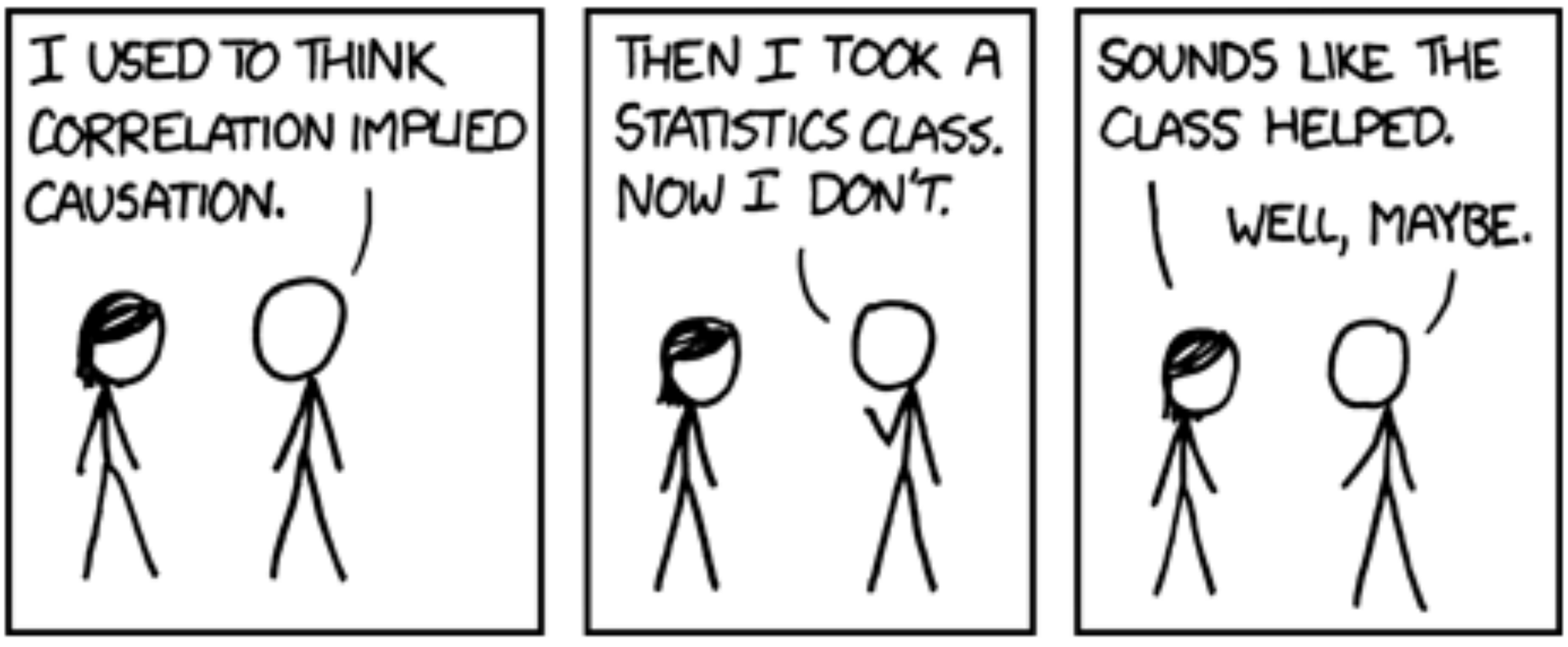
Causality and Machine Learning

Kun Zhang
Mingming Gong

Carnegie Mellon University

Causality vs. Dependence

- Causality ➜ dependence ! Dependence ➜ causality



(http://imgs.xkcd.com/comics/correlation.png)

X and Y are associated iff

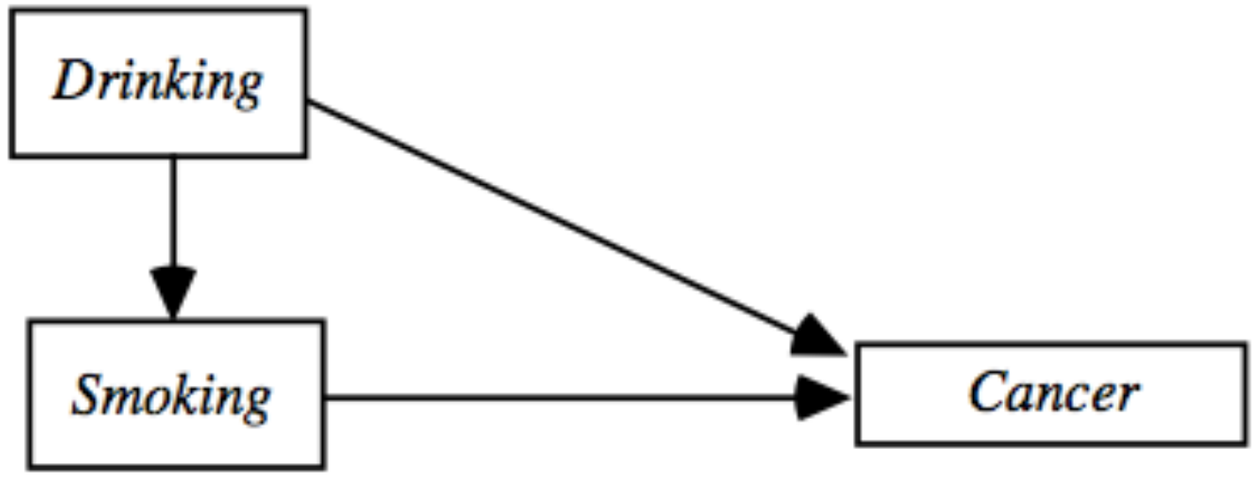
$\exists x_1 \neq x_2 \; P(Y|X=x_1) \neq P(Y|X=x_2)$

X is a cause of Y iff

$\exists x_1 \neq x_2 \; P(Y|do\,(X=x_1)) \neq P(Y|do\,(X=x_2))$

Representing Causal Relations with Directed Graphs

- A directed graph represents a causally sufficient causal structure



(adapted from "Causation, Prediction, and Search" by SGS, 1995)

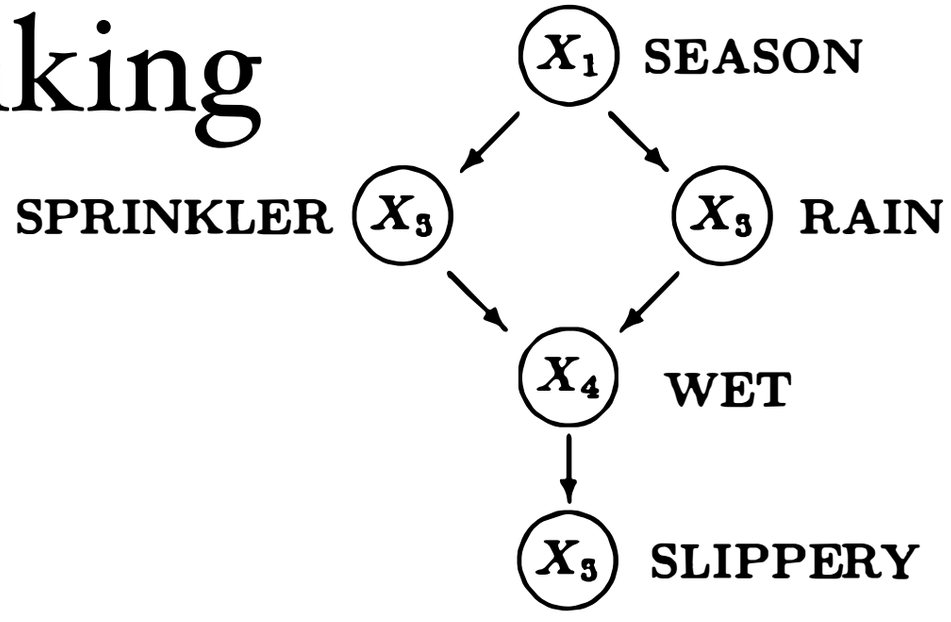
- Directed edge from A to B means A is a direct cause of B relative to the given variable set $\boldsymbol{V}$

Motivating Example: Simpson's Paradox

	Treatment A	Treatment B
Small Stones	*Group 1* **93% (81/87)**	*Group 2* 87% (234/270)
Large Stones	*Group 3* **73% (192/263)**	*Group 4* 69% (55/80)
Both	78% (273/350)	**83% (289/350)**

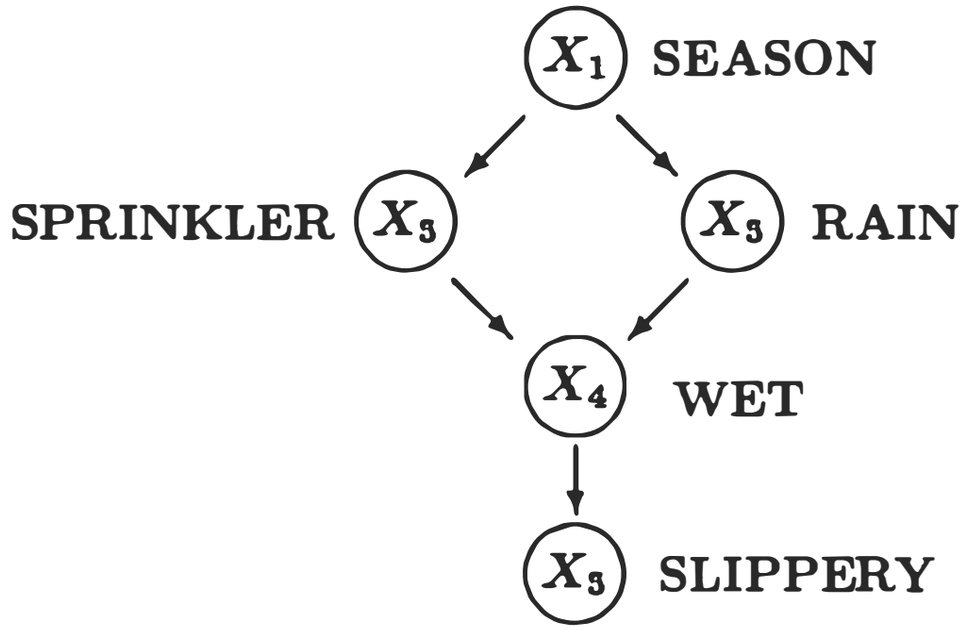
(from wikipedia)

Conditioning, Manipulating, and Counterfactual Thinking



- Three questions:
- **Prediction**: Would the pavement be slippery if we *find* the sprinkler off?
- **Intervention**: Would the pavement be slippery if we *make sure* that the sprinkler is off?
- **Counterfactual**: Would the pavement be slippery *had* the sprinkler been off, *given that the pavement is in fact not slippery and the sprinkler is on*?

Conditioning vs. Manipulating



- Three questions:
  - **Prediction**: Would the pavement be slippery if we *find* the sprinkler off?

