Class 1 (January 14, 2025)

Causality and Machine Learning

(*80-816/516*)

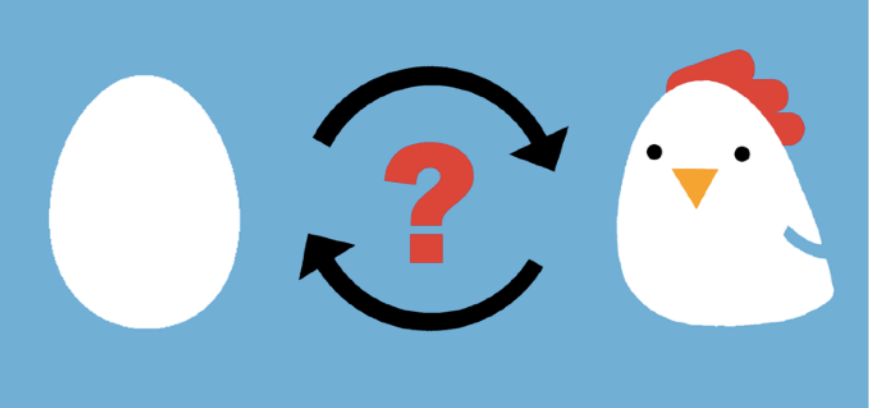
Course time & location: Tuesdays & Thursdays 12:30 – 1:50PM, TEP 1308
https://www.andrew.cmu.edu/course/80-516/

Instructor:

Kun Zhang (kunz1@cmu.edu)

Zoom link: https://cmu.zoom.us/j/8214572323)

Office Hours: W 3:00–4:00PM (on Zoom or in person); other times by appointment



Causality and Machine Learning
(80-816/516)

Classes 1 & 2 (Jan 14 & 16, 2025)

Introduction to Causality: Why, What, and How?

Instructor:

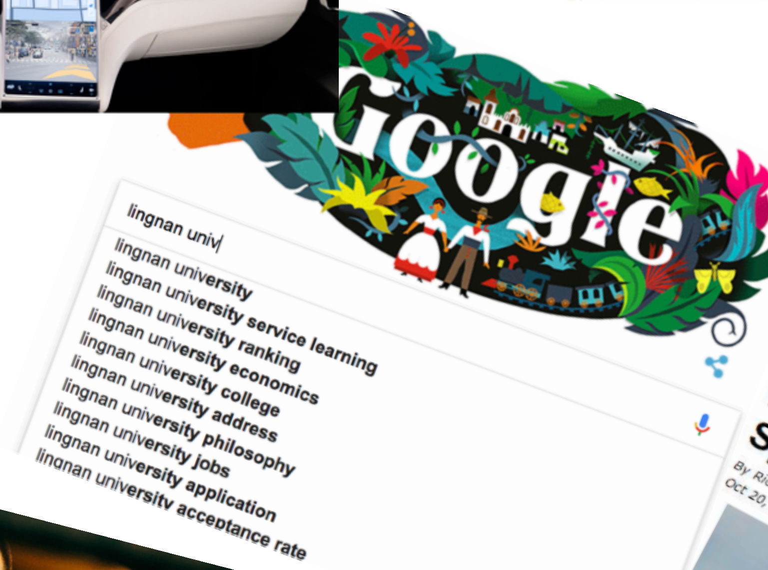
Kun Zhang (kunz1@cmu.edu)

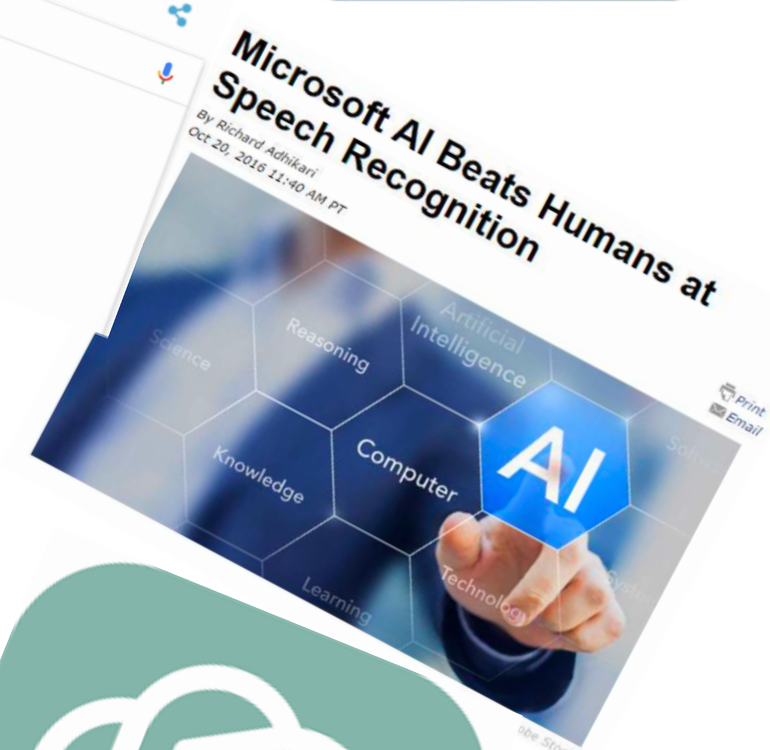
Zoom link: https://cmu.zoom.us/j/8214572323)

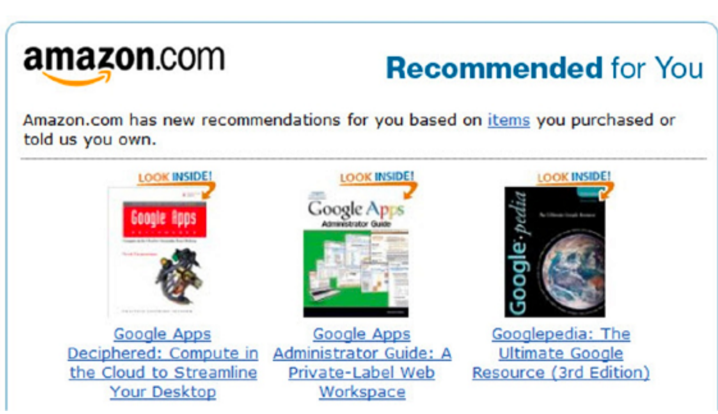
Office Hours: W 3:00–4:00PM (on Zoom or in person); other times by appointment

Grading Policy

- Grading
  - Participation: 5%; in-class discussion: 10%
    - Critique me and each other, please (also, feel free to suggest papers to discuss)
  - 4 homework assignments questions (available on Canvas): 40%
  - Project/essay proposal for each individual or team of two students (due on 03/14, 11:59 pm): 10%
  - Project report/essay (due on 05/02, 11:59 pm): 35%
    - Decide on the topic by 02/28
- We grade undergraduate students on a curve entirely separately from graduate students






amazon.com **Recommended for You**

Amazon.com has new recommendations for you based on items you purchased or told us you own.

Google Apps Deciphered: Compute in the Cloud to Streamline Your Desktop	Google Apps Administrator Guide: A Private-Label Web Workspace	Googlepedia: The Ultimate Google Resource (3rd Edition)

By Geospatial World (https://youtu.be/X-3Oq_82XNA?si=DuW7WRheBetK-OsK)

(Automated) Scientific Discovery: A Story

- Consider puerperal fever in the mid-19th century
- Two clinics used almost the same techniques but had very different mortality rates
- Semmelweis discovered the only major difference was the individuals who worked there
  - *Hypothesis*: Unknown "cadaverous material" caused puerperal fever
  - Proposed *intervention*: washing hands
  - *Conflicted* with the established scientific and medical opinions of the time
  - Rejected by the medical community until years after his death, when Louis Pasteur *confirmed* the germ theory

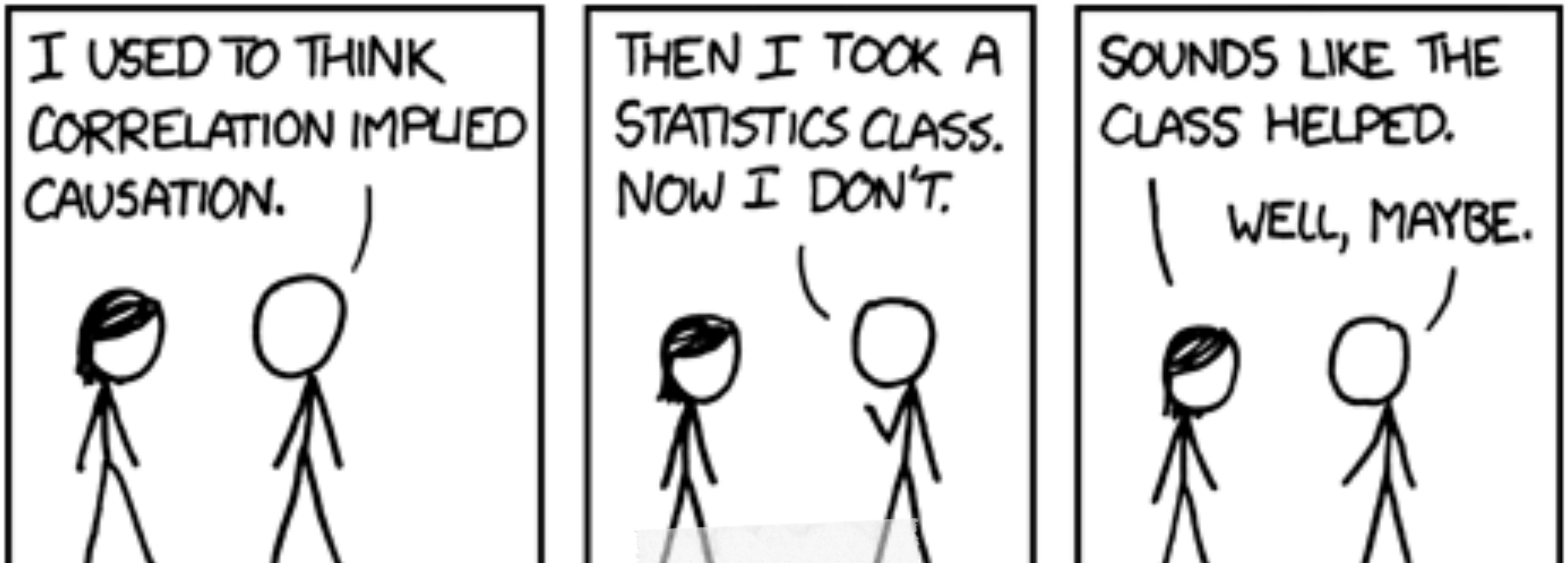
https://amol-kulkarni.com/project/semmelweis/

Ignaz Semmelweis

Semmelweis, aged 42 in 1860, photograph by Borsos and Doctor

Causality vs. Dependence

- Causality ➜ dependence ! Dependence ➜ causality



An intervention on X changes only the target variable X, leaving any other variable unchanged, at least for the moment.

X and Y are associated iff

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

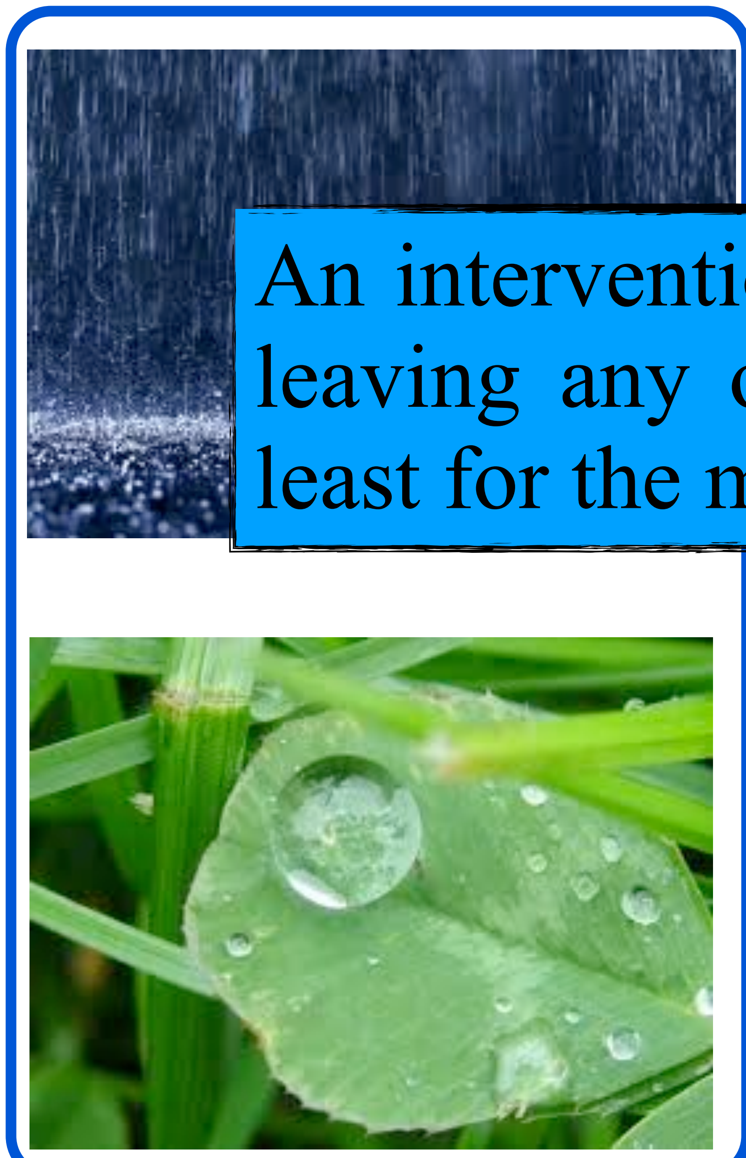
X is a cause of Y iff

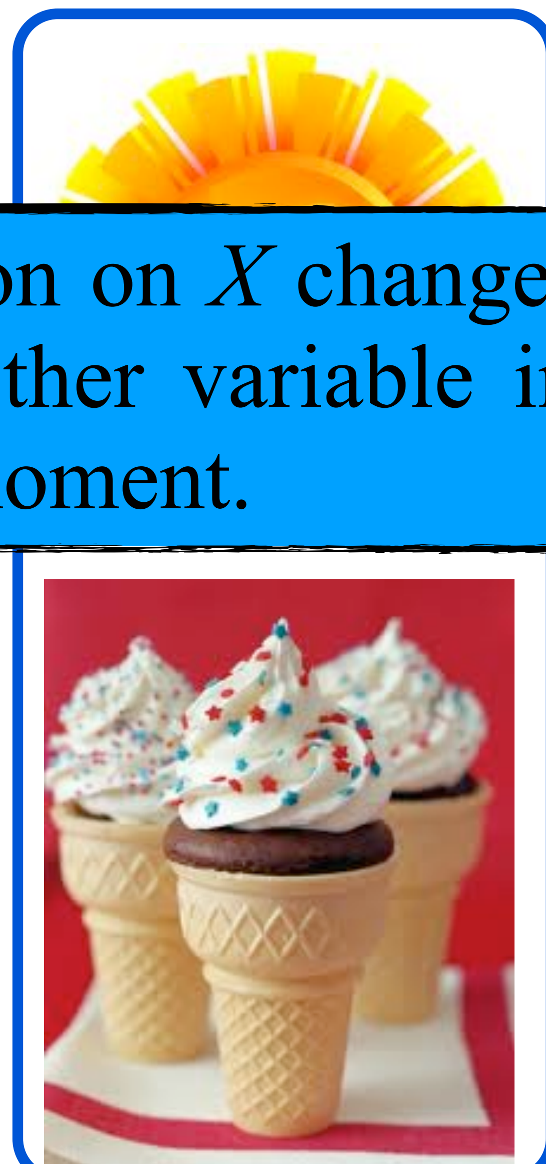
$\exists x_1 \neq x_2 \; P(Y|do\ X=x_1) \neq P(Y|do\ X=x_2)$

intervention

Classic Ways to Find Causal Information (i.i.d. Case)

- What if X and Y are dependent?
- What if you change X and see Y also changes?
  - A **manipulation/ intervention** directly changes only the target variable X





An intervention on X changes only the target variable X, leaving any other variable in the system unchanged, at least for the moment.

