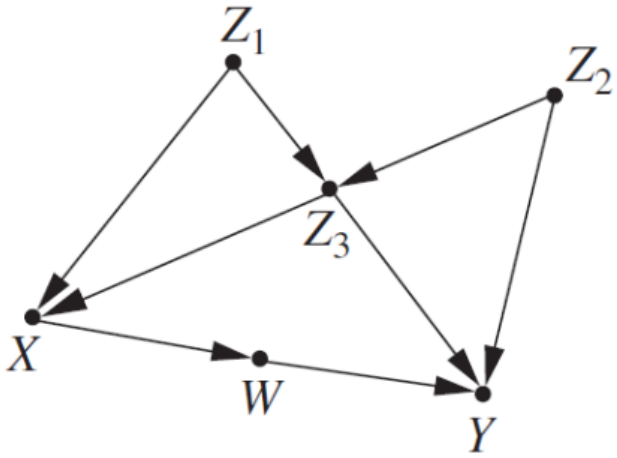
If they are conditionally independent, on which variables we should condition on ?

Definition 2.4.1 (*d*-separation) *A path p is blocked by a set of nodes Z if and only if*

1. *p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., B is conditioned on), or*
2. *p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.*

If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.





Structural Causal Model (SCM)

A causal model is triple $\mathcal{M} =< \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F} >$, where

- $\boldsymbol{U}$ is a set of exogenous (hidden) variables whose values are determined by factors outside the model;
- $\boldsymbol{V} = \{X_1, \cdots, X_i, \cdots\}$ is a set of endogenous (observed) variables whose values are determined by factors within the model;
- $\boldsymbol{F} = \{f_1, \cdots, f_i, \cdots\}$ is a set of deterministic functions where each f_i is a mapping from $\boldsymbol{U} \times (\boldsymbol{V} \setminus X_i)$ to X_i. Symbolically, f_i can be written as

$$x_i = f_i(\boldsymbol{pa}_i, \boldsymbol{u}_i)$$

where $\boldsymbol{pa}_i$ is a realization of X_i's parents in $\boldsymbol{V}$, i.e., $\boldsymbol{Pa}_i \subseteq \boldsymbol{V}$, and $\boldsymbol{u}_i$ is a realization of X_i's parents in $\boldsymbol{U}$, i.e., $\boldsymbol{U}_i \subseteq \boldsymbol{U}$.

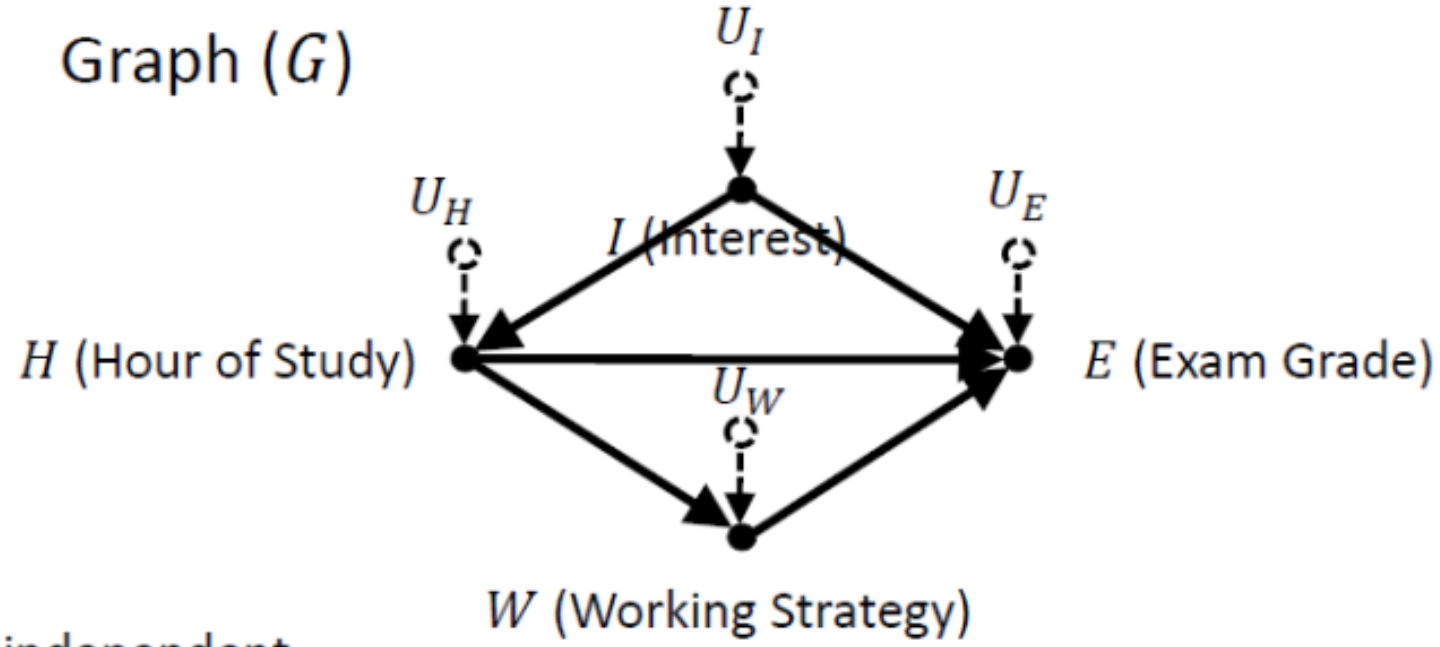
Structural Causal Model (SCM)