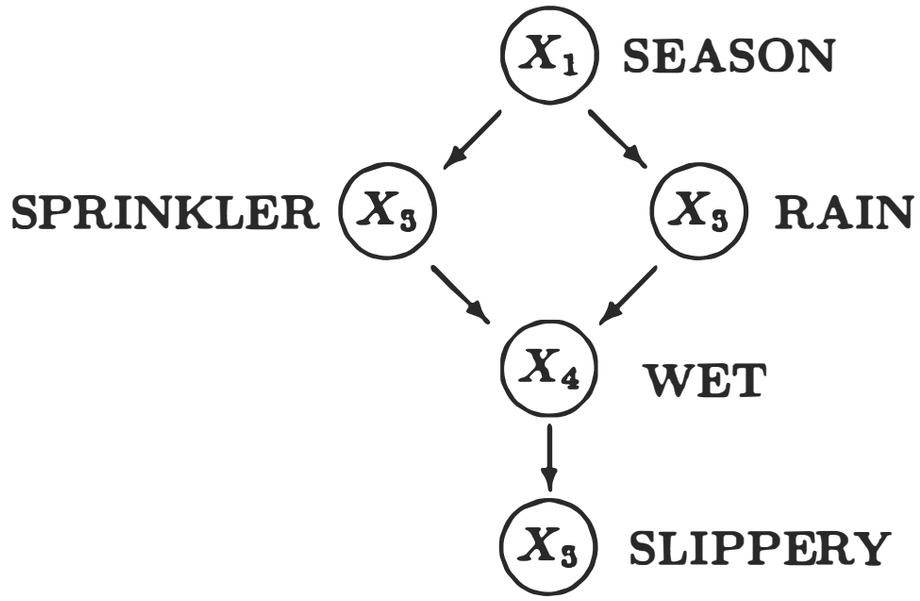
P(Slippery | Sprinkler=off)

  - **Intervention**: Would the pavement be slippery if we *make sure* that the sprinkler is off?

P(Slippery | do(Sprinkler=off))

  - **Counterfactual**: Would the pavement be slippery *had* the sprinkler been off, *given that the pavement is in fact not slippery and the sprinkler is on*?

Counterfactural Reasoning



- Three questions:
  - **Prediction**: Would the pavement be slippery if we *find* the sprinkler off?

$$P(Slippery \mid Sprinkler{=}off)$$

  - **Intervention**: Would the pavement be slippery if we *make sure* that the sprinkler is off?

$$P(Slippery \mid do(Sprinkler{=}off))$$

  - **Counterfactual**: Would the pavement be slippery *had* the sprinkler been off, *given that the pavement is in fact not slippery and the sprinkler is on*?

$$P(Slippery_{\,Sprinkler=off} \mid Sprinkler = on, Slippery = no)$$

Artificial "Intelligence"

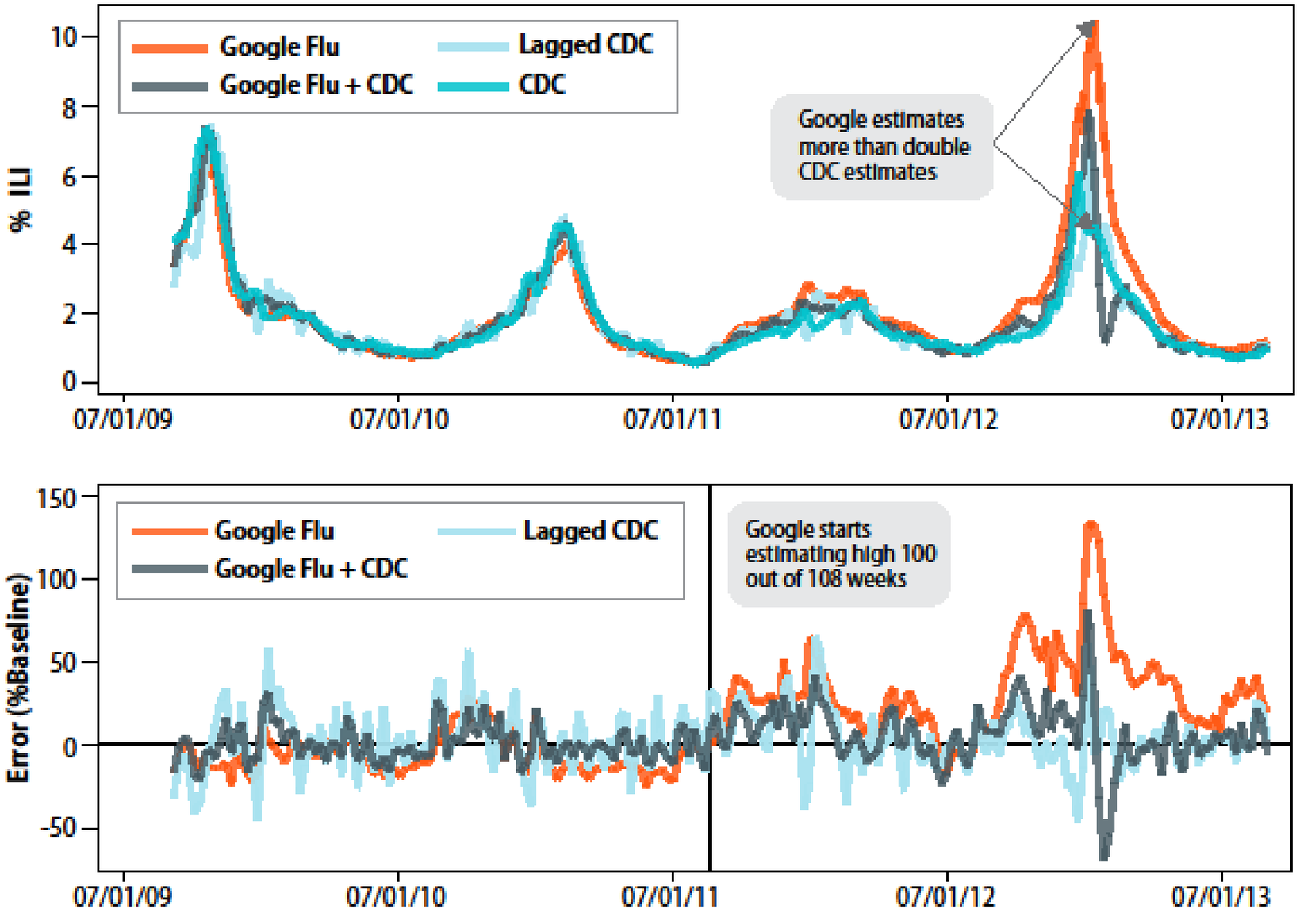
- Traditional machine learning usually assumes a fixed data distribution; tries to avoid overfitting
- Is deep learning "intelligent"?

- Intelligence: understanding; control/intervention; decomposability; information fusion, learning with few examples, extrapolation

Problem with Google Flu Trends

- Google flu trends (Nature, 2008; Science, 2014)
  - Based on frequency of queries including “fever”, “cough”, etc.
- Wrong weighting? How?

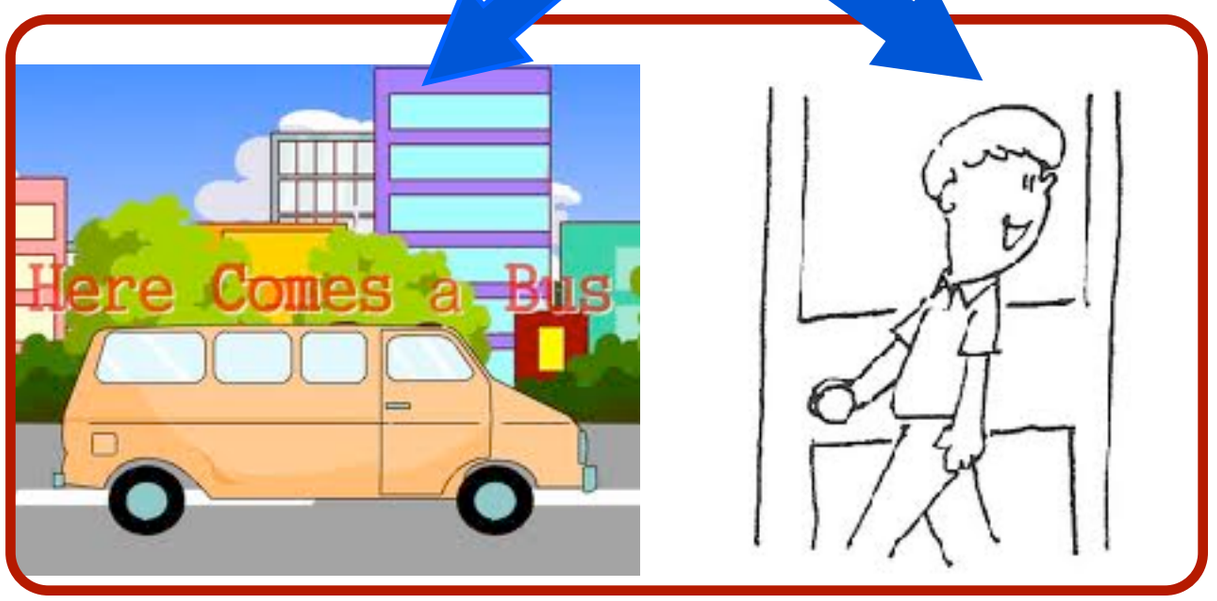
GFT overestimated the prevalence of flu



Classic Ways to Find Causal Information

- What if you manipulate *X* and see *Y* also *changes*?
  - A manipulation directly changes only the target variable *X*
- Manipulate vs. change

Bus timetable



Causal Discovery from Data: Examples

The Telegraph

Home | News | World | Sport | Finance | Comment | Culture | Travel | Life | Women | Fashion | Lu

USA | Asia | China | Europe | Middle East | Australasia | Africa | South America | Central Asia

France | Francois Hollande | Germany | Angela Merkel | Russia | Vladimir Putin | Greece | Spa

HOME » NEWS » WORLD NEWS » EUROPE

Couples who share the housework are more likely to divorce, study finds

Divorce rates are far higher among "modern" couples who share the housework than in those where the woman does the lion's share of the chores, a Norwegian study has found.

Causal Discovery from Data: Examples

Example 1: Conditional Independence

Science AAAS March, 2014

Large-Scale Psychological Differences Within China Explained by Rice Versus Wheat Agriculture

T. Talhelm,[1]* X. Zhang,[2,3] S. Oishi,[1] C. Shimin,[4] D. Duan,[2] X. Lan,[5] S. Kitayama[5]

Cross-cultural psychologists have mostly contrasted East Asia with the West. However, this study shows that there are major psychological differences within China. We propose that a history of farming rice makes cultures more interdependent, whereas farming wheat makes cultures more independent, and these agricultural legacies continue to affect people in the modern world. We tested 1162 Han Chinese participants in six sites and found that rice-growing southern China is more interdependent and holistic-thinking than the wheat-growing north. To control for confounds like climate, we tested people from neighboring counties along the rice-wheat border and found differences that were just as large. We also find that modernization and pathogen prevalence theories do not fit the data.