** Definition of "interventions"*

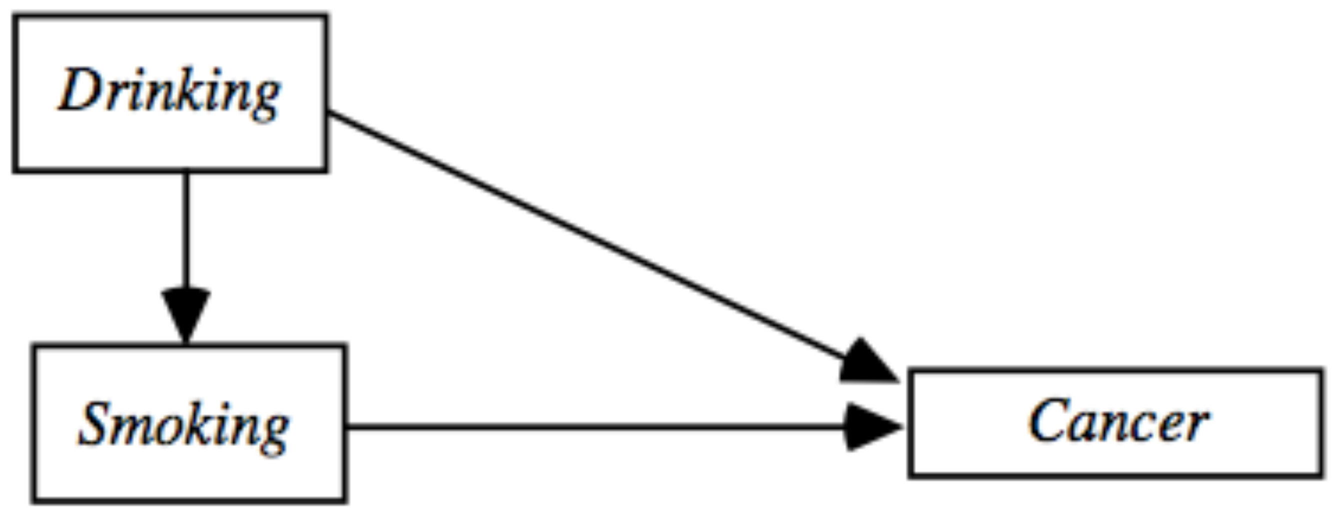
Course Objectives

As an outcome of this course, participants are expected to

- Understand how causality is different from association and why it is useful
- Get familiar with graphical models, causality-related concepts and principles, and emerging approaches to causal discovery or causal representation learning from observational data
- Be acquainted with the state-of-the-art of causality research in different disciplines
- Be able to develop suitable methods for causal representation learning or causal discovery to address problems in specific domains
- Properly leverage causality in understanding and solving advanced machine learning and artificial intelligence problems
- Identify and formulate causal problems in your respective fields, and be able to find potential solutions

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 $\boldsymbol{V}$

Course Objectives

As an outcome of this course, participants are expected to

- Understand how causality is different from association and why it is useful
- Get familiar with graphical models, causality-related concepts and principles, and emerging approaches to causal discovery or causal representation learning from observational data
- Be acquainted with the state-of-the-art of causality research in different disciplines
- Be able to develop suitable methods for causal representation learning or causal discovery to address problems in specific domains
- Properly leverage causality in understanding and solving advanced machine learning and artificial intelligence problems
- Identify and formulate causal problems in your respective fields, and be able to find potential solutions

Outline of Class 1 & 2

- What is causality?
- Why causality (and ML)? A broad picture
    - Everyday life examples, generative AI, adaptive/robust prediction, recommender systems, culture…
- Typical causal problems
    - Identification of causal effects, correcting selection bias, counterfactual reasoning, causal discovery (causal representation learning)

Causality vs. Association

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 | Spai

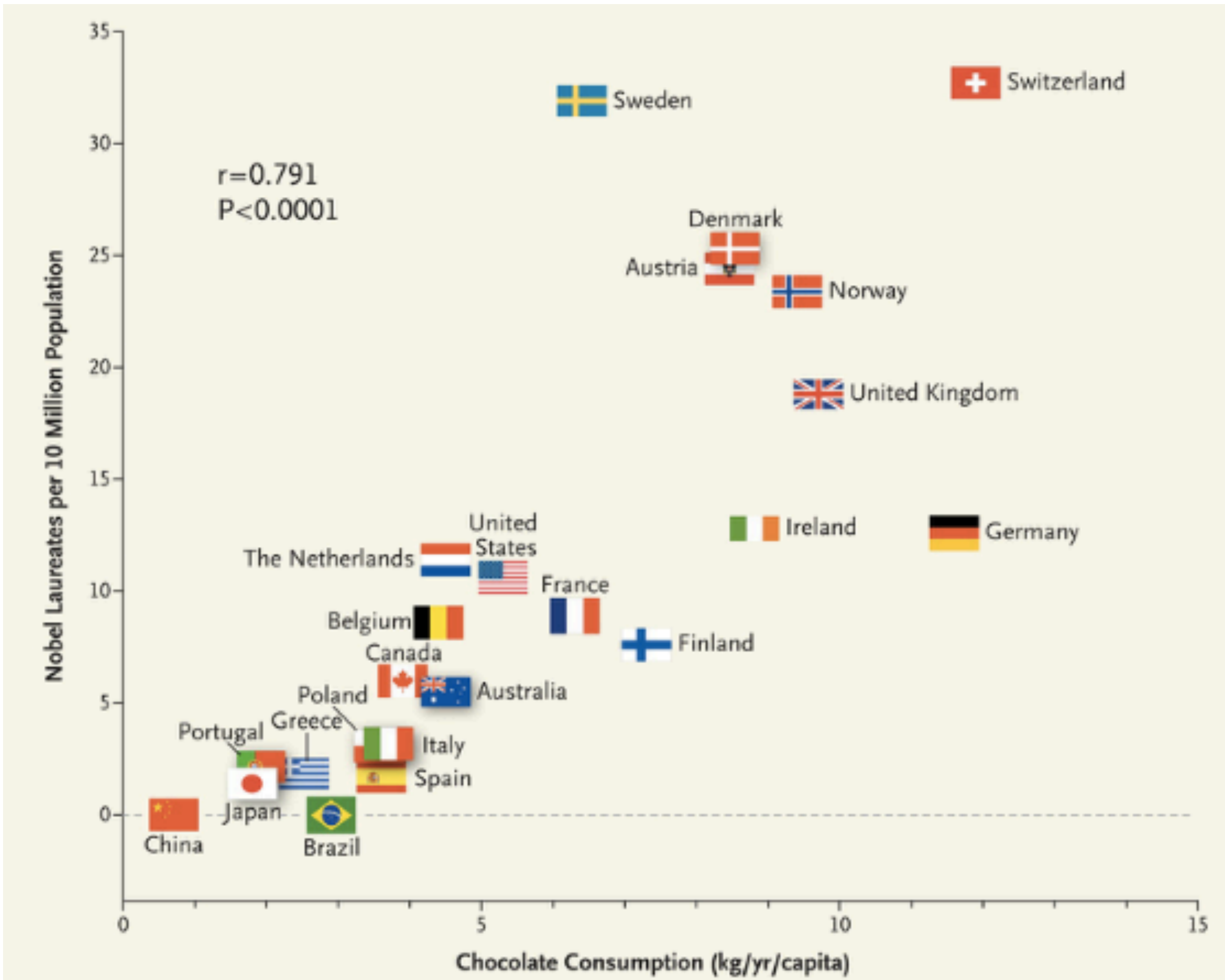
HOME » NEWS » WORLD NEWS » EUROPE

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.

Causality vs. Association

Another Example



Find Causal Relations from Observational Data: An Example

Thanks to collaborator Marlijn Noback

- 8 variables of 250 skeletons collected from different locations

	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=	Average at	Attrition p	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	AUSM8173	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

Let’s Look at AI Image Generator

- Prompt: a peacock eating ice cream

By Stable Diffusion: One Year Ago

- Prompt: a peacock eating ice cream

By DALL·E 3: Three Months Ago

- Prompt: a peacock eating ice cream

Understanding Dependence in Biology, Music, Text…

- Genes are often dependent—are they causally related?
- Causal process behind music, text…

Autoregresstive Generation in Music, Text, etc.?

Detecting and Identifying Selection Structure in Sequential Data

Yujia Zheng[1] **Zeyu Tang**[1] **Yiwen Qiu**[1] **Bernhard Schölkopf**[2] **Kun Zhang**[1 3]

Abstract

We argue that the selective inclusion of data points based on latent objectives is common in practical situations, such as music sequences. Since this selection process often distorts statistical analysis, previous work primarily views it as a bias to be corrected and proposes various methods to mitigate its effect. However, while controlling this bias is crucial, selection also offers an opportunity to provide a deeper insight into the hidden generation process, as it is a fundamental mechanism underlying what we observe. In particular, overlooking selection in sequential data can lead to an incomplete or overcomplicated inductive bias in modeling, such as assuming a universal autoregressive structure for all dependencies. Therefore, rather than merely viewing it as a bias, we explore the causal structure of selection in sequential data to delve deeper into the complete causal process.

generating process in various applications. For instance, in composing music, composers are guided by specific artistic goals or themes, leading them to selectively choose certain patterns of musical combinations (as combinations of basic elements) from their mind, thereby introducing dependencies among the basic elements in the music sequences (Schoenberg et al., 1967). These intentional but unmeasured selections, together with the contextual information, shape the structure of the compositions. A comprehensive understanding of the selection structure is essential for uncovering the underlying causal process and making use of it.

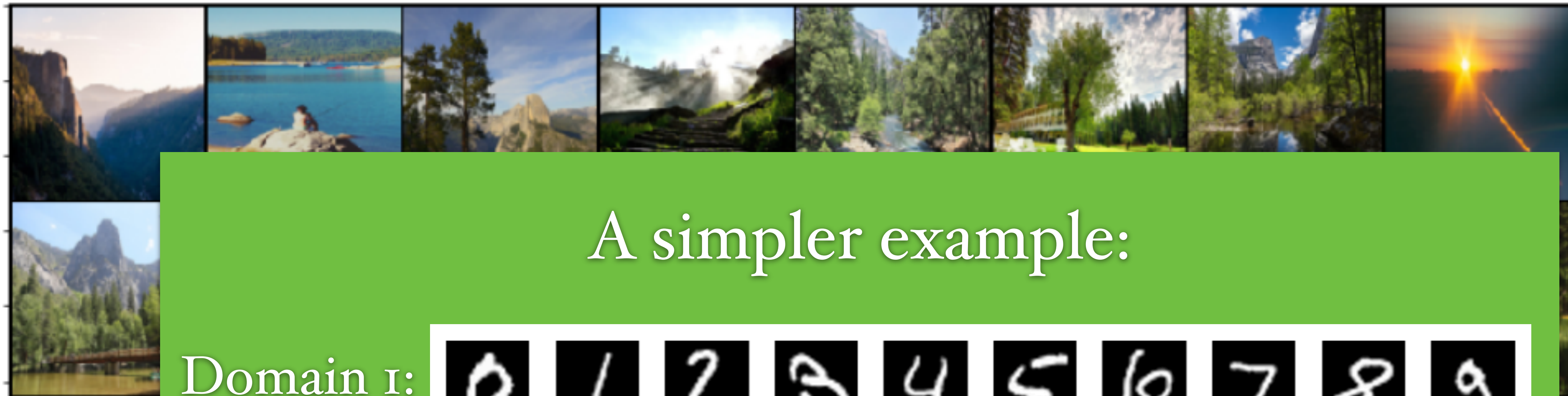
In sequential data, the understanding of selection plays a vital role. One essential question is whether selection leaves unique data dependence patterns that cannot be well explained by direct causal relations or latent confounding. Interestingly, as we will see in this paper, the answer is yes. Consequently, overlooking selection in such data can result in the introduction of incomplete or overcomplicated dependence models for the data. For instance, due to the

Translation by GPT-4o: An Example

Apologies!

- English: Athens was an aggressive city-state that conquered and subdued as much of the Greek peninsula and islands as it could, and the effort brought slaves to the city.
- Translated to: 雅典是一个好战的城邦，尽可能地征服和征服希腊半岛和岛屿，这一努力给城市带来了奴隶。

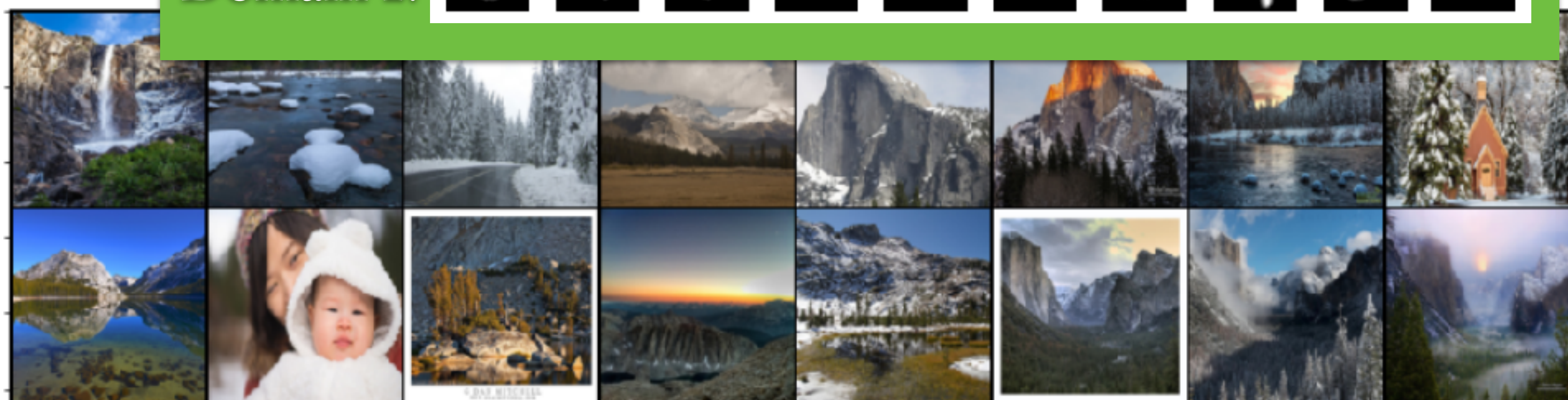
Unsupervised Image-to-Image Translation



A simpler example:

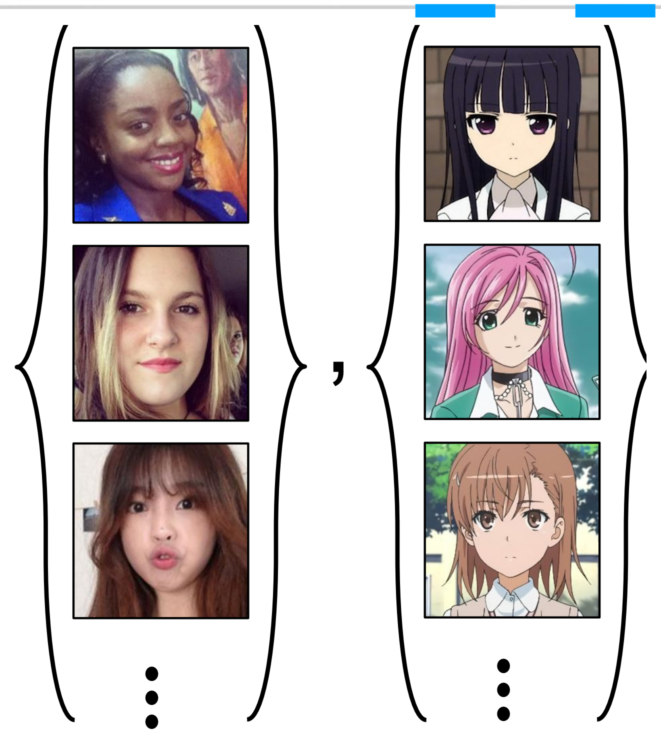
Domain 1:

Domain 2:

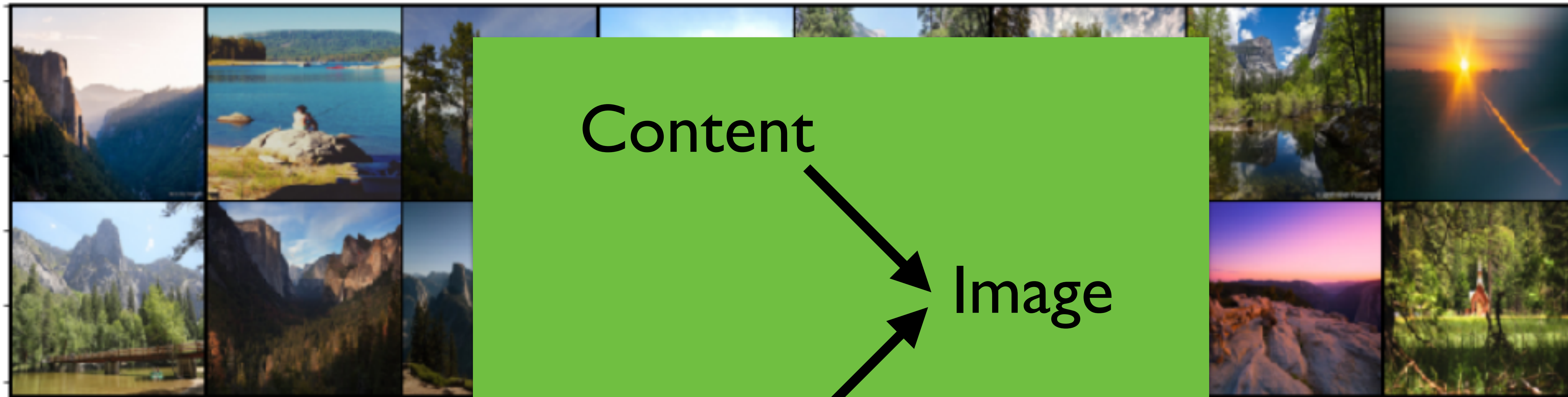


Images from the winter season domain.


Zebras Horses
Summer Winter
zebra → horse
summer → winter
horse → zebra
winter → summer

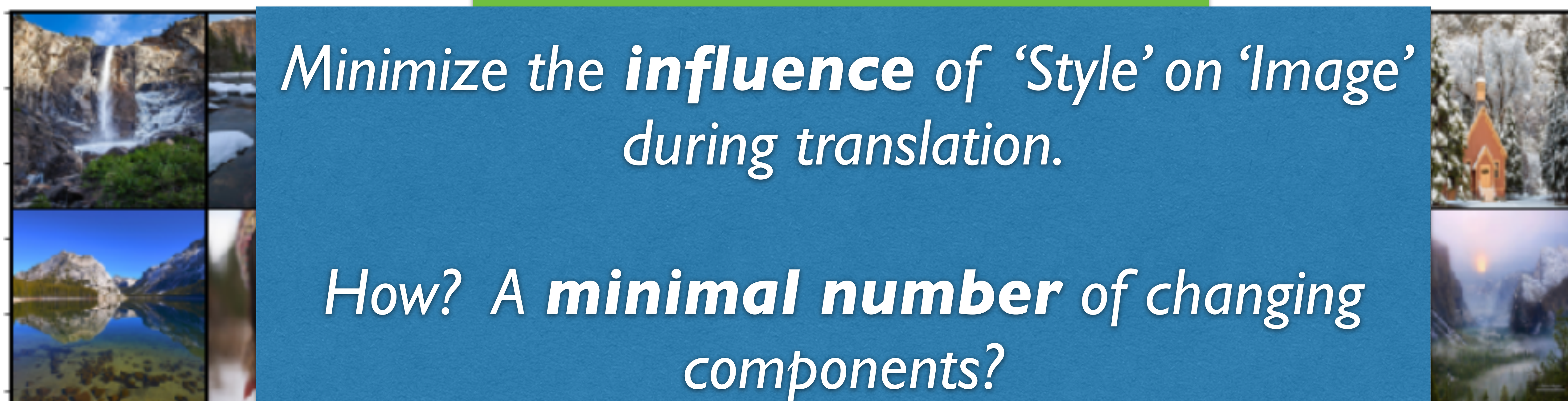


Unsupervised Image-to-Image Translation



Content → Image ← Style

Im ain.



*Minimize the **influence** of 'Style' on 'Image' during translation.*

*How? A **minimal number** of changing components?*

Images from the winter season domain.

Multi-domain Image Generation & Translation with Identifiability Guarantees

- Example: Generating female & male images with the same “content”

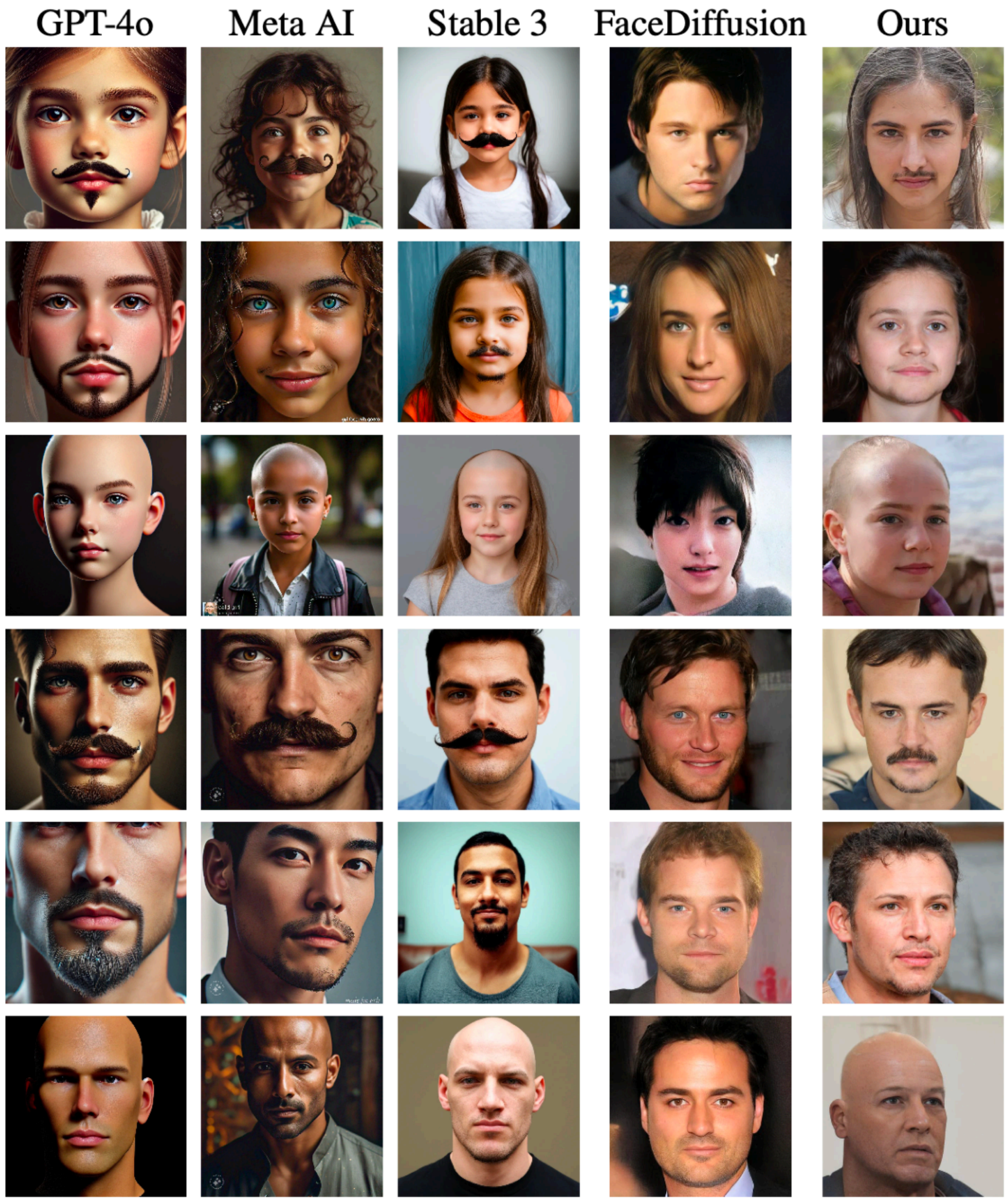
Ours

StyleGAN2-ADA

TGAN

- *Xie, Kong, Gong, Zhang, “Multi-domain image generation and translation with identifiability guarantees”, ICLR 2023*
- *Yan, Kong, Gui, Chi, Xing, He, Zhang, Counterfactual Generation with Identifiability Guarantee, NeurIPS 2023*

Minimal Changes + *Causal Modeling* for Generation



Corresponding to prompt:

girl with mustache,

girl with goatee,

bald girl,

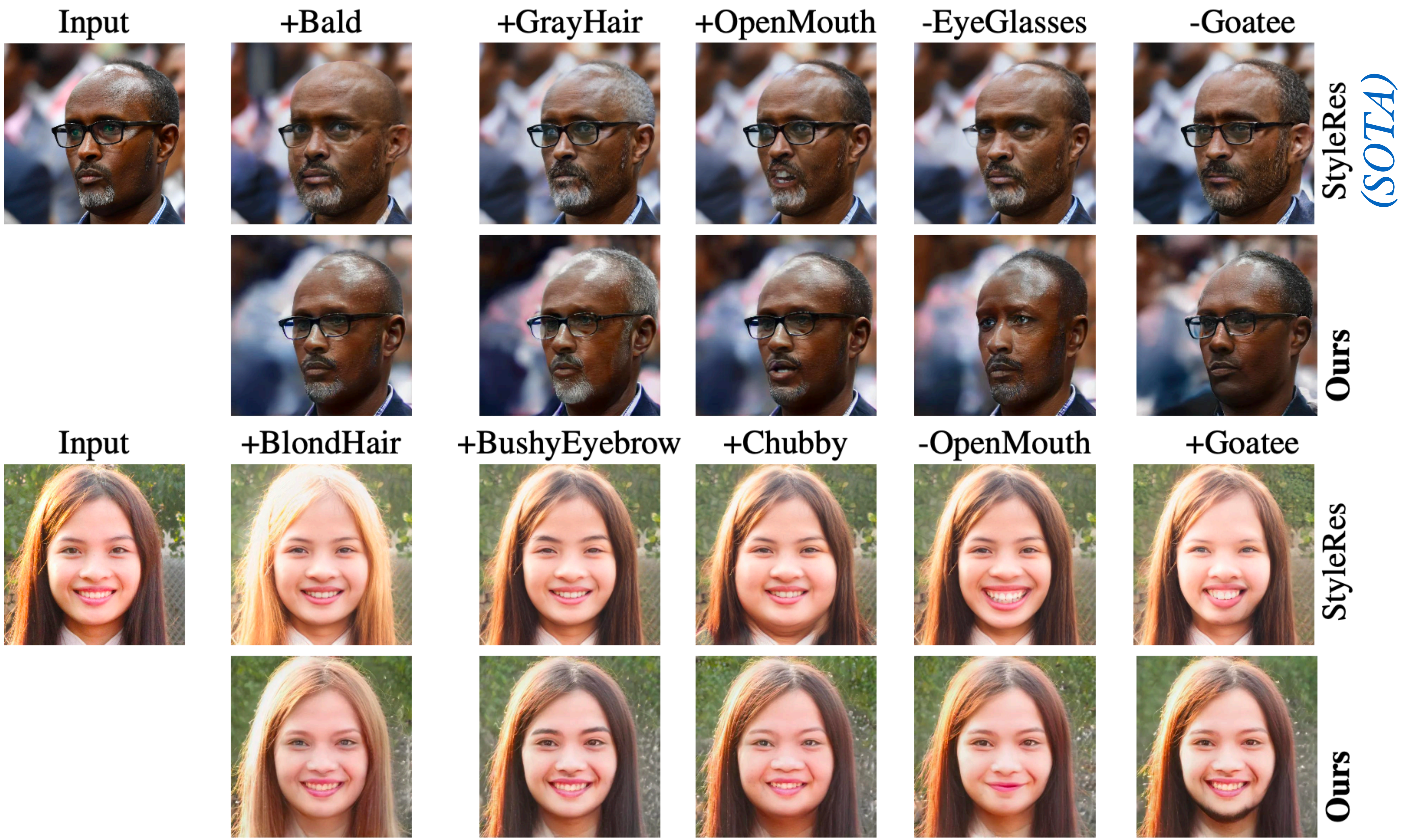
male with mustache,

male with goatee, and

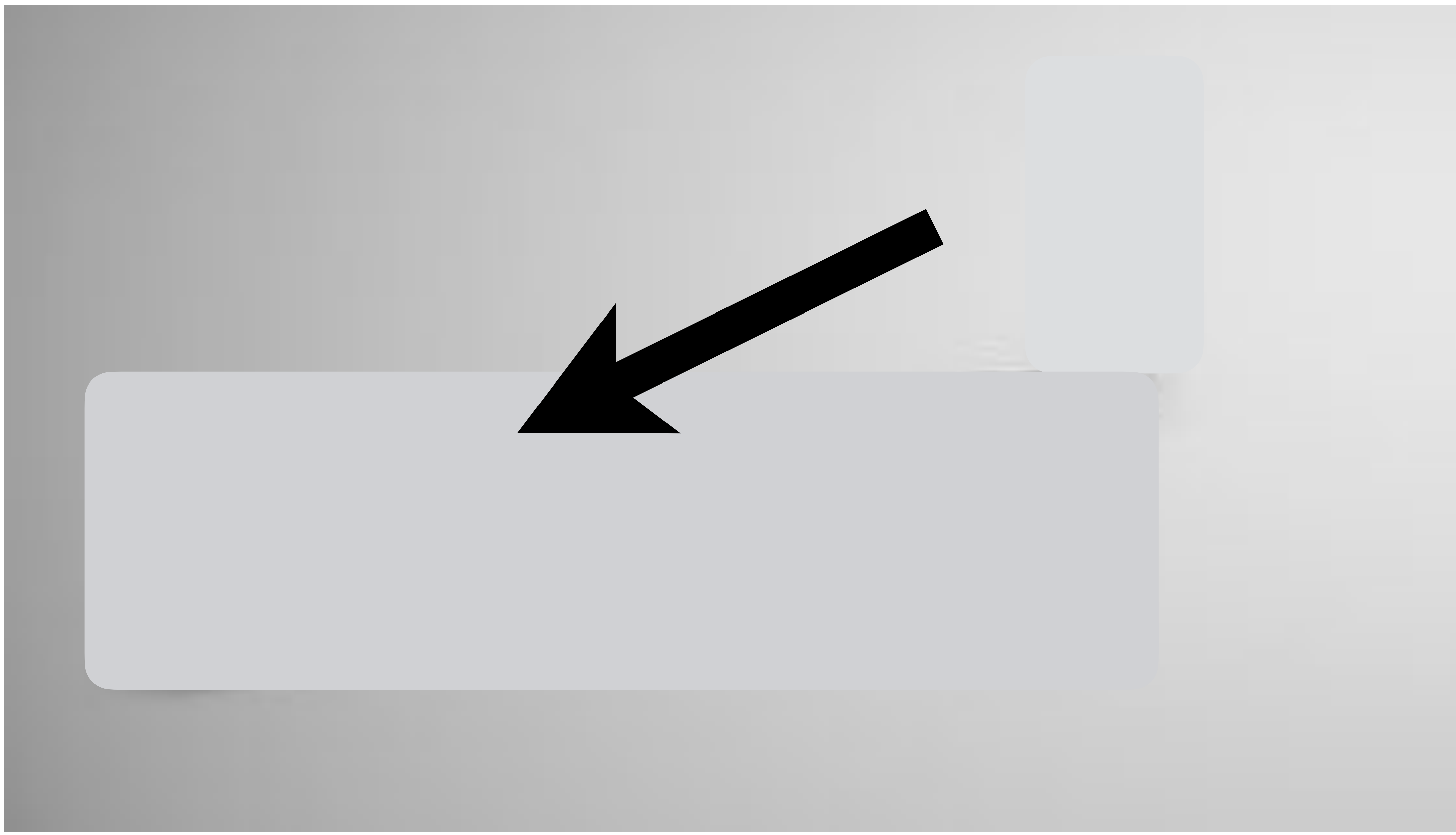
bald male

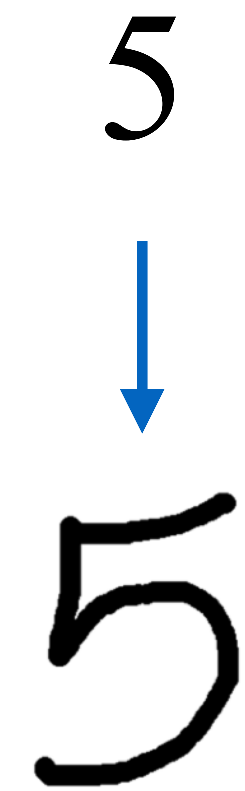
- S. Xie, Y. Zheng, I. Ng, K. Zhang, Causal Compositional Image Generation with Minimal Change, under submission

For Real Image Editing



Making Prediction in Nonstationary Environments





Understanding connections between different scenarios & *modeling* differences

Foreword to the I Ching

by Carl Gustav Jung

HTML Edition by Dan Baruth

developed what we call science. Our science, however, is based upon the principle of causality, and causality is considered to be an axiomatic truth. But a great change in ou[illegible]*e of Pure Reason* failed to do, is being accomplished by modern physics. The axioms of causality are being shaken to their foundations: we know now that what we term natural laws are merely statistical truths and thus must necessarily allow for exceptions. We have not sufficiently taken into account as yet that we need the laboratory with its incisive restrictions in order to demonstrate the invariable validity of natural law. If we leave things to nature, we see a very different picture: every process is partially or totally interfered with by chance, so much so that under natural circumstances a course of events absolutely conforming to specific laws is almost an exception.

The Chinese mind, as I see it at work in the *I Ching*, seems to be exclusively preoccupied with the chance aspect of events. What we call coincidence seems to be the chief concern of this peculiar mind, and what we worship as causality passes almost unnoticed. We must admit that there is something to be said for the immense importance of chance. An incalculable amount of human effort is directed to combating and restricting the nuisance or danger represented by chance. Theoretical considerations of cause and effect often look pale and dusty in comparison to the practical results of chance. It is all very well to say that the crystal of quartz is a hexagonal prism. The statement is quite true in so far as an ideal crystal is envisaged. But in nature one finds no two crystals exactly alike, although all are unmistakably hexagonal. The actual form, however, seems to appeal more to the Chinese sage than the ideal one. The jumble of natural laws constituting empirical reality holds more significance for him than a causal explanation of events that, moreover, must usually be separated from one another in order to be properly dealt with.

The manner in which the *I Ching* tends to look upon reality seems to disfavor our causalistic procedures. The moment under actual observation appears to the ancient Chinese view more of a chance hit than a clearly defined result of concurring causal chain processes. The matter of interest seems to be the configuration formed by chance events in the moment of observation, and not at all the hypothetical reasons that seemingly account for the coincidence. While the Western mind carefully sifts, weighs, selects, classifies, isolates, the Chinese picture of the moment encompasses everything down to the minutest nonsensical detail, because all of the ingredients make up the observed moment.

Thus it happens that when one throws the three coins, or counts through the forty-nine yarrow stalks, these chance details enter into the picture of the moment of observation and form a part of it -- a part that is insiguificant to us, yet most meaningful to the Chinese mind. With us it would be a banal and almost meaningless statement (at least on the face of it) to say that whatever happens in a given moment possesses inevitably the quality peculiar to that moment. This is not an abstract argument but a very practical one. There are certain connoisseurs who can tell you merely from the appearance, taste, and behavior of a wine the site of its vineyard and the year of its origin. There are antiquarians who with almost uncanny accuracy will name the time and place of origin and the maker of an *objet d'art* or piece of furniture on merely looking at it. And there are even astrologers who can tell you, without any previous knowledge of your nativity, what the position of sun and moon was and what zodiacal sign rose above the horizon in the moment of your birth. In the face of such facts, it must be admitted that moments can leave long-lasting traces.

In other words, whoever invented the *I Ching* was convinced that the hexagram worked out in a certain moment coincided with the latter in quality no less than in time. To him the hexagram was the exponent of the moment in which it was cast -- even more so than the hours of the clock or the divisions of the calendar could be -- inasmuch as the hexagram was understood to be an indicator of the essential situation prevailing in the moment of its origin.

This assumption involves a certain curious principle that I have termed synchronicity,[2] a concept that formulates a point of view diametrically opposed to that of causality. Since the latter is a merely statistical truth and not absolute, it is a sort of working hypothesis of how events evolve one out of another, whereas synchronicity takes the coincidence of events in space and time as meaning something more than mere chance, namely, a peculiar interdependence of objective events among themselves as well as with the subjective (psychic) states of the observer or observers.

The ancient Chinese mind contemplates the cosmos in a way comparable to that of the modern physicist, who cannot deny that his model of the world is a decidedly psychophysical structure. The microphysical event includes the observer just as much as the reality underlying the *I Ching* comprises subjective, i.e., psychic conditions in the totality of the momentary situation. Just as causality describes the sequence of events, so synchronicity to the Chinese mind deals with the coincidence of events. The causal point of view tells us a dramatic story about how *D* came into existence: it took its origin from *C*, which existed before *D*, and *C* in its turn had a father, *B*, etc. The synchronistic view on the other hand tries to produce an equally meaningful picture of coincidence. How does it happen that *A'*, *B'*, *C'*, *D'*, etc., appear all in the same moment and in the same place? It happens in the first place because the physical events *A'* and *B'* are of the same quality as the psychic events *C'* and *D'*, and further because all are the exponents of one and the same momentary situation. The situation is assumed to represent a legible or understandable picture.

Now the sixty-four hexagrams of the *I Ching* are the instrument by which the meaning of sixty-four different yet typical situations can be determined. These interpretations are equivalent to causal explanations. Causal connection is statistically necessary and can therefore be subjected to experiment. Inasmuch as situations are unique and cannot be repeated, experimenting with synchronicity seems to be impossible under ordinary

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 + ...

Remember the Scientific Revolution?

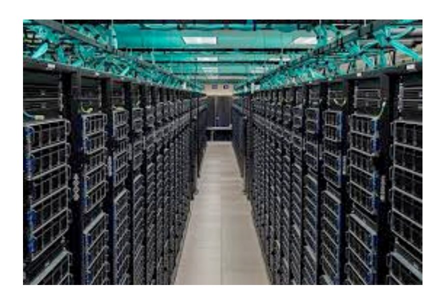
 

- By Copernicus, Galilei, Newton, Bacon, Harvey…
- Book production, observational data, the ability to do some experiments, basic inference rules…
- Quantitative vs. qualitative view of nature; new experimental, scientific method seeking definite answers; “how” instead of “why”…
- This revolution in human thought changed the world

Causal ML Facilitates the Second Scientific Revolution (I Believe)


- Analogy to Scientific Revolution
  - By Copernicus, Galilei, Newton, Bacon, Harvey…
  - Book production, observational data, the ability to do some experiments, basic inference rules…
  - Quantitative vs. qualitative view of nature; new experimental, scientific method seeking definite answers; "how" instead of "why"…

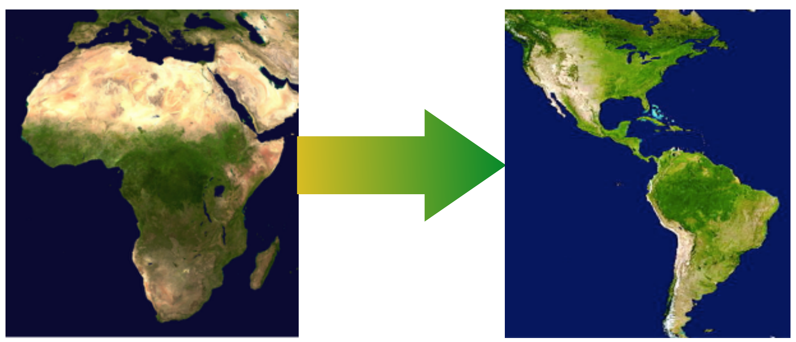
- Available: Internet, data, statistical tools, computational resources…
- Goals? Learning paradigms? Methodology?
- (Causal ML) will impact each scientific discipline, every industry, and human society

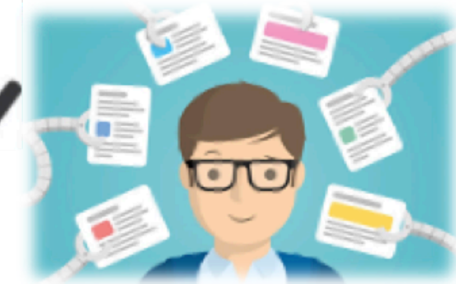


Answering Why Questions: A View

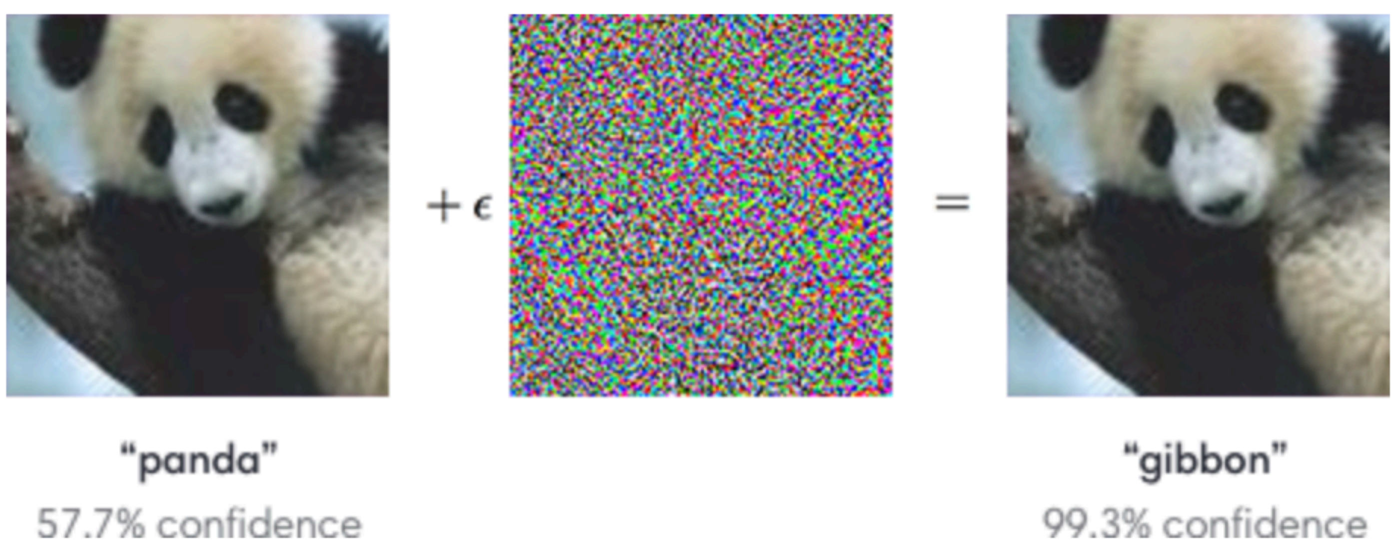
Good Representations Are Needed…

- Generalization/adaptation, decision making, fairness, recommendations, generative AI…



- Dealing with adversarial attacks?



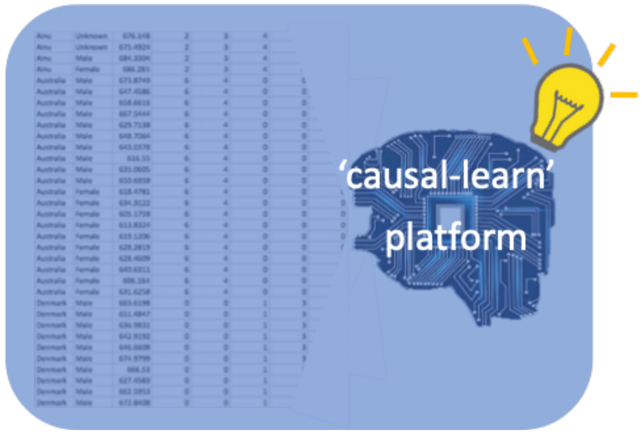
(Goodfellow et al., 2014)

An adversarial input, overlaid on a typical image, can cause a classifier to miscategorize a panda as a gibbon.