Example:

Model (M)

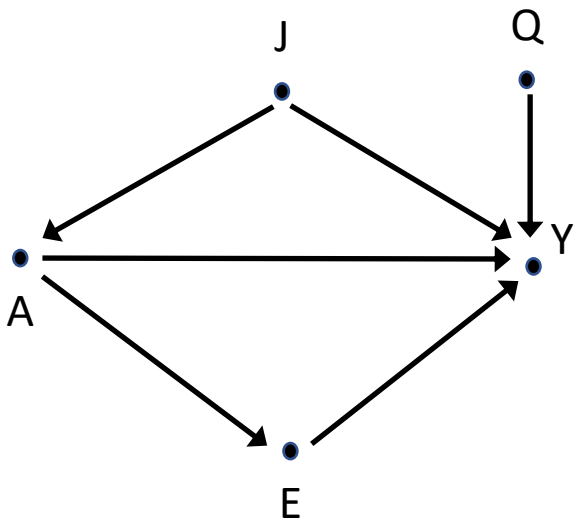
$$i = f_I(u_I)$$
$$h = f_H(i, u_H)$$
$$w = f_W(h, u_W)$$
$$e = f_E(i, h, w, u_E)$$

Graph (G)



Assume U_I, U_H, U_W, U_E are mutually independent.

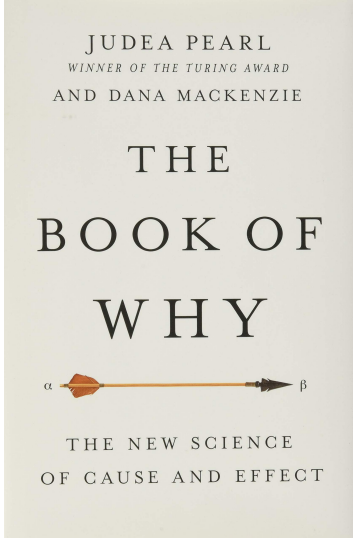
How strong is the causal dependence of Y on A (causal effect of A on Y)?



Why not P(Y|A)? → Selection bias

TV = P(Y=1|A=1) - P(Y=1|A=0)

The illusion of correlation



"The correlation we observe is an illusion. An illusion we brought upon ourselves by choosing which events to include in our dataset and which to ignore."

Example 1:

Flip two coins 100 times, and write down the results only when at least one of them comes up head

Coin 1	Coin 2
Head	head
Tail	head
head	tail
Tail	head
Head	head

Notice the dependence: every time coin1 lands tail, coin2 lands head !

Example 2:

Did you notice that among the people you date, the attractive ones are more likely to be jerks ?

You are dating from these:

Attractive	Jerk
Attractive	Nice
Not attractive	Nice
Not attractive	Jerk