Over the past 20 years, psychologists have cataloged a long list of differences be- more insular and collectivistic (*6*). Studies have found that historical pathogen prevalence

founded with rice—a possibility that prior research did not control for.

X: rice/wheat agriculture;
Y: culture;
Z: climate etc.:

$X \not\perp\!\!\!\perp Y$;
$X \not\perp\!\!\!\perp Y \mid Z$.

Under what conditions can we say $X \rightarrow Y$?

subsistence crops—rice and wheat—are very dif-

Example 2: Conditional Independence

Thanks to collaborator Marlijn Noback

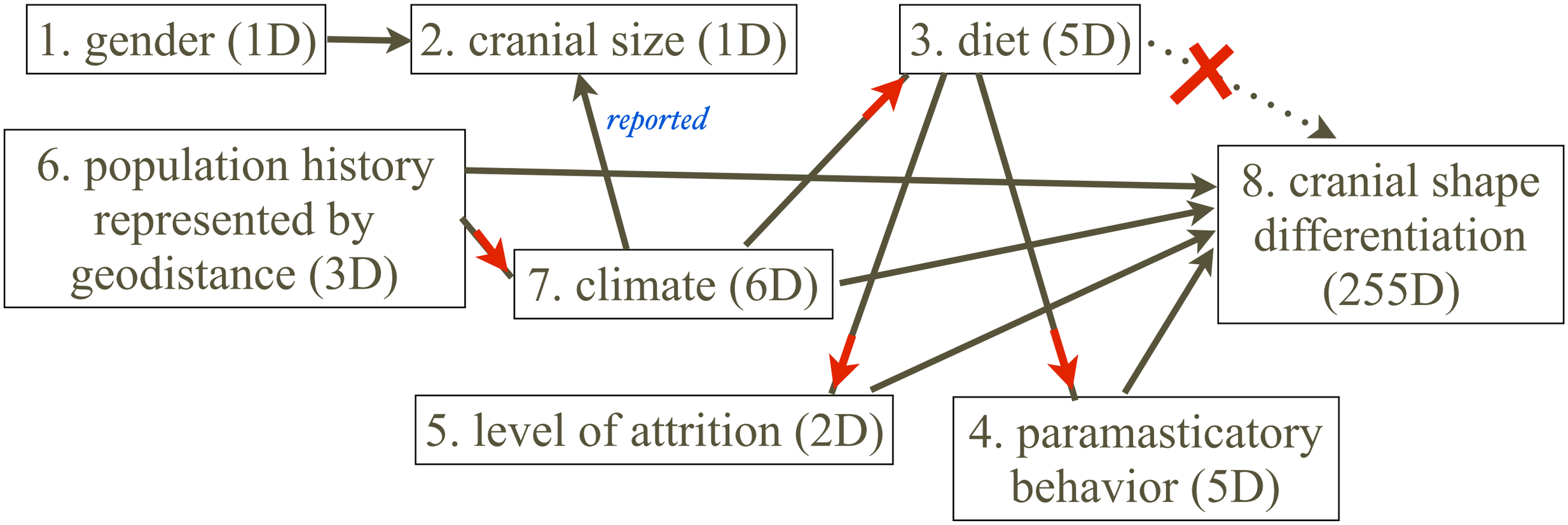
- 8 variables of 250 skeletons collected from different locations
- Different dimensions (from 1 to 255) with nonlinear dependence

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Id	Population	Sex	Cranial size	Diet or subsistence					Paramastic	Dental wear		Geographic location per population			Climate per population					
2			(Male, fem	(Centroid S	Gathering	Hunting	Fishing	Pastoralism	Agriculture	Yes=1, no=0	Average att	Attrition pc	Distance to	Longitude	Latitude	Tmean	Tmin	Tmax	Vpmean	Vpmin	Vpmax
3	AINU31_1	Ainu	Unknown	713.2942	2	3	4	0	1	0	1.5	2	16464	43.548548	142.639159	2.86	-11.19	17.01	7.43	2.27	16.83
4	AINU7_1	Ainu	Unknown	676.148	2	3	4	0	1	0	1.5	1	16464	43.548548	142.639159	2.86	-11.19	17.01	7.43	2.27	16.83
5	AINU7_2	Ainu	Unknown	675.4924	2	3	4	0	1	0	1.5	1	16464	43.548548	142.639159	2.86	-11.19	17.01	7.43	2.27	16.83
6	AINU_1016	Ainu	Male	684.3304	2	3	4	0	1	0	1.5	2.5	16464	43.548548	142.639159	2.86	-11.19	17.01	7.43	2.27	16.83
7	AINU_1016	Ainu	Female	686.285	2	3	4	0	1	0	1.5	4	16464	43.548548	142.639159	2.86	-11.19	17.01	7.43	2.27	16.83
8	AUSM245	Australia	Male	673.8749	6	4	0	0	0	1	2.5	1	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
9	AUSM246	Australia	Male	647.4586	6	4	0	0	0	1	2.5	4	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
10	AUSM8217	Australia	Male	658.6616	6	4	0	0	0	1	2.5	2	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
11	AUSM8177	Australia	Male	667.5444	6	4	0	0	0	1	2.5	4	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
12	AUSM8179	Australia	Male	629.7138	6	4	0	0	0	1	2.5	3.5	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
13	AUSM8173	Australia	Male	648.7064	6	4	0	0	0	1	2.5	3.5	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
14	AUSM8171	Australia	Male	643.0378	6	4	0	0	0	1	2.5	2	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
15	AUSM8165	Australia	Male	616.55	6	4	0	0	0	1	2.5	3.5	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
16	AUSM8154	Australia	Male	635.0605	6	4	0	0	0	1	2.5	2	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
17	AUSM8153	Australia	Male	650.6959	6	4	0	0	0	1	2.5	3	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
18	AUSF1412	Australia	Female	618.4781	6	4	0	0	0	1	2.5	1	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
19	AUSF8179	Australia	Female	634.3122	6	4	0	0	0	1	2.5	3.5	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
20	AUSF8175	Australia	Female	605.1759	6	4	0	0	0	1	2.5	1.5	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
21	AUSF8172	Australia	Female	613.8324	6	4	0	0	0	1	2.5	3	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
22	AUSF8169	Australia	Female	619.1206	6	4	0	0	0	1	2.5	2.5	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
23	AUSF8157	Australia	Female	628.2819	6	4	0	0	0	1	2.5	2	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
24	AUSF8155	Australia	Female	628.4609	6	4	0	0	0	1	2.5	3.5	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
25	AUSF1578	Australia	Female	640.6311	6	4	0	0	0	1	2.5	2	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
26	AUSF243	Australia	Female	606.164	6	4	0	0	0	1	2.5	2.5	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
27	AUSF8158	Australia	Female	631.6258	6	4	0	0	0	1	2.5	2	20164	-24.287027	135.615234	22.46	13.33	30.27	11.10	7.55	15.96
28	DENM1432	Denmark	Male	663.6198	0	0	1	3	6	0	2.1	2	10440	55.717055	11.711426	8.01	-0.02	16.66	9.67	5.59	15.27
29	DENM1011	Denmark	Male	651.4847	0	0	1	3	6	0	2.1	3	10440	55.717055	11.711426	8.01	-0.02	16.66	9.67	5.59	15.27
30	DENM1205	Denmark	Male	636.9831	0	0	1	3	6	0	2.1	1.5	10440	55.717055	11.711426	8.01	-0.02	16.66	9.67	5.59	15.27
31	DENM116_	Denmark	Male	642.9192	0	0	1	3	6	0	2.1	3	10440	55.717055	11.711426	8.01	-0.02	16.66	9.67	5.59	15.27
32	DENM116_	Denmark	Male	646.6609	0	0	1	3	6	0	2.1	2.5	10440	55.717055	11.711426	8.01	-0.02	16.66	9.67	5.59	15.27
33	DENM116_	Denmark	Male	674.9799	0	0	1	3	6	0	2.1	2	10440	55.717055	11.711426	8.01	-0.02	16.66	9.67	5.59	15.27
34	DENM7_77	Denmark	Male	666.53	0	0	1	3	6	0	2.1	2.5	10440	55.717055	11.711426	8.01	-0.02	16.66	9.67	5.59	15.27
35	DENM1_58	Denmark	Male	627.4583	0	0	1	3	6	0	2.1	1.5	10440	55.717055	11.711426	8.01	-0.02	16.66	9.67	5.59	15.27
36	DENM903	Denmark	Male	662.5953	0	0	1	3	6	0	2.1	2	10440	55.717055	11.711426	8.01	-0.02	16.66	9.67	5.59	15.27