Outline of Class 1 & 2

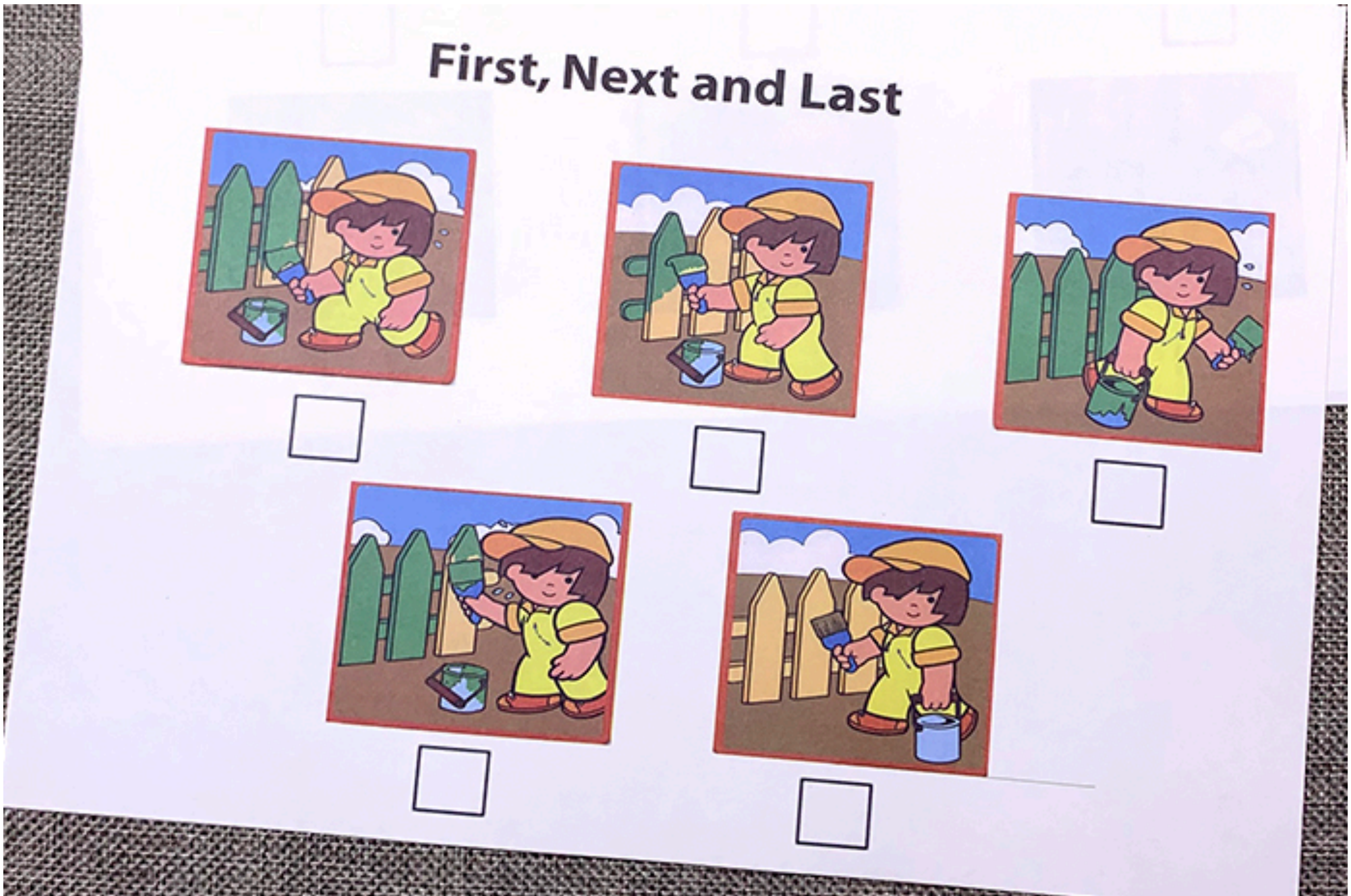
- What is causality?
- Why causality (and ML)? A broad picture
  - Everyday life examples, generative AI, adaptive/robust prediction, recommender systems, culture...
- Typical causal problems
  - Identification of causal effects, correcting selection bias, counterfactual reasoning, causal discovery (causal representation learning)

Uncover Causality from Observational Data: Task?

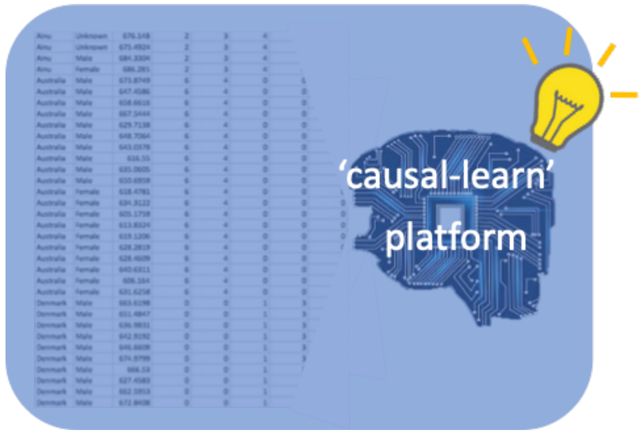


- Causal discovery (Spirtes et al., 1993)/ causal representation learning (Schölkopf et al., 2021): find such representations with identifiability guarantees

Temporal Order? Assumptions are Needed…

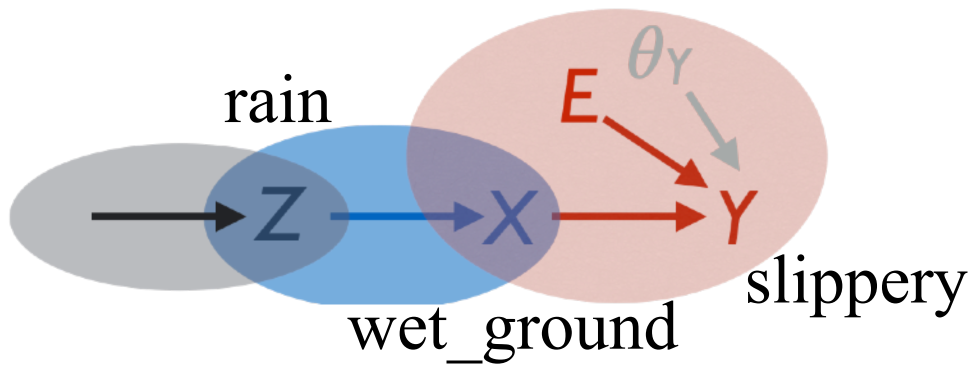


Uncover Causality from Observational Data: How?



- Causal discovery (Spirtes et al., 1993)/ causal representation learning (Schölkopf et al., 2021): find such representations with identifiability guarantees

- Causal system has "irrelevant" modules (Spirtes et al., 1993; Pearl, 2000)



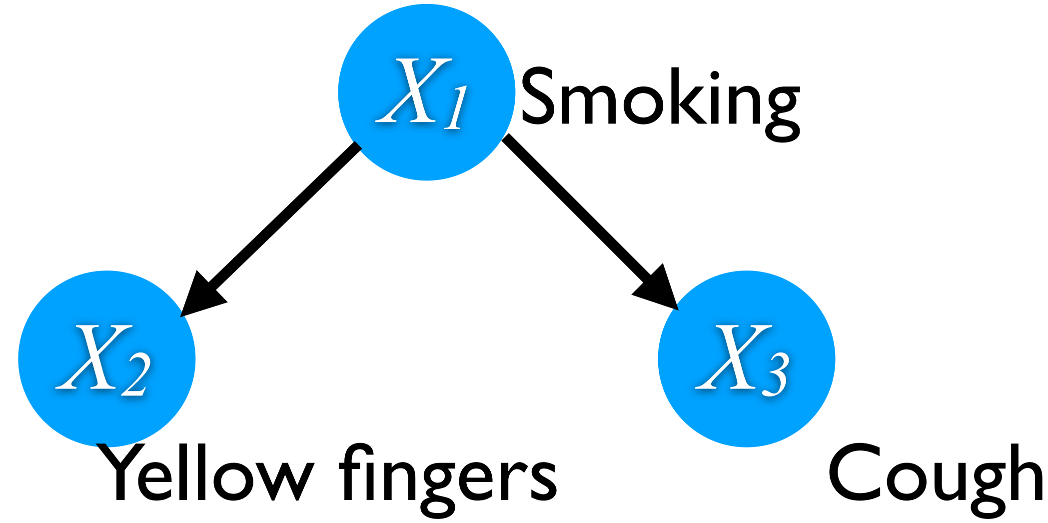
- conditional independence among variables;
- independent noise condition;
- minimal (and independent) changes…

Footprint of causality in data

- Three dimensions of the problem:

i.i.d. data?	**Parametric constraints?**	**Latent confounders?**
Yes	No	No
No	Yes	Yes

Formulation: Three Types of Problems in Current AI



- Three questions:

X_1	X_2	X_3
1	0	0
0	0	1
0	1	1
1	1	1
0	0	0
0	1	0
1	1	1
1	1	1
0	0	0
1	0	0
…	…	…

- **Prediction**: Would the person cough if we *find* he/she has yellow fingers?

$$P(X3 \mid X2=1)$$

- **Intervention**: Would the person cough if we *make sure* that he/she has yellow fingers?

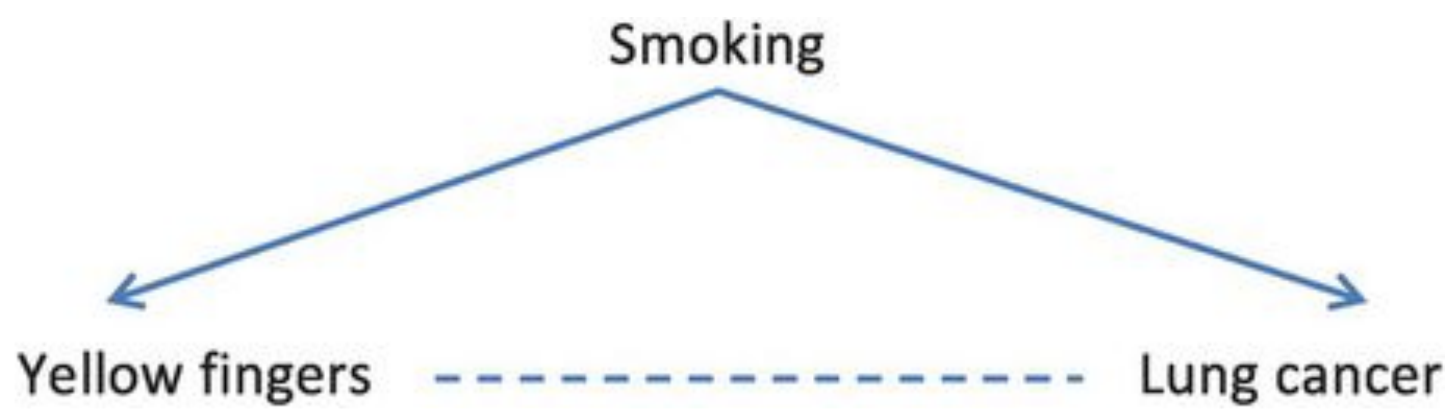
$$P(X3 \mid do\,(X2=1))$$

- **Counterfactual**: Would George cough *had* he had yellow fingers, *given that he does not have yellow fingers and coughs*?

$$P(X3_{X2=1} \mid X2 = 0, X3 = 1)$$

Causal Thinking: Making Changes?

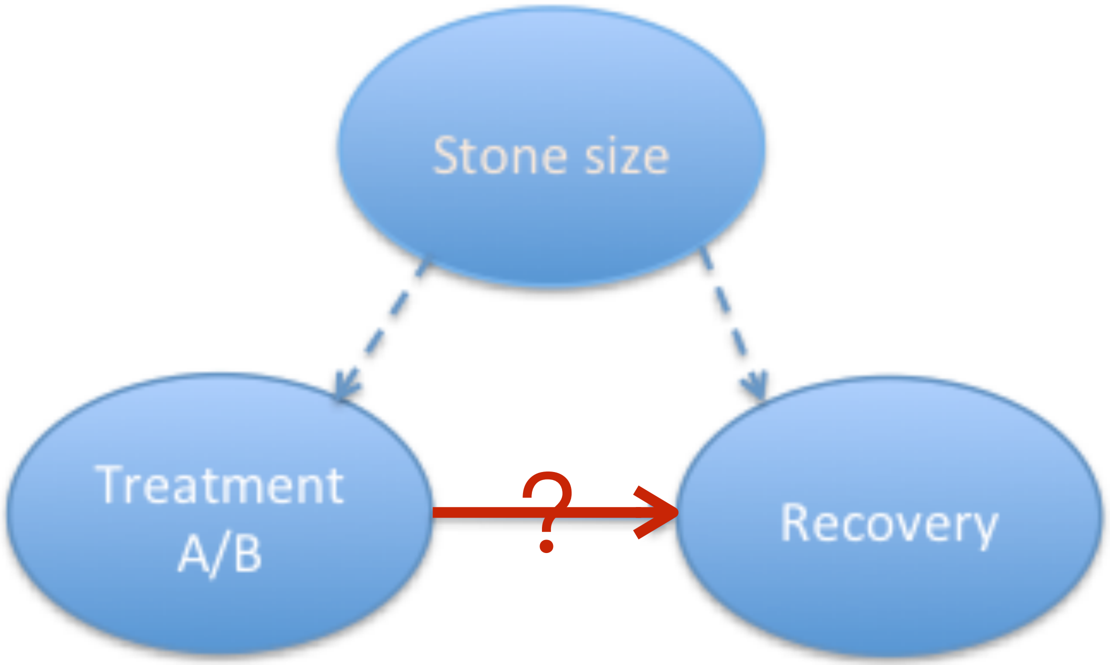
- Dependence vs. causality



Causal Thinking: Why "Paradox"?