Example 2: Conditional Independence

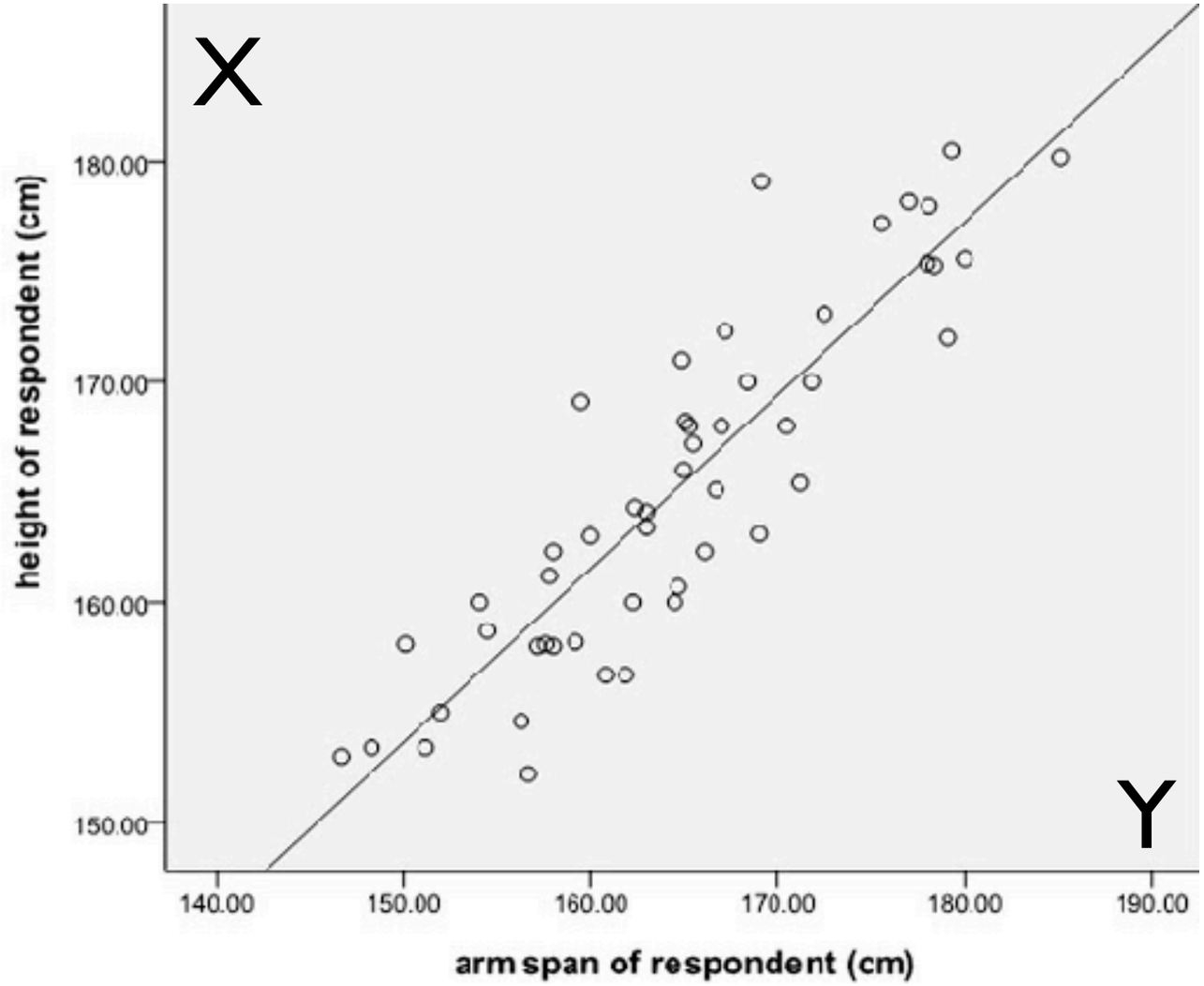
Thanks to collaborator Marlijn Noback

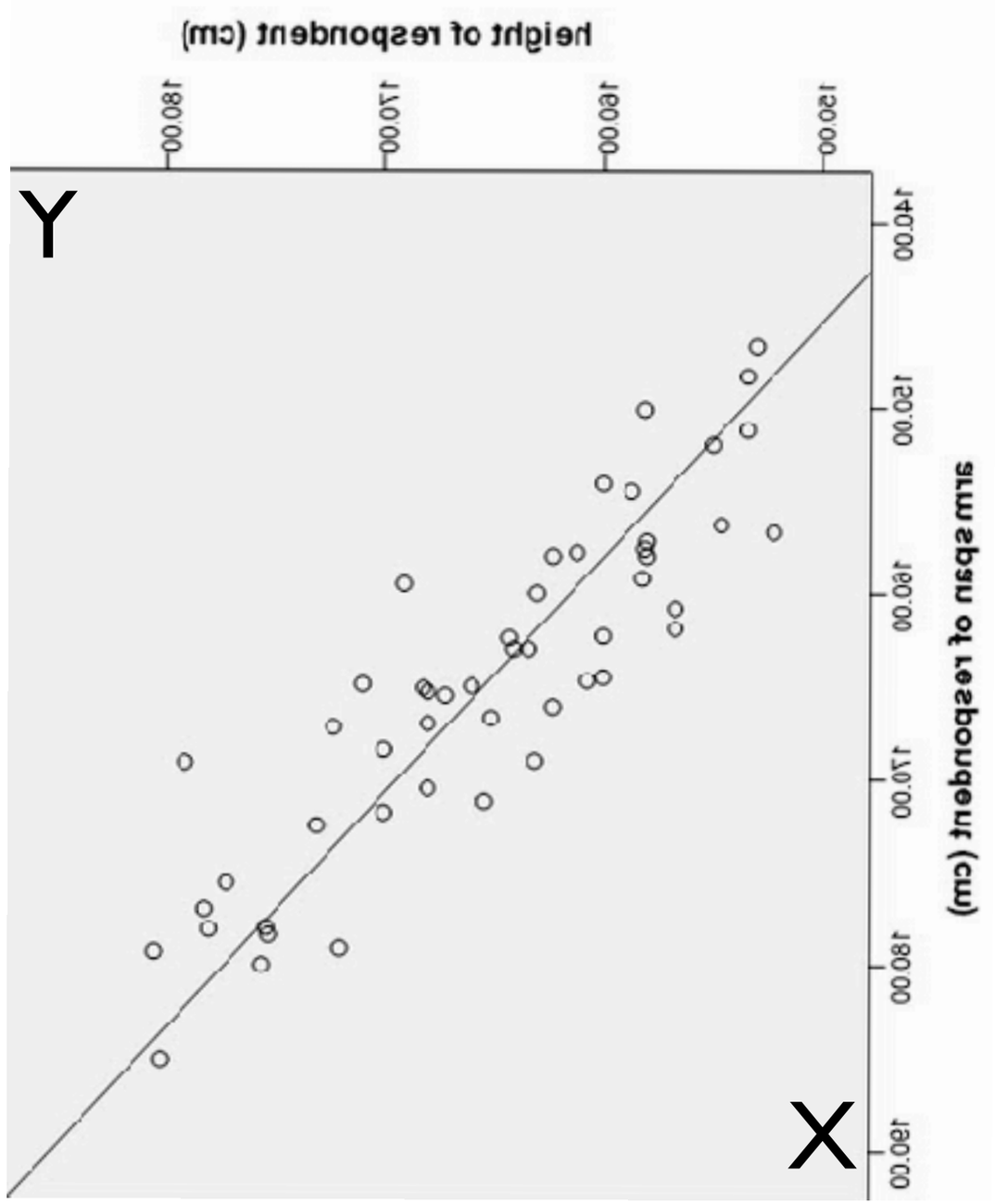
- 8 variables of 250 skeletons collected from different locations
- Different dimensions (from 1 to 255) with nonlinear dependence



Example 3: Linear Non-Gaussian Case

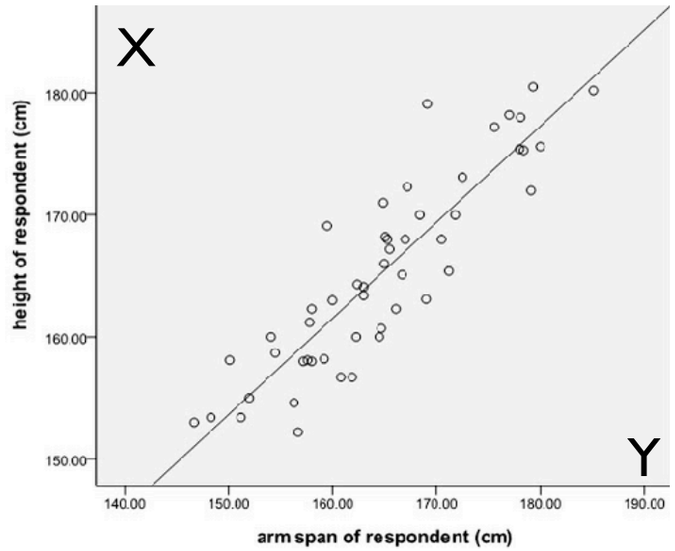
Can we see the causal direction from joint distribution?

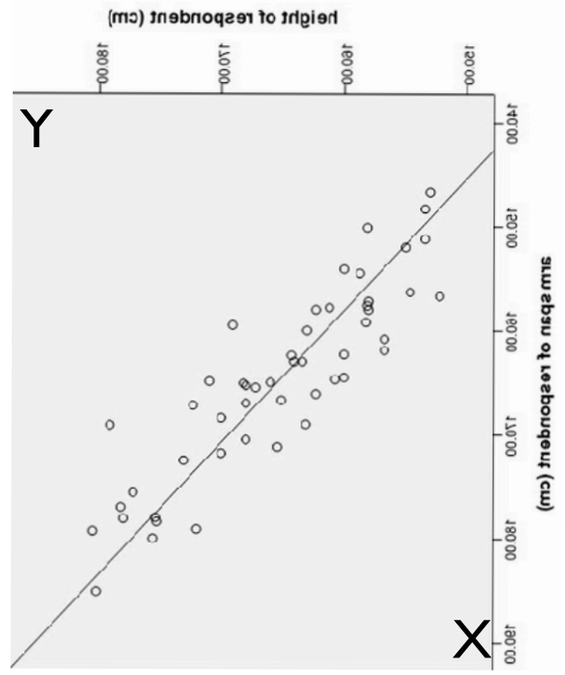


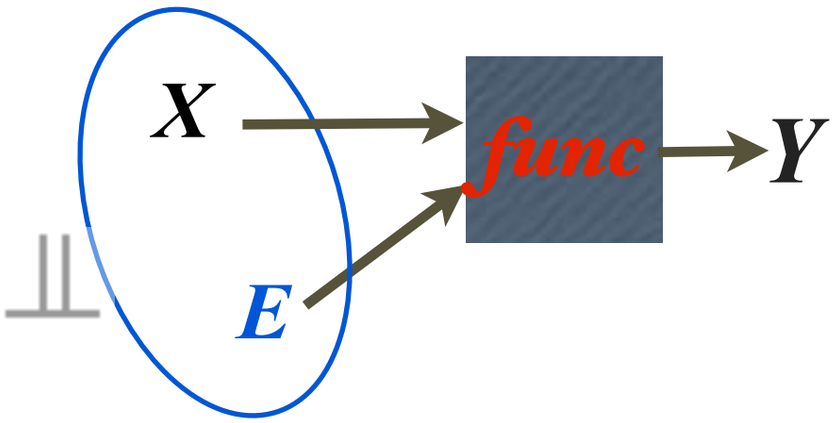


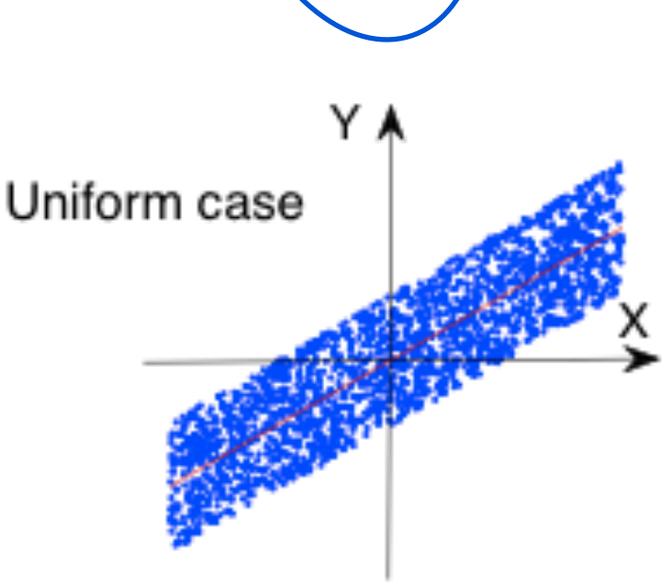
(Figure adapted from http://hs2135statistics.blogspot.de/p/data-analysis.html)

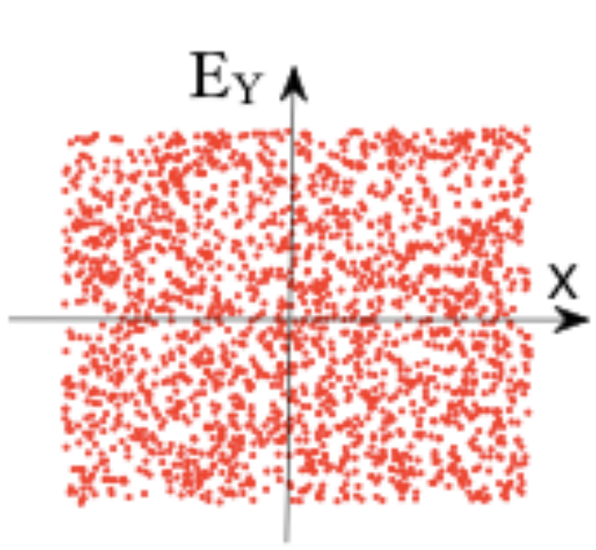
Example 3: Linear Non-Gaussian Case

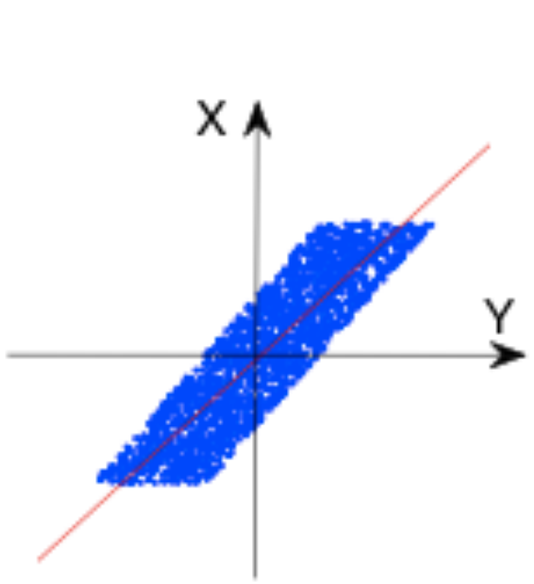


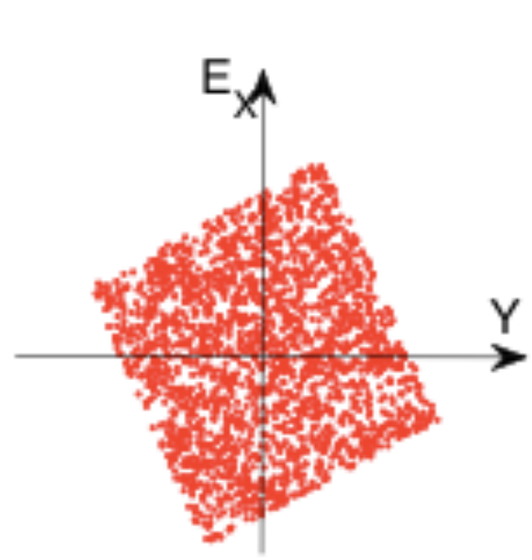






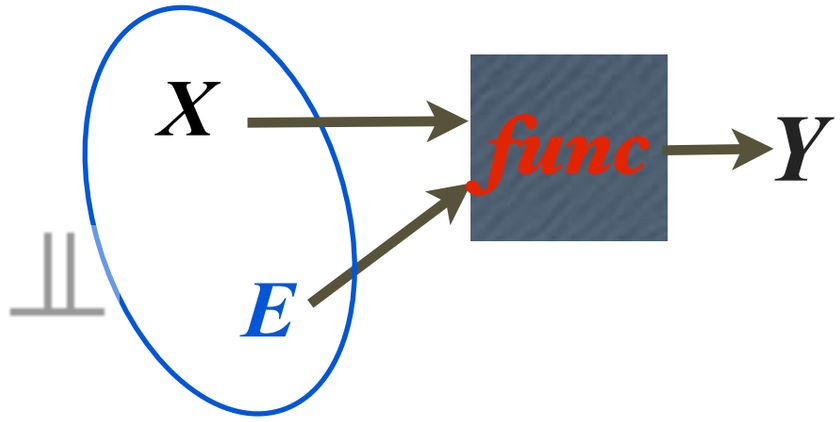


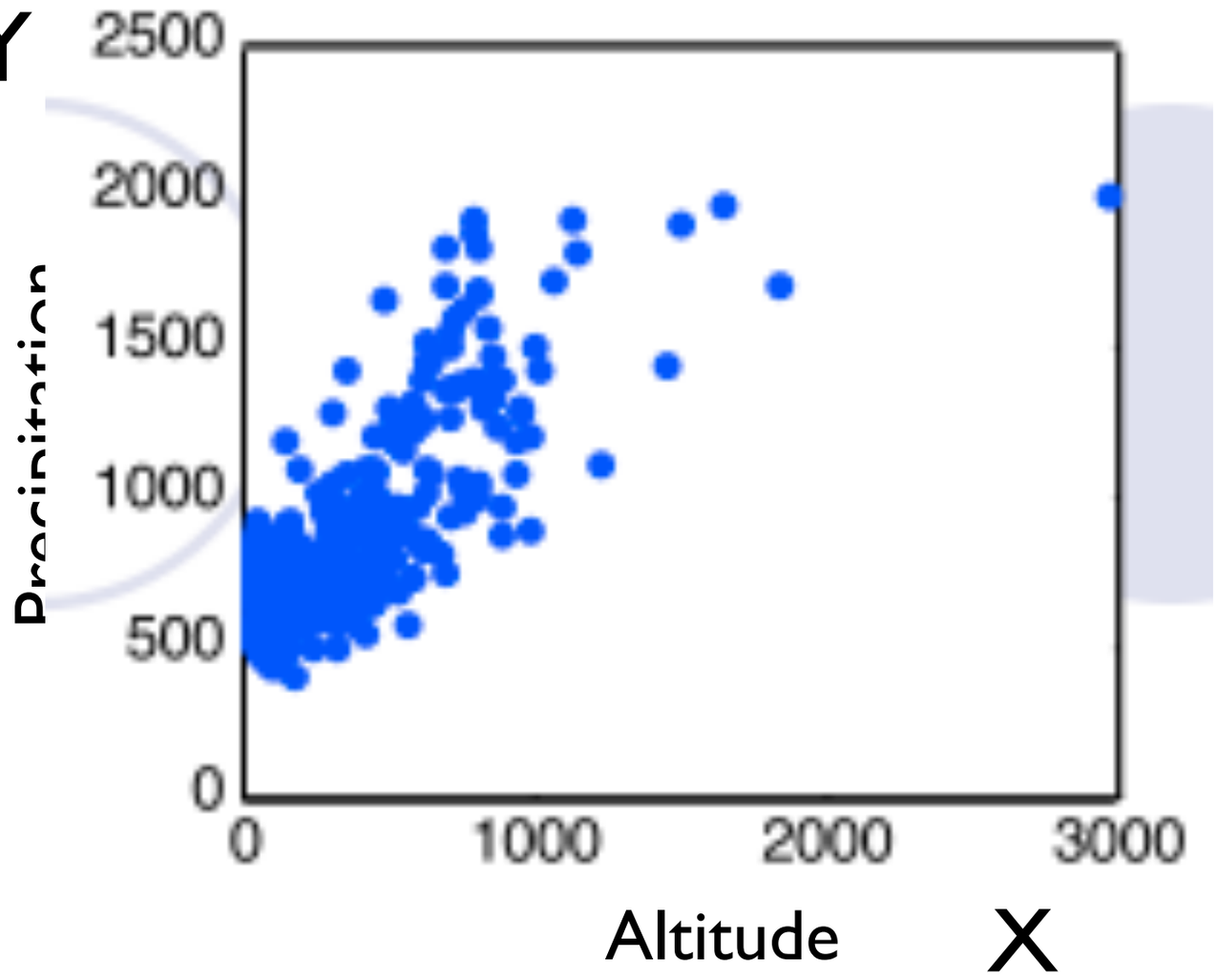




Example 4: Post-Nonlinear Model

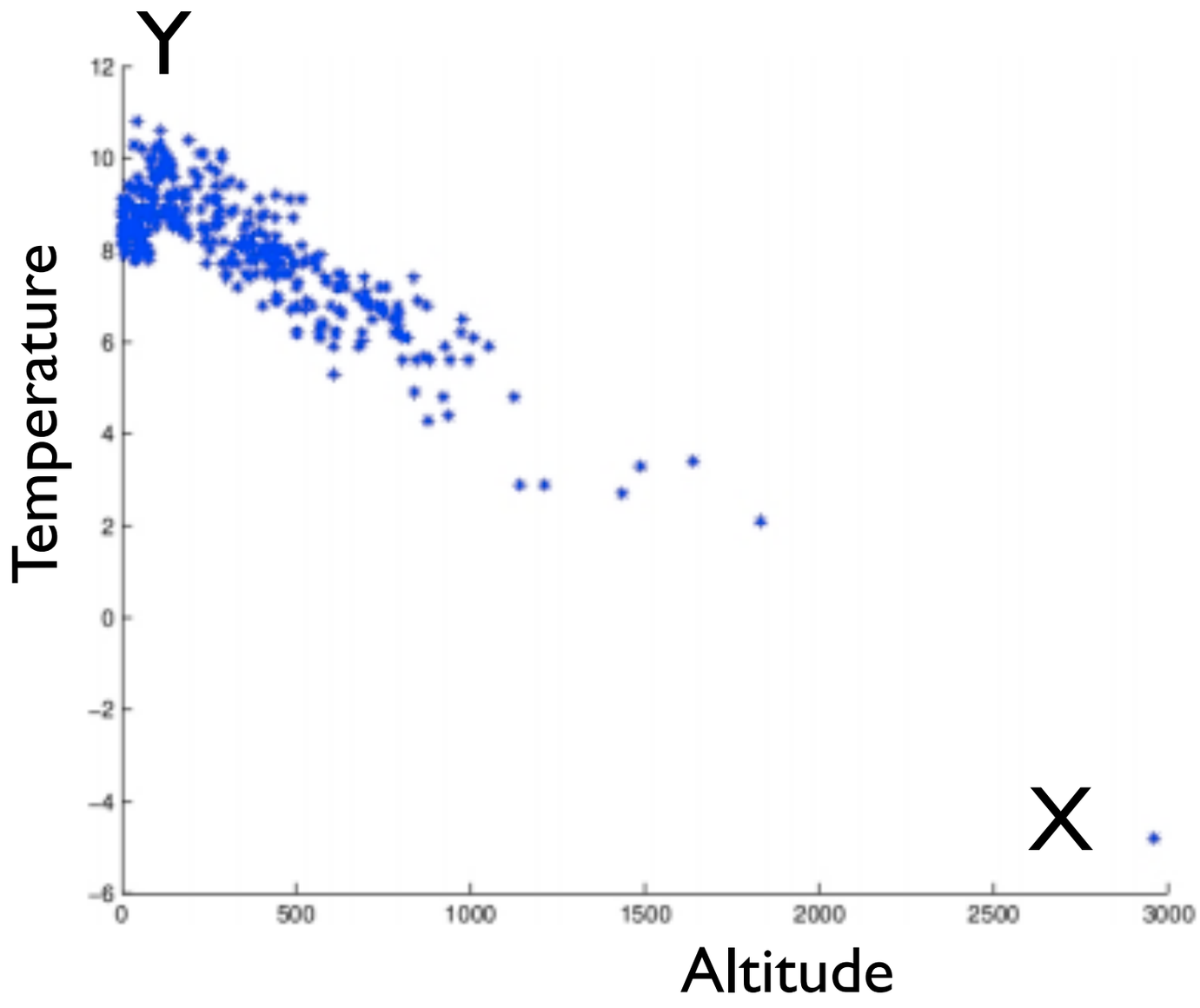
$$Y=f_2\left(f_1(X)+\boldsymbol{E}_Y\right) \text{ with } \boldsymbol{E}_Y \perp\!\!\!\perp \boldsymbol{X}$$