- Dependence vs. causality
- Simpson's paradox

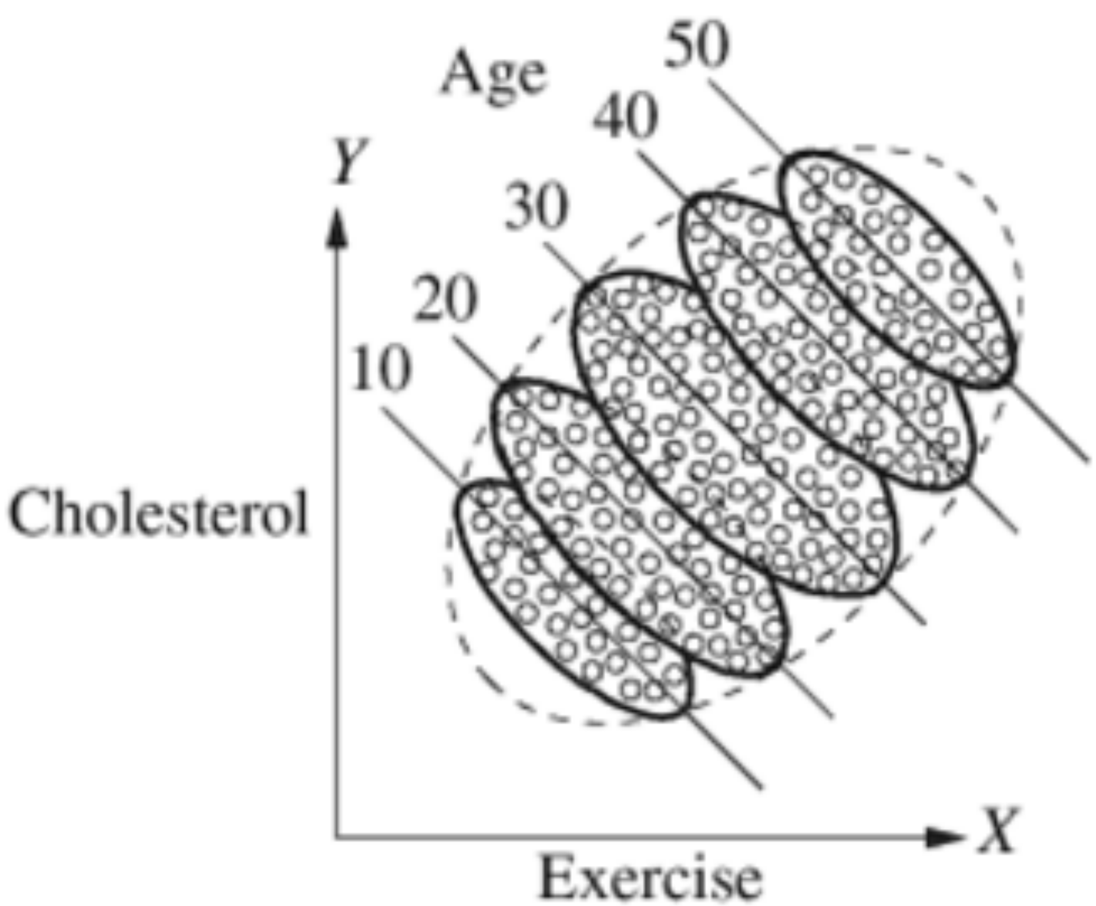
	Treatment A	Treatment B
Small Stones	*Group 1* **93% (81/87)**	*Group 2* 87% (234/270)
Large Stones	*Group 3* **73% (192/263)**	*Group 4* 69% (55/80)
Both	78% (273/350)	**83% (289/350)**

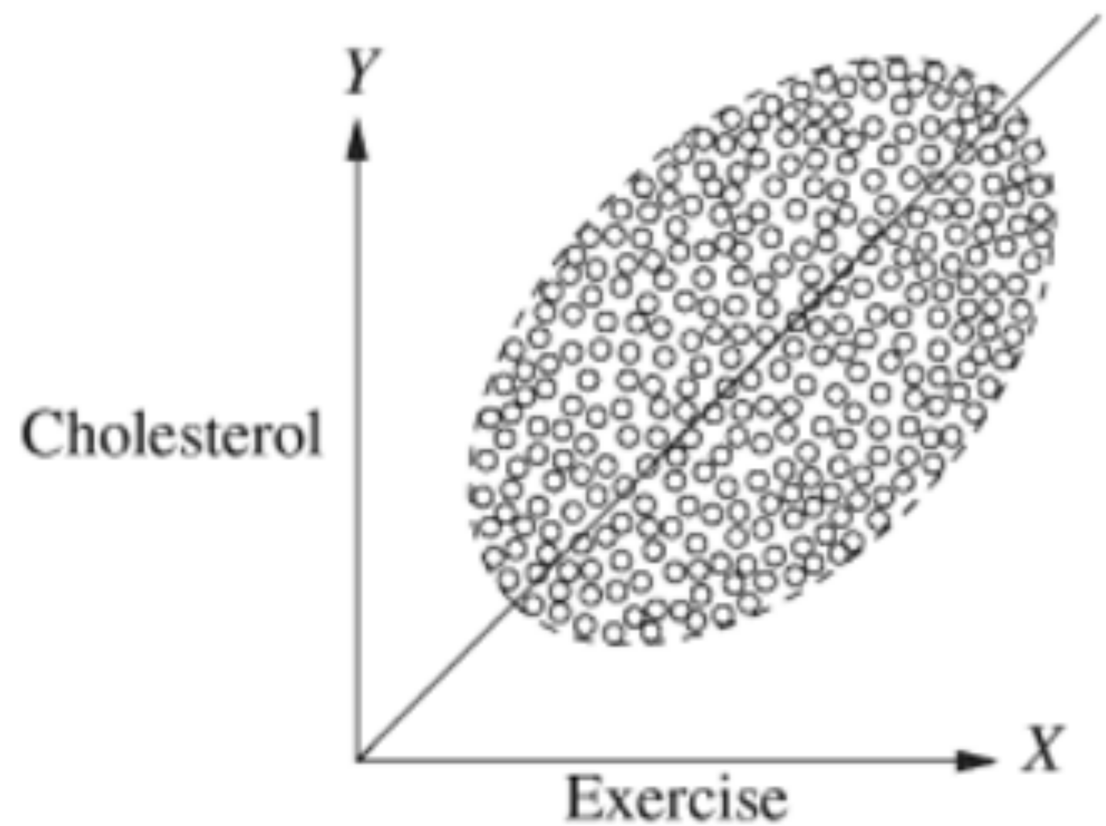


Causal Thinking: Why "Paradox"?

- Dependence vs. causality

- Simpson's paradox



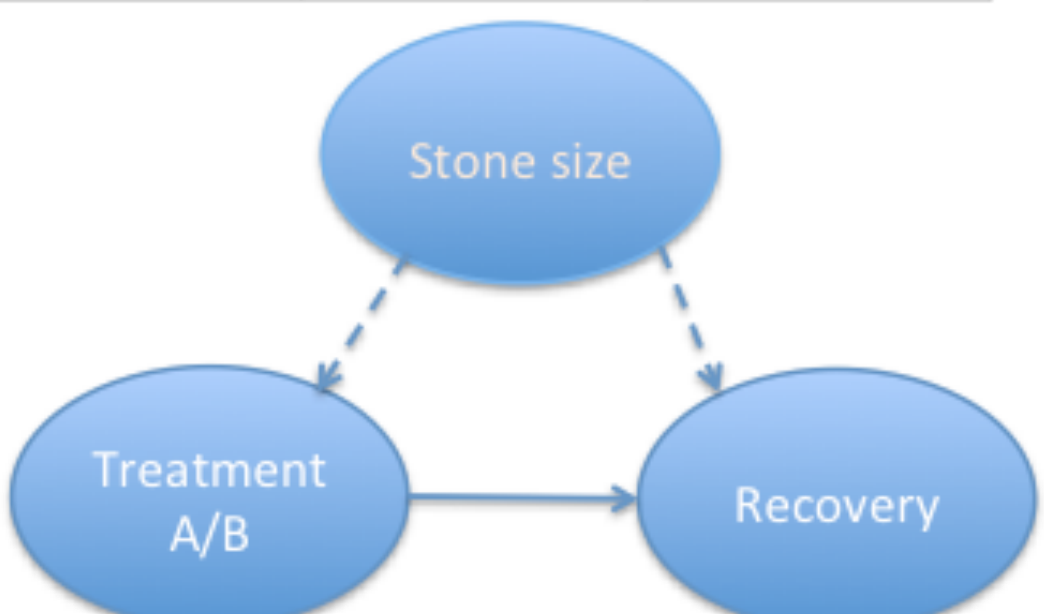


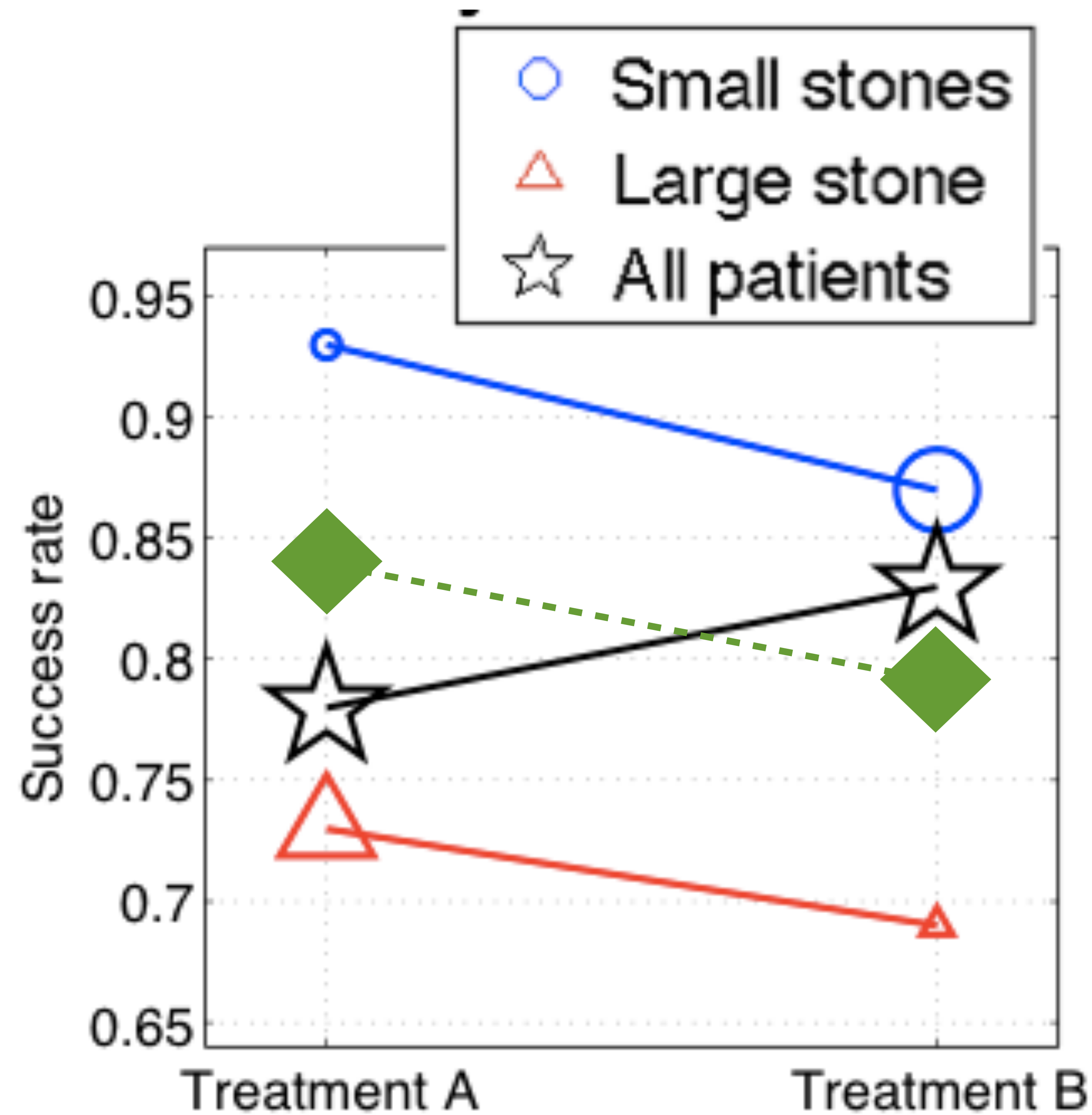
Causal Thinking: Why "Paradox"?

- Dependence vs. causality

- 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)**

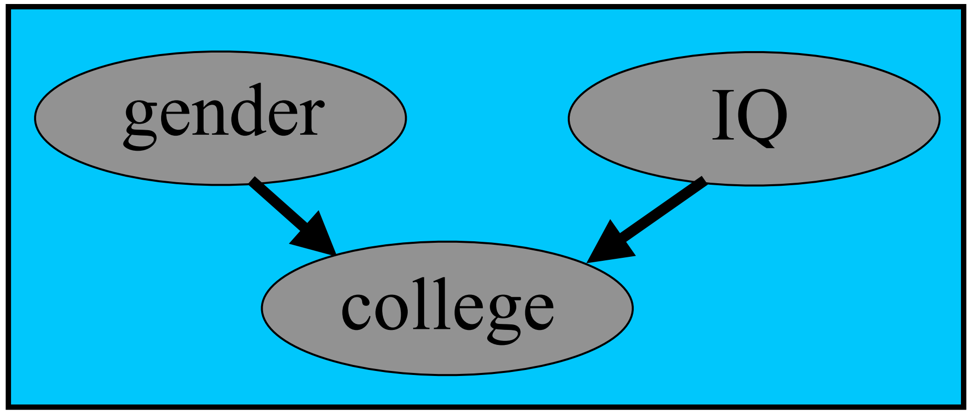




Prediction vs. causal effect

Causal Thinking: Sample vs. Population

- Dependence vs. causality
- Simpson's paradox
- "Strange" dependence
  - Go back 50 years; female college students were smarter than male ones on average. Why?