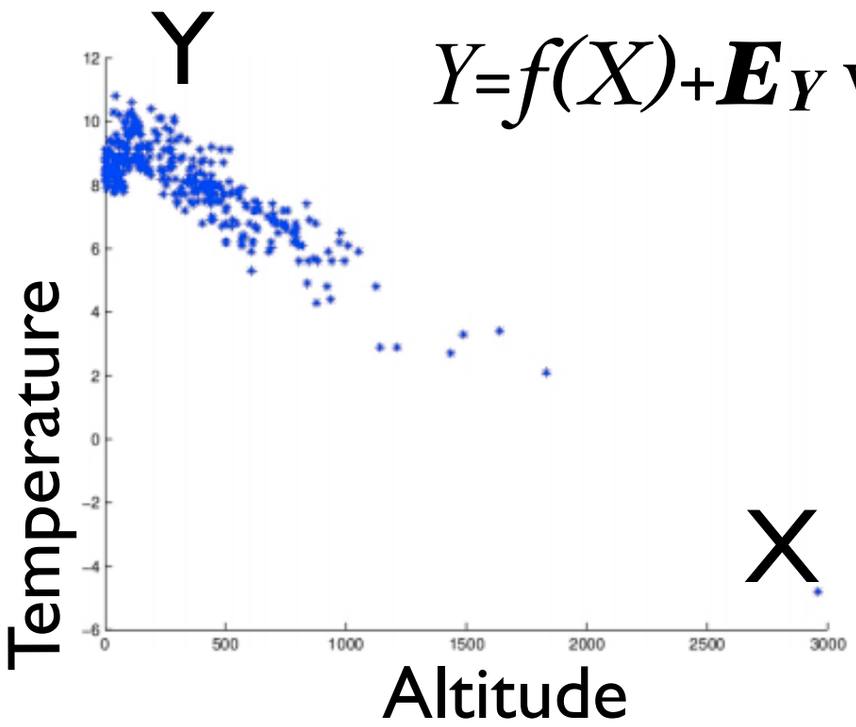


Example 5: Nonlinear Additive Noise

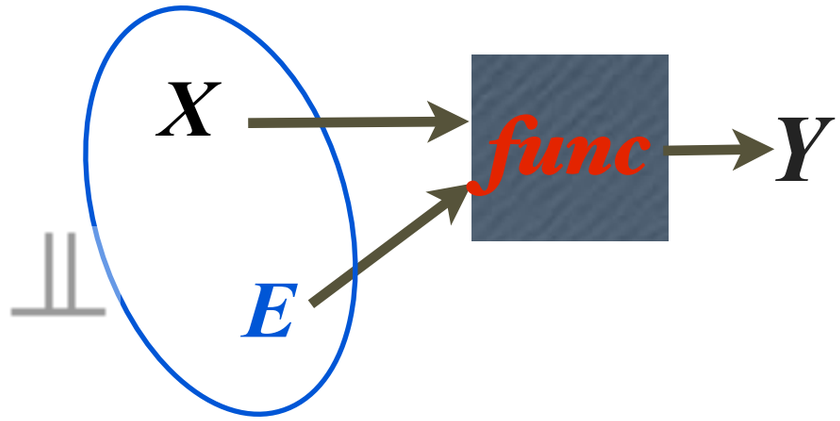
$$Y = f(X) + \boldsymbol{E}_Y \text{ with } \boldsymbol{E}_Y \perp\!\!\!\perp \boldsymbol{X}$$

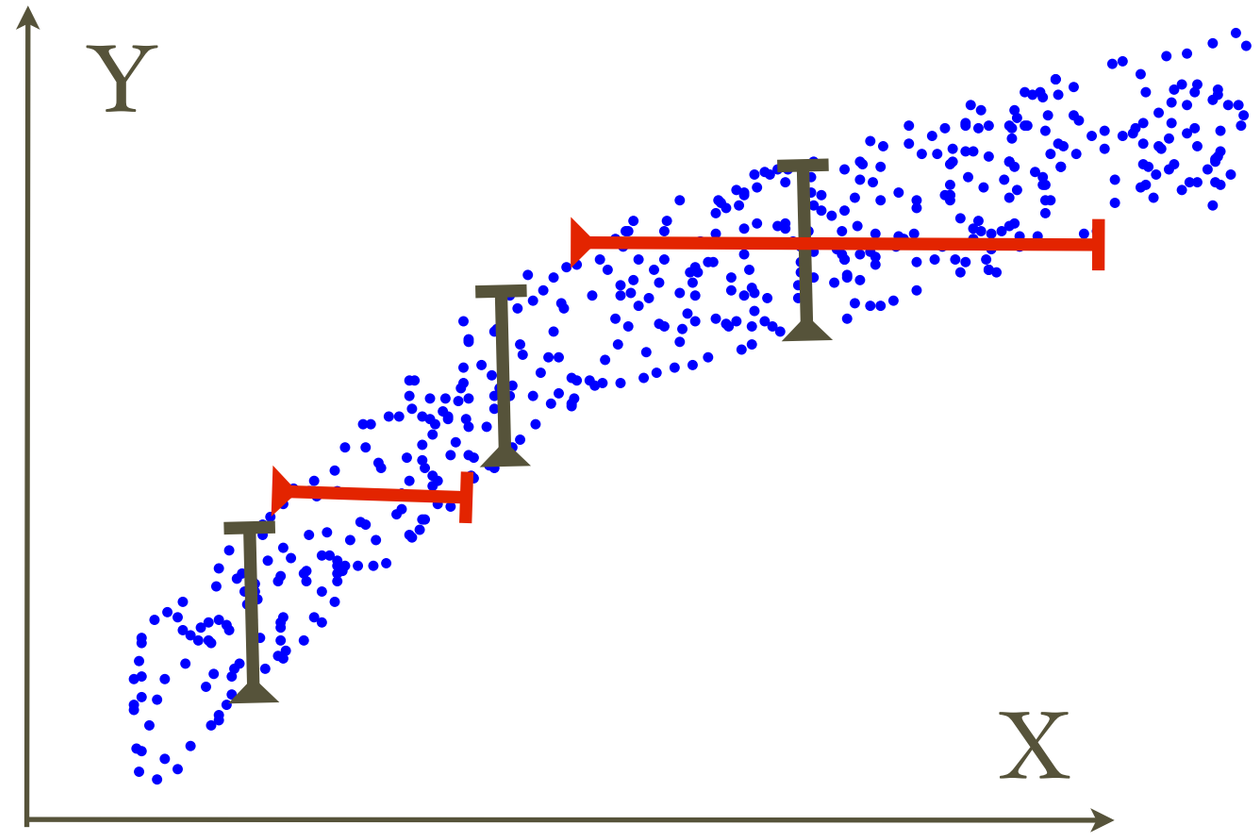


Example 4: Nonlinear additive noise

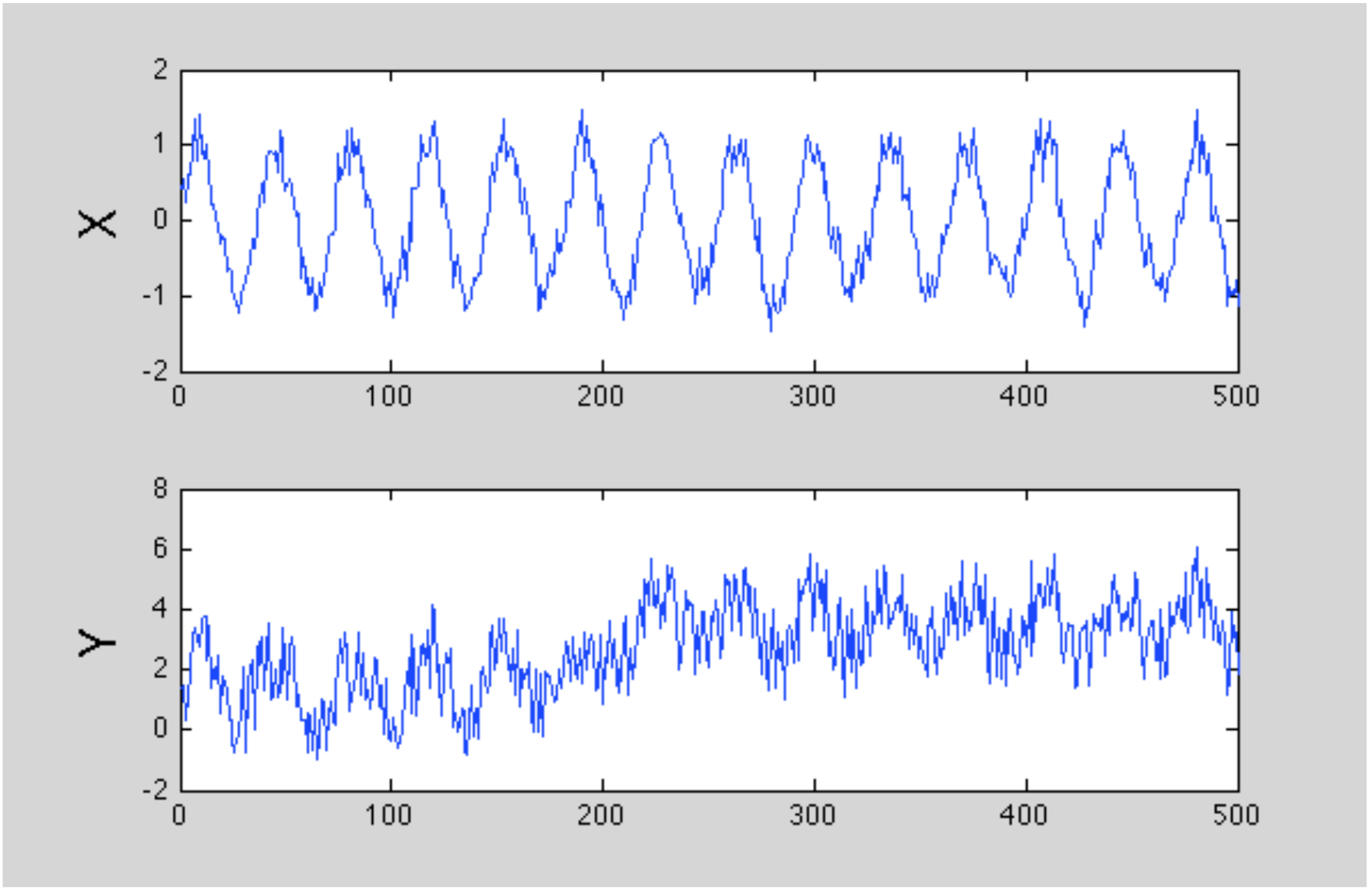


$Y=f(X)+\boldsymbol{E}_Y$ with $\boldsymbol{E}_Y \perp\!\!\!\perp \boldsymbol{X}$

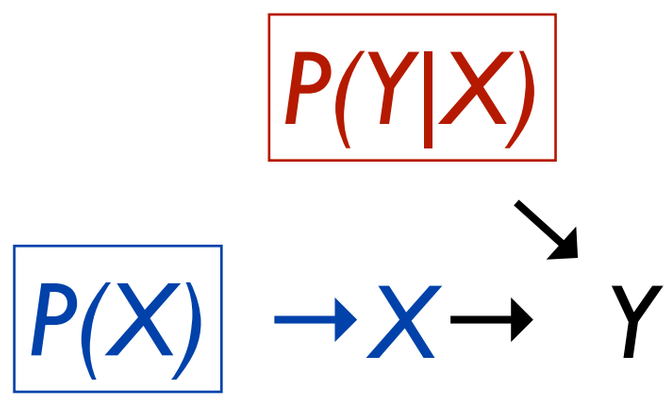




Example 6: From Distribution Changes



Generating process for cause X is irrelevant to that generating effect Y from X: $P(X)$ and $P(Y|X)$ **change independently**



Outline

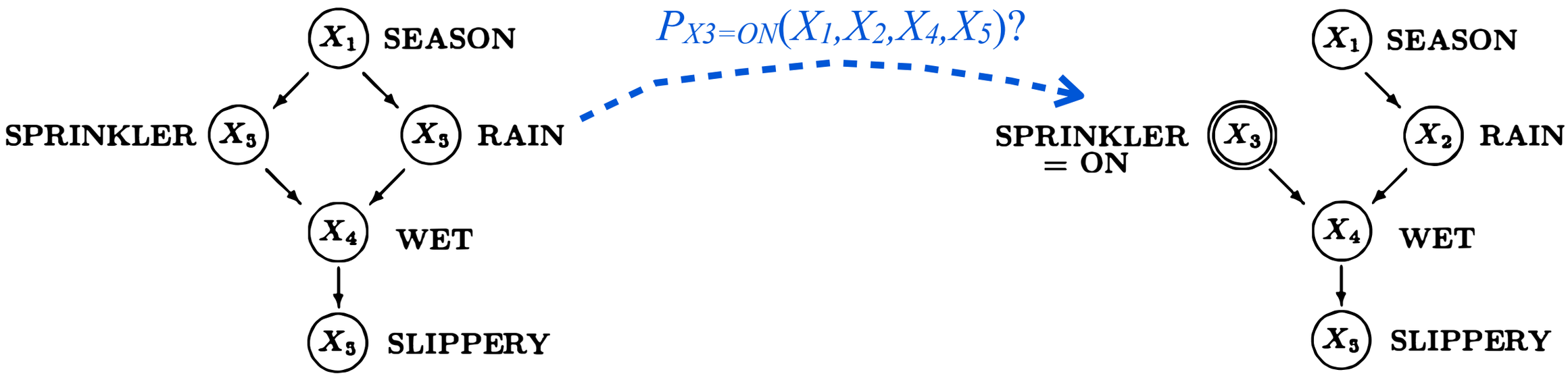
- Causal thinking
- Preliminaries (probability and graphical models)
- Identification of causal effects
- Causal discovery 1: Traditional constraint- or score-based methods

- Causal discovery 2: Linear, non-Gaussian models
- Causal discovery 3: Nonlinear models and noiseless cases
- Causal discovery 4: Practical issues
- Causality-based machine learning

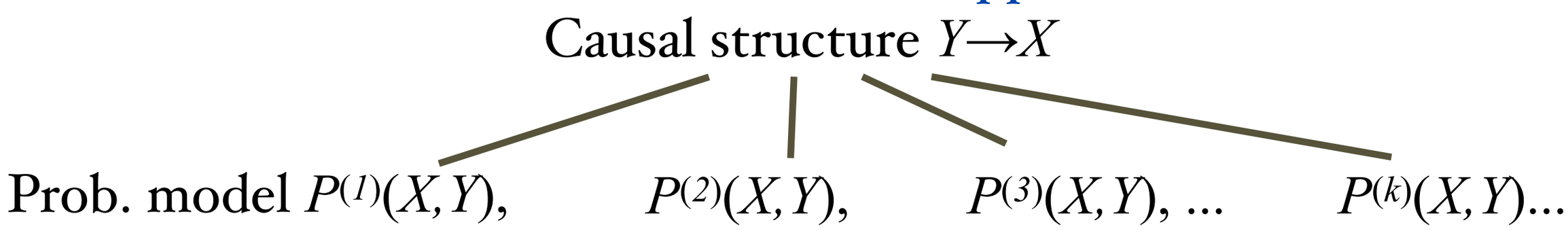
Part I
Causal Thinking

Why Causal Models?

- Infer effect of interventions:

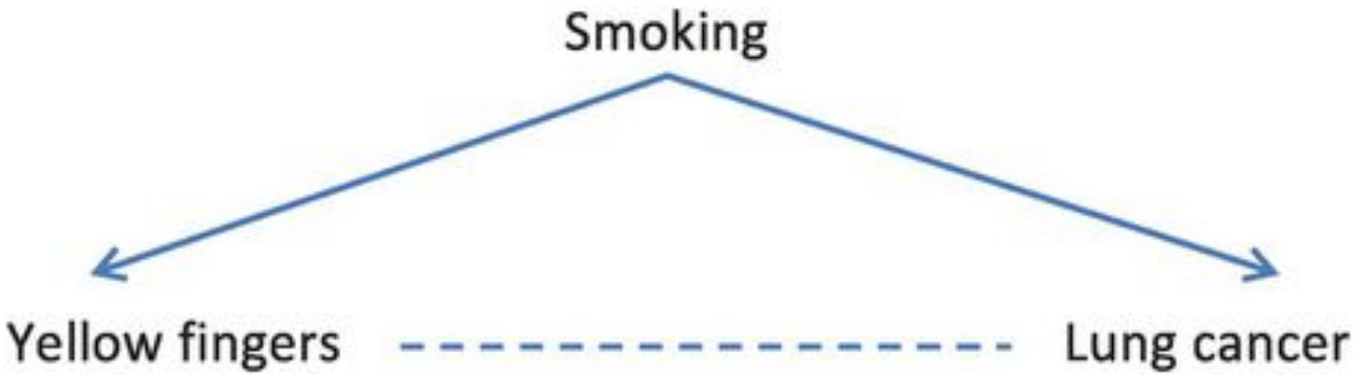


- Causal model: compact description of the properties of the joint distribution
- Derived from one or few distributions; applied to other scenarios



Causal Thinking (1)

- Dependence vs. causality



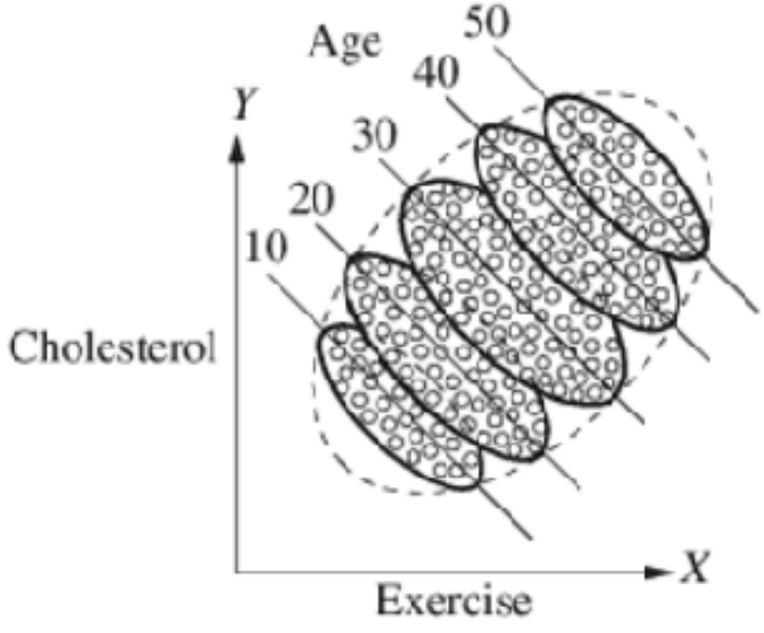
- Simpson's paradox

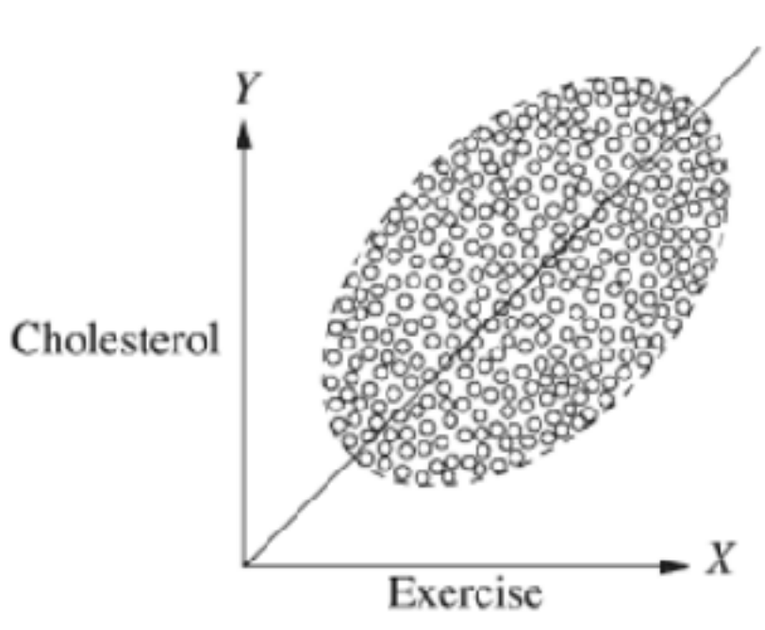
- "Strange" dependence

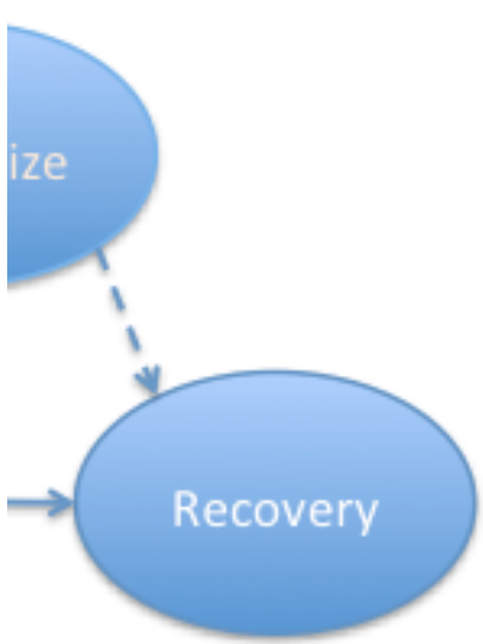
Causal Thinking (2)