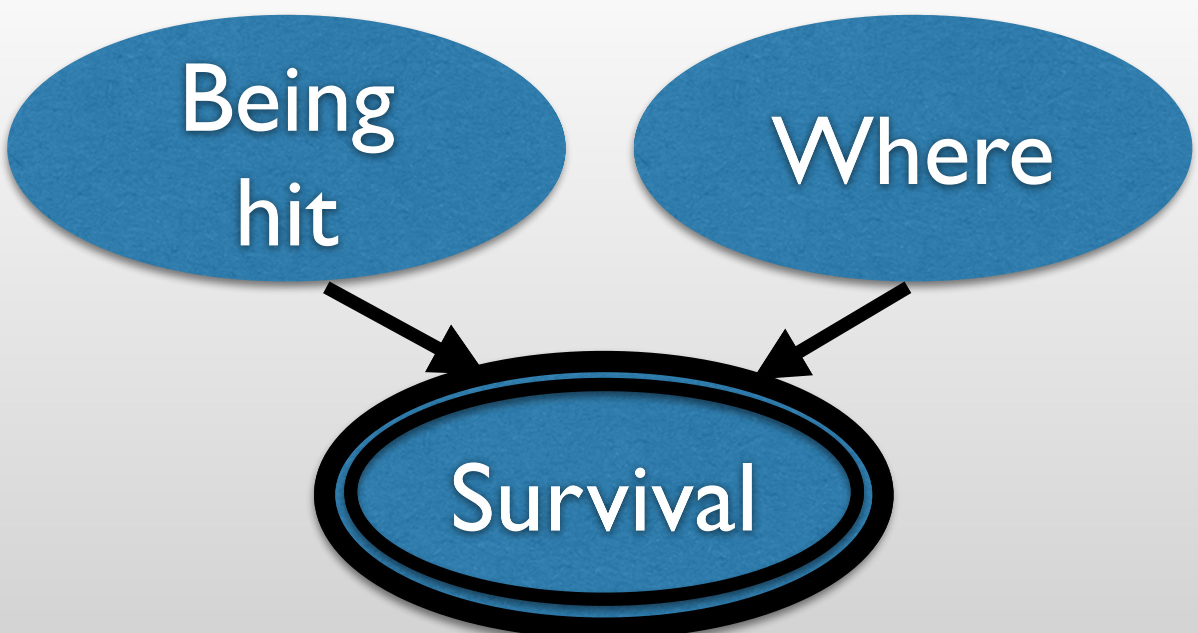


Causal Thinking: Sample vs. Population

- Dependence vs. causality
- Simpson's paradox
- "Strange" dependence

Causal Thinking: Sample vs. Population

- Dependence vs. causality

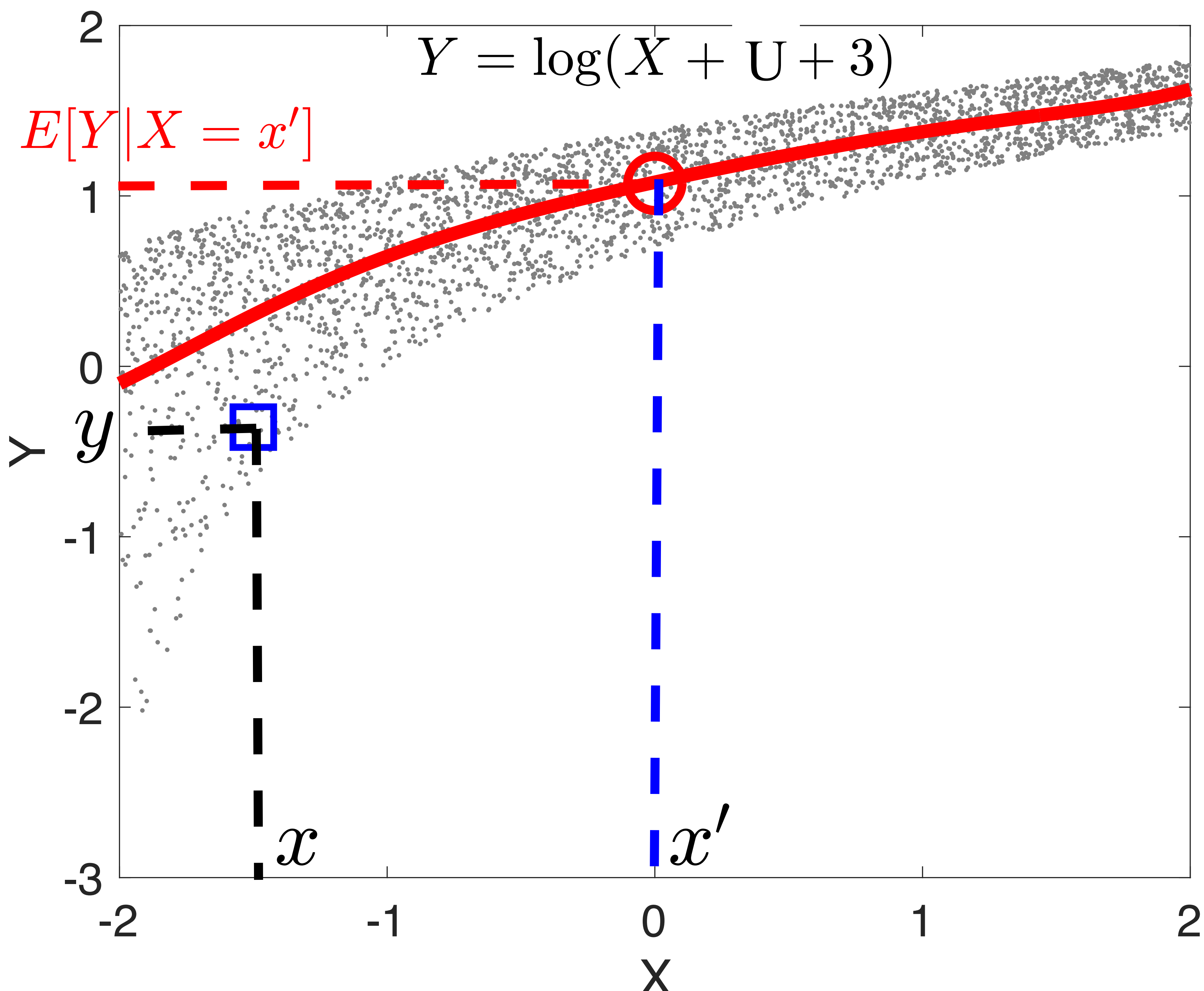


- Observe P(B,W | S=1)
- Infer P(B,W | S=0)

Counterfactual Inference vs. Prediction

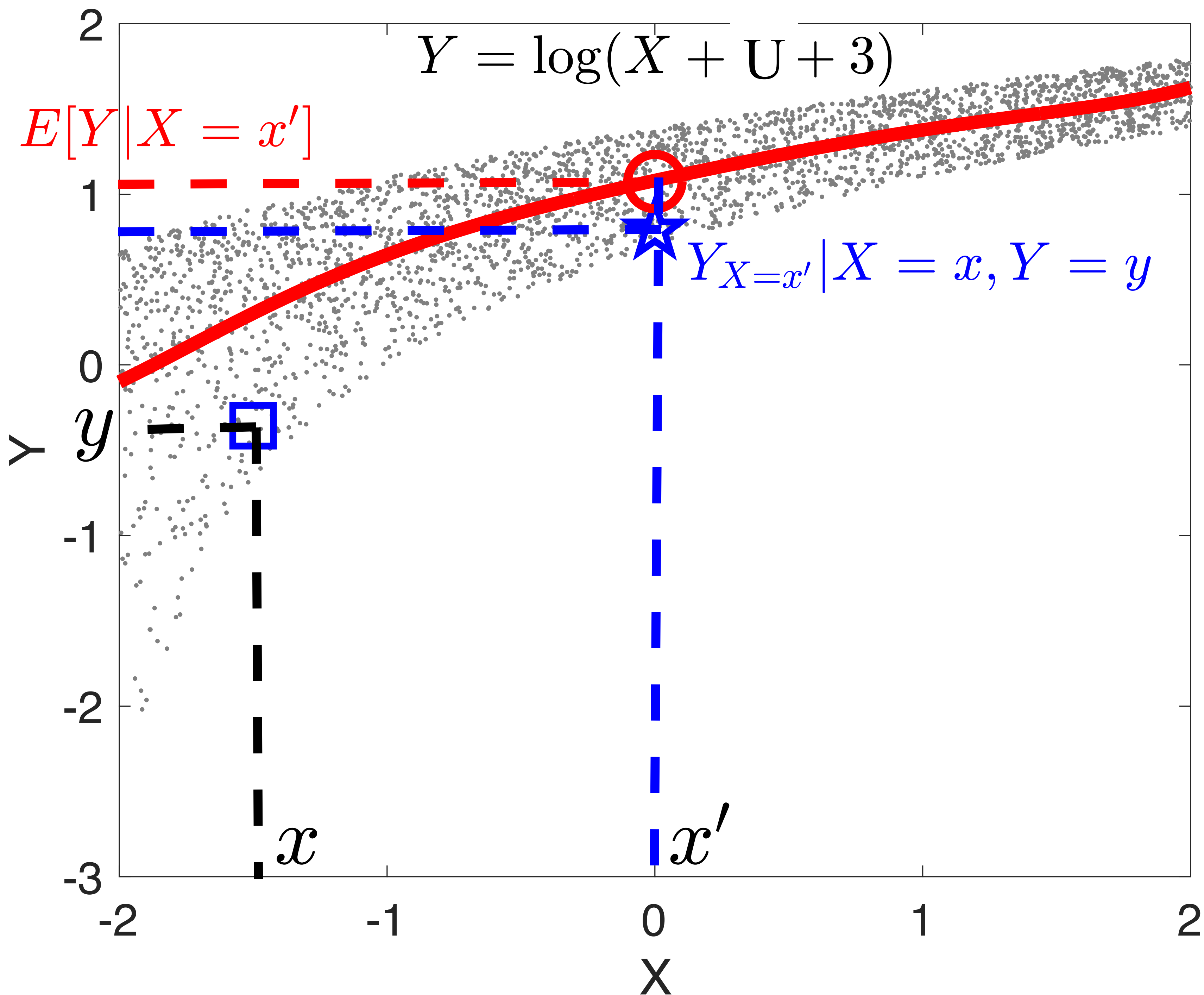
- Suppose X→Y with Y = log(X + U + 3). For an individual with (*x*,*y*), what would Y be if X had been *x'* ?

  attendance (X) *grade* (Y)



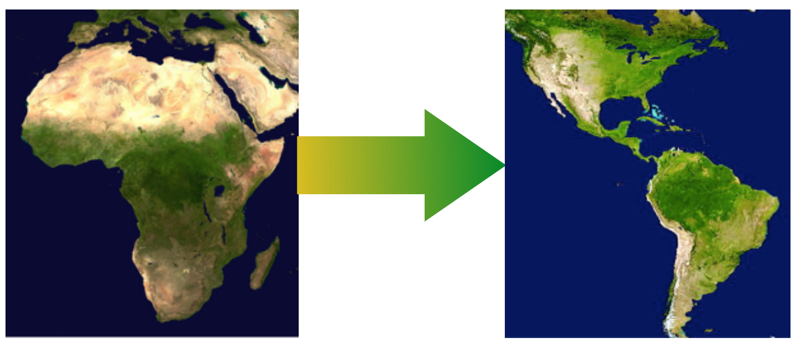
Counterfactual Inference vs. Prediction

- Suppose X→Y with Y = log(X + U + 3). For an individual with (*x*,*y*), what would Y be if X had been *x'* ?

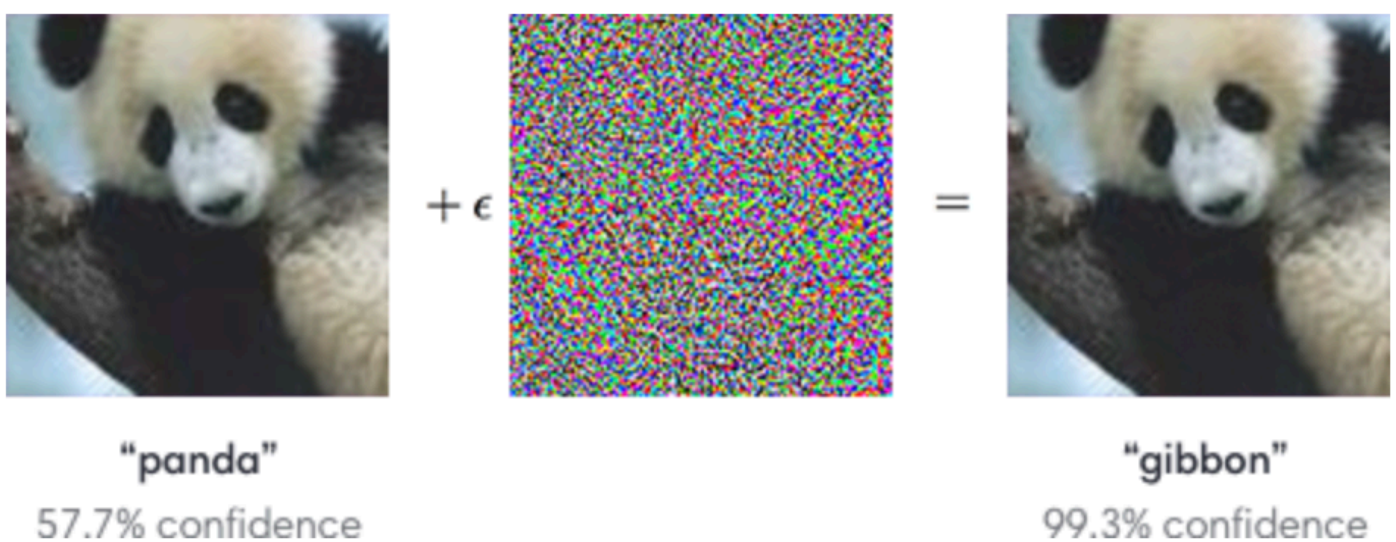


Good Representations Are Needed…

- Generalization/adaptation, decision making, fairness, recommendations, generative AI…



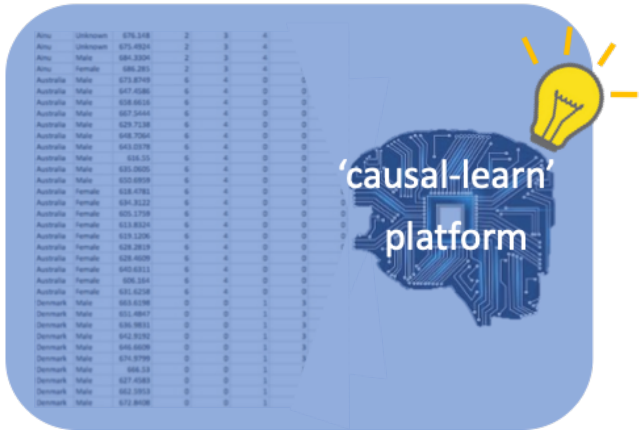

- Dealing with adversarial attacks?



(Goodfellow et al., 2014)

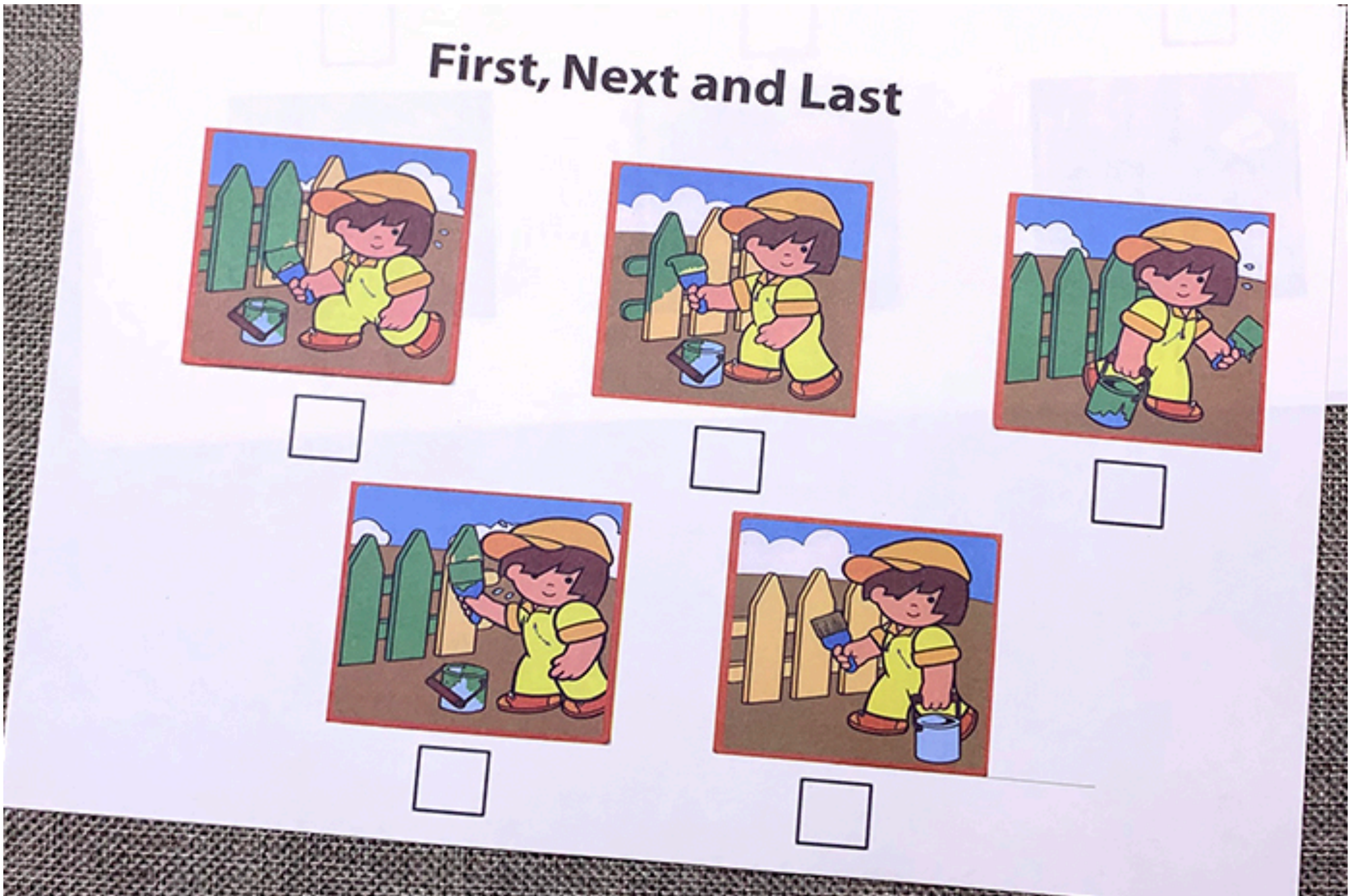
An adversarial input, overlaid on a typical image, can cause a classifier to miscategorize a panda as a gibbon.