- Dependence vs. causality

- Simpson's paradox







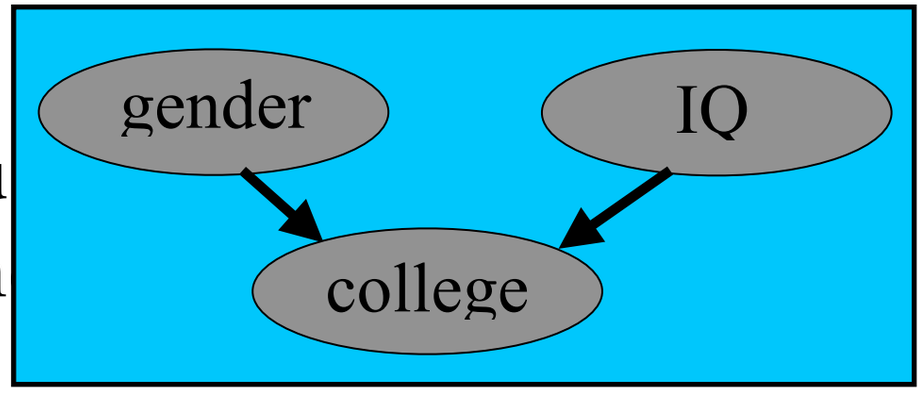
- "Strange" dependence

Causal Thinking (3)

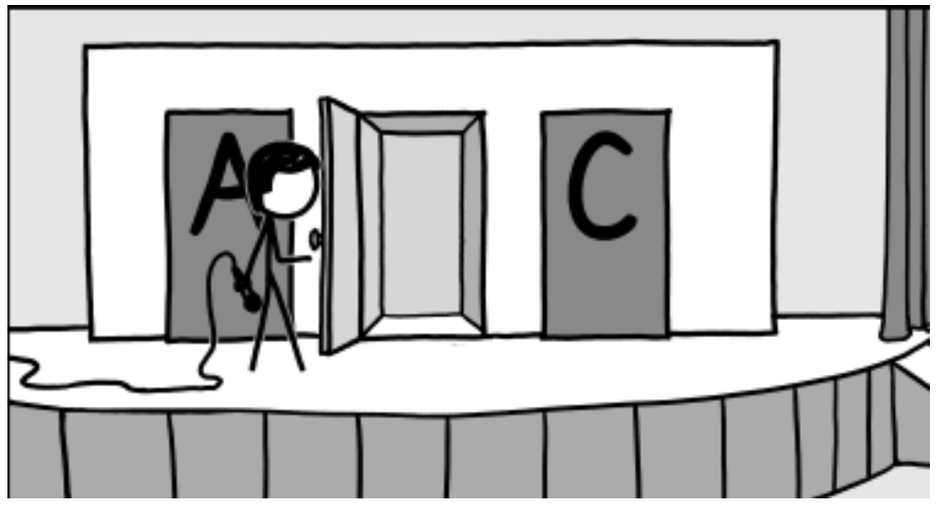
- Dependence vs. causality

- Simpson's paradox

- "Stranger" dependence
  - Let's go back 50 years; maybe you
    students are smarter than male on



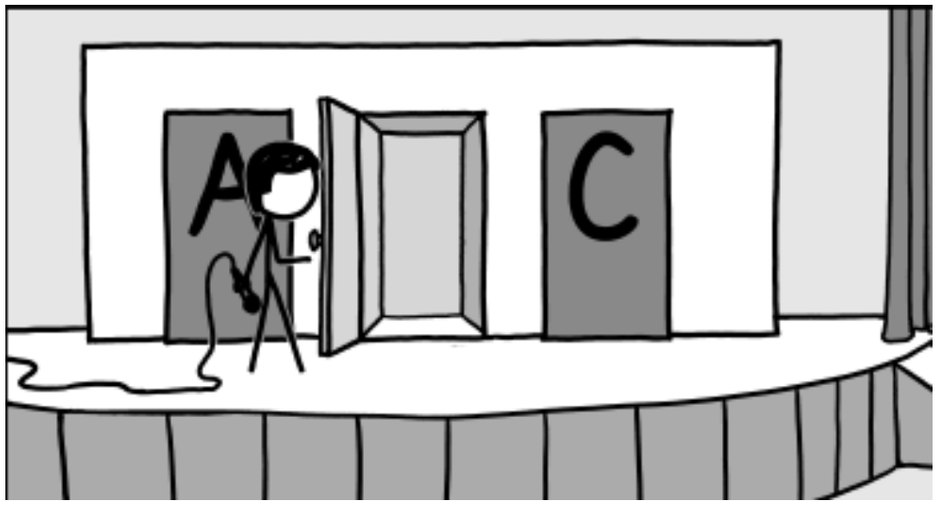
Question 3
Monty Hall Problem



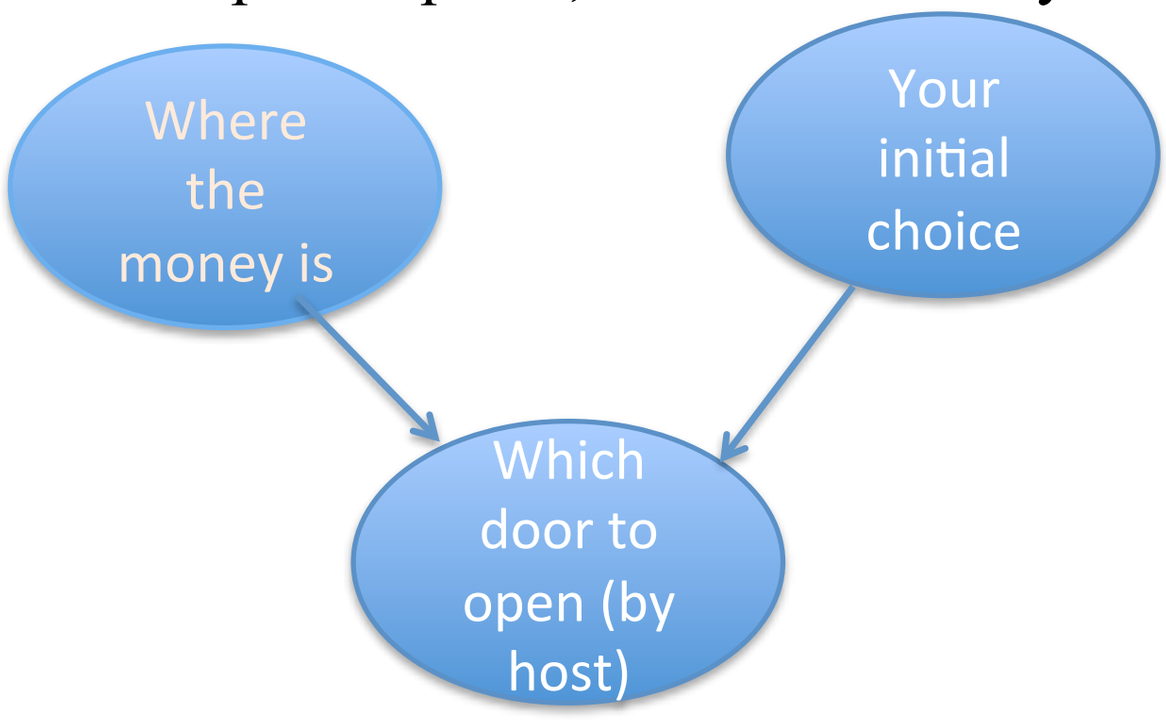
- You are a game show contestant. Before the game begins, the host, Monty Hall, has placed $1,000 dollars behind one of three doors. Nothing is behind the other two doors. The game is played as followed. You, the contestant, choose one of the doors, say, door A. Then Monty opens a door that is not the door you chose and does not have the money behind it, say B. If you want to maximize the expected profit, which door will you finally choose?
  - A
  - C

Excerpt from "The Mind's Arrows"

Question 3
Monty Hall Problem



- You are a game show contestant. Before the game begins, the host, Monty Hall, has placed $1,000 dollars behind one of three doors. Nothing is behind the other two doors. The game is played as followed. You, the contestant, choose one of the doors, say, door A. Then Monty opens a door that is not the door you chose and does not have the money behind it, say B. If you want to maximize the expected profit, which door will you finally choose?
  - A
  - C

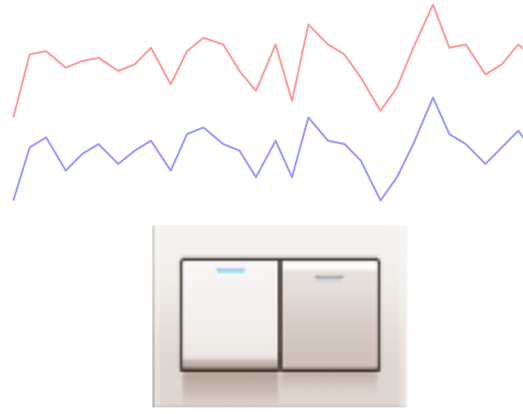


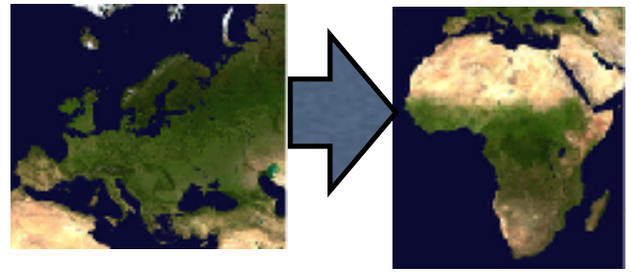
Causal Thinking Makes a Difference

- Active manipulation /control vs. passive prediction
- Generalization / adaptation ability in new environments?
- Integration of causal information: what is the causal model for *X*, *Y*, and *Z* if
  - $X \rightarrow Y$, $Y \rightarrow Z$ (expansion) or $X \rightarrow Z$, $Y \rightarrow Z$ (refinement)...
- Creativity
  - Thoughts consist of the "What if?" and the "If I had only..." + knowledge integration + ...

Use associational or causal knowledge?

- Make passive predictions in stationary environments ?
- Manipulation and control ?
  - E.g., advertisements & recommendation
    - What products to recommend?
- Predictions in non-stationary environments ?





Knowing Effect may Be More Informative