Uncover Causality from Observational Data: Task?

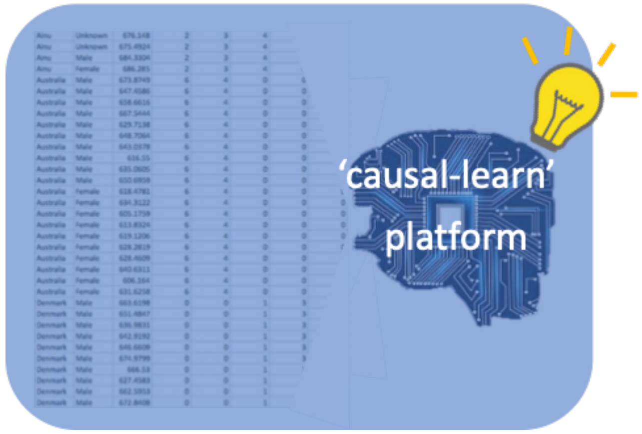


- Causal discovery (Spirtes et al., 1993)/ causal representation learning (Schölkopf et al., 2021): find such representations with identifiability guarantees

Temporal Order? Assumptions are Needed…

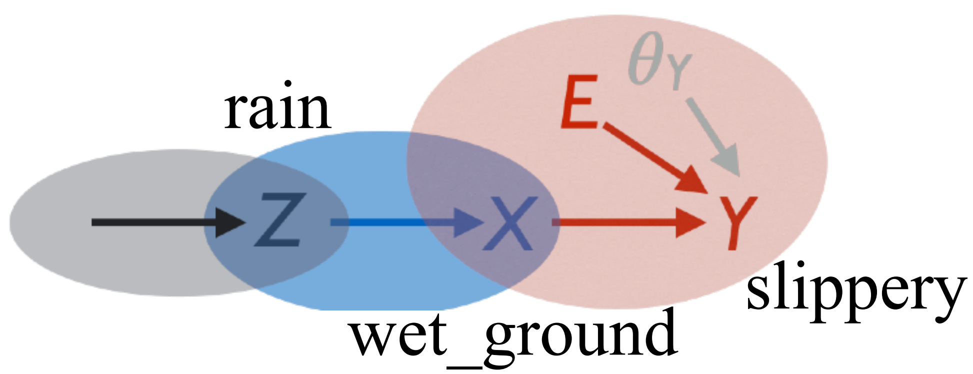


Uncover Causality from Observational Data: How?



- Causal discovery (Spirtes et al., 1993)/ causal representation learning (Schölkopf et al., 2021): find such representations with identifiability guarantees

- Causal system has "irrelevant" modules (Spirtes et al., 1993; Pearl, 2000)



- conditional independence among variables;
- independent noise condition;
- minimal (and independent) changes...

Footprint of causality in data

- Three dimensions of the problem:

i.i.d. data?	**Parametric constraints?**	**Latent confounders?**
Yes	No	No
No	Yes	Yes

Causal Representation Learning: Recent Advances

<table>
<tr><th>i.i.d. data?</th><th>Parametric constraints?</th><th>Latent confounders?</th><th>What can we get?</th></tr>
<tr><td rowspan="4">Yes</td><td rowspan="2">No</td><td>No</td><td rowspan="2">(Different types of) equivalence class</td></tr>
<tr><td>Yes</td></tr>
<tr><td rowspan="2">Yes</td><td>No</td><td rowspan="2">Unique identifiability (under structural conditions)</td></tr>
<tr><td>Yes</td></tr>
<tr><td rowspan="2">Non-I, but I.D.</td><td rowspan="2">No/Yes</td><td>No</td><td>(Extended) regression</td></tr>
<tr><td>Yes</td><td>Latent temporal causal processes identifiable!</td></tr>
<tr><td rowspan="4">I., but non-I.D.</td><td>No</td><td rowspan="2">No</td><td>More informative than MEC (CD-NOD)</td></tr>
<tr><td>Yes</td><td>May have unique identifiability</td></tr>
<tr><td>No</td><td rowspan="2">Yes</td><td>Changing subspace identifiable</td></tr>
<tr><td>Yes</td><td>Variables in changing relations identifiable</td></tr>
</table>

Causal Discovery in Archeology: An Example

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

Thanks to Marlijn Noback

- 8 variables of 250 skeletons collected from different locations

	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=	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	AUSM8173	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

(Typical) Constraint-Based Causal Discovery

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

- Conditional independence constraints between each variable pair
  - Illustration: the PC algorithm
  - Extensions: the FCI algorithm...

X1	X2	X3	X4	X5
-1.1	1	1.3	0.2	-0.7
2.1	2	3.1	-1.3	-1.6
3.1	4.2	-2.6	0.6	2.1
2.3	-0.6	-3.5	0.8	2.3
1.3	-1.7	0.9	2.4	-1.4
-1.8	0.9	-1.3	0.9	0.7
...	...	...	...	...

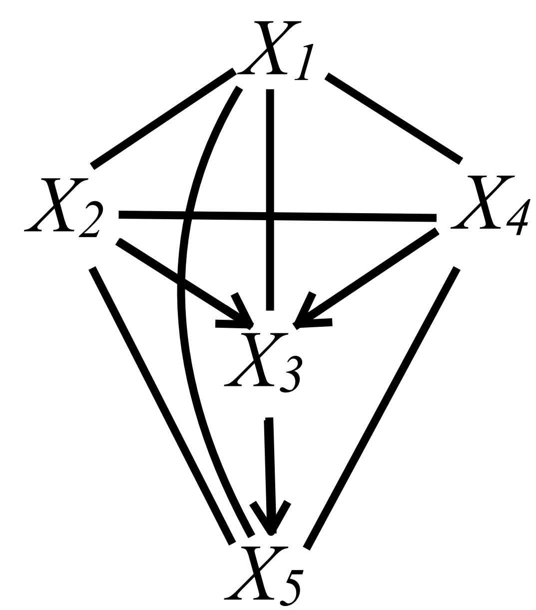
$X_1 \perp\!\!\!\perp X_5 \mid X_3$

$X_2 \perp\!\!\!\perp X_4 \mid X_1$

$X_2 \perp\!\!\!\perp X_5 \mid X_3$

$X_4 \perp\!\!\!\perp X_5 \mid X_3$

$X_1 \perp\!\!\!\perp X_3 \mid \{X_2, X_4\}$

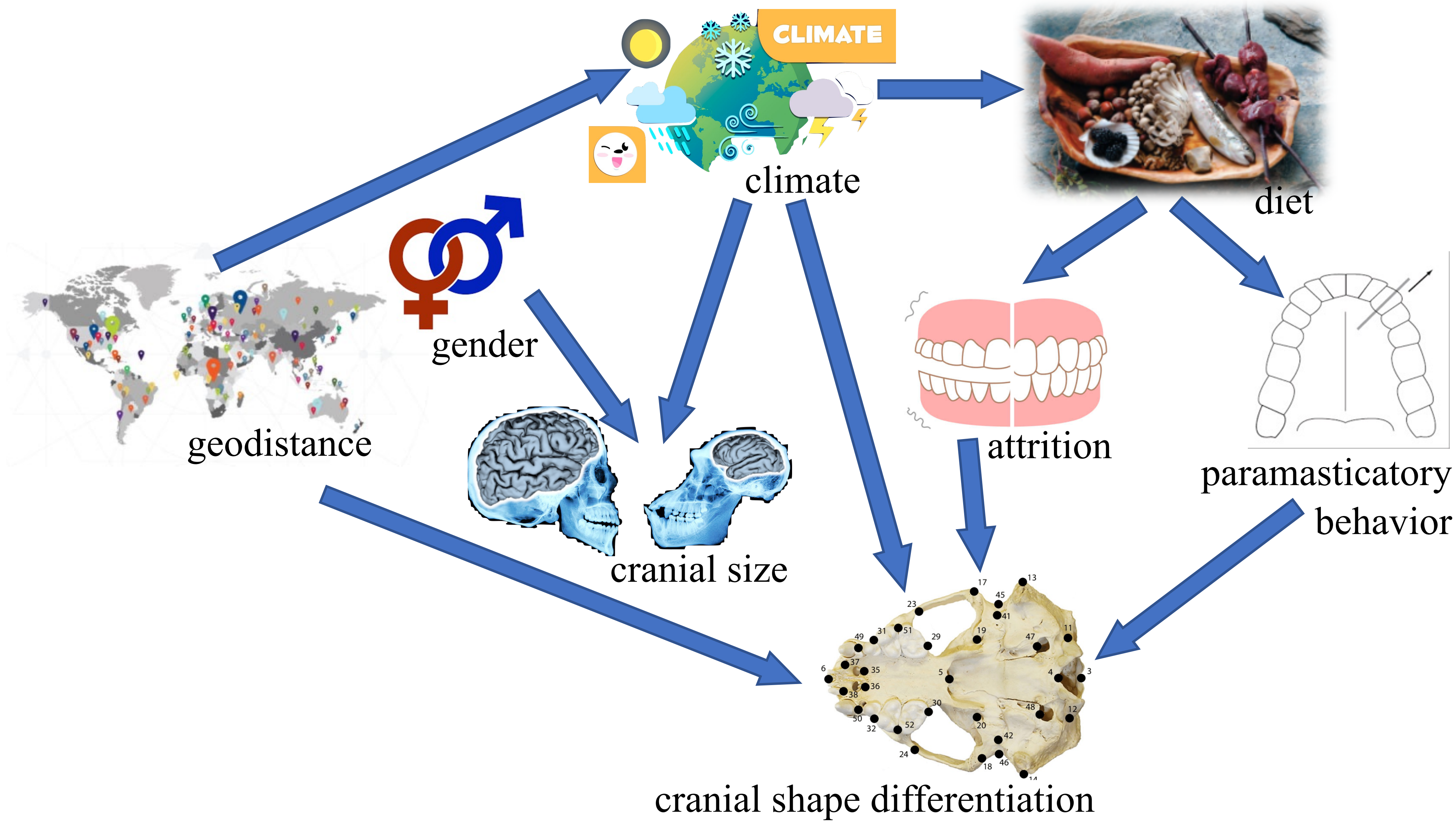


- Spirtes, Glymour, and Scheines. Causation, Prediction, and Search. 1993.

Result of PC on the Archeology Data

Thanks to collaborator Marlijn Noback

- By PC algorithm (Spirtes et al., 1993) + kernel-based conditional independence test (Zhang et al., 2011)

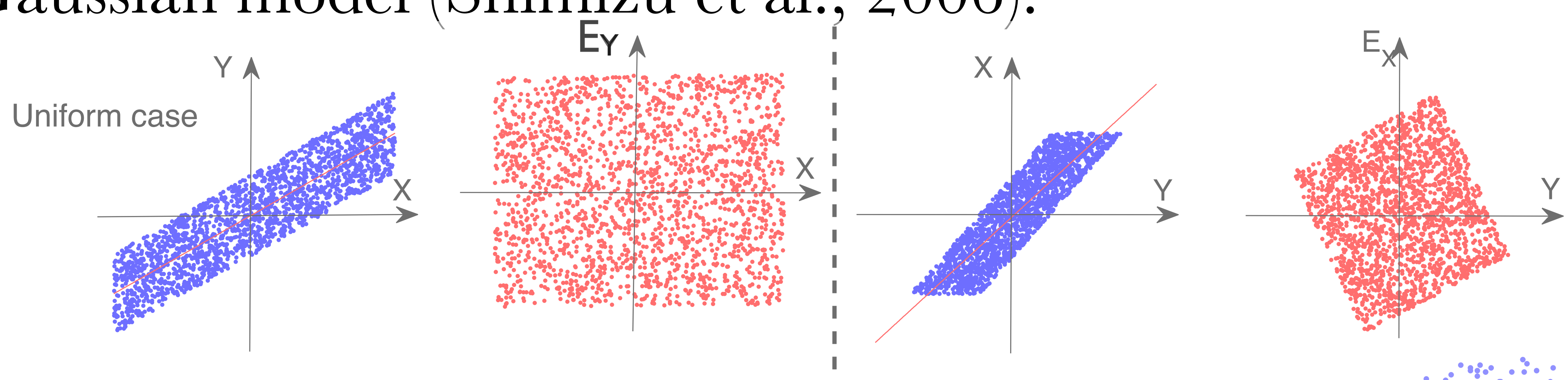


Functional Causal Model-Based Causal Discovery

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

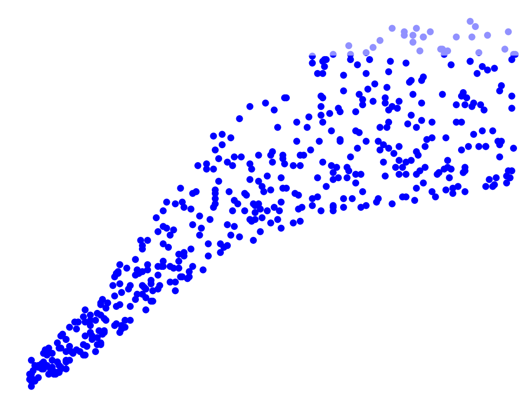
"Independent changes" renders causal direction identifiable

- Linear non-Gaussian model (Shimizu et al., 2006):
  $Y = aX + E$



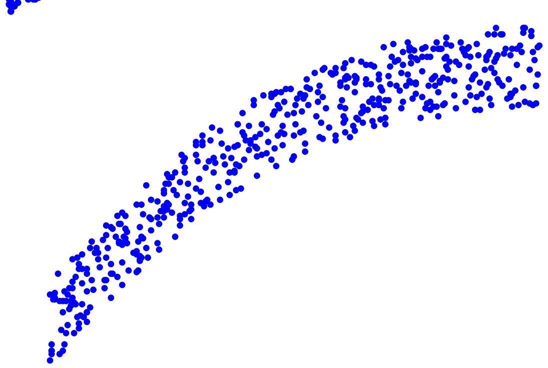
- Post-nonlinear causal model (Zhang & Chan, 2006):

$$Y = f_2\,(\,f_1(X) + E\,)$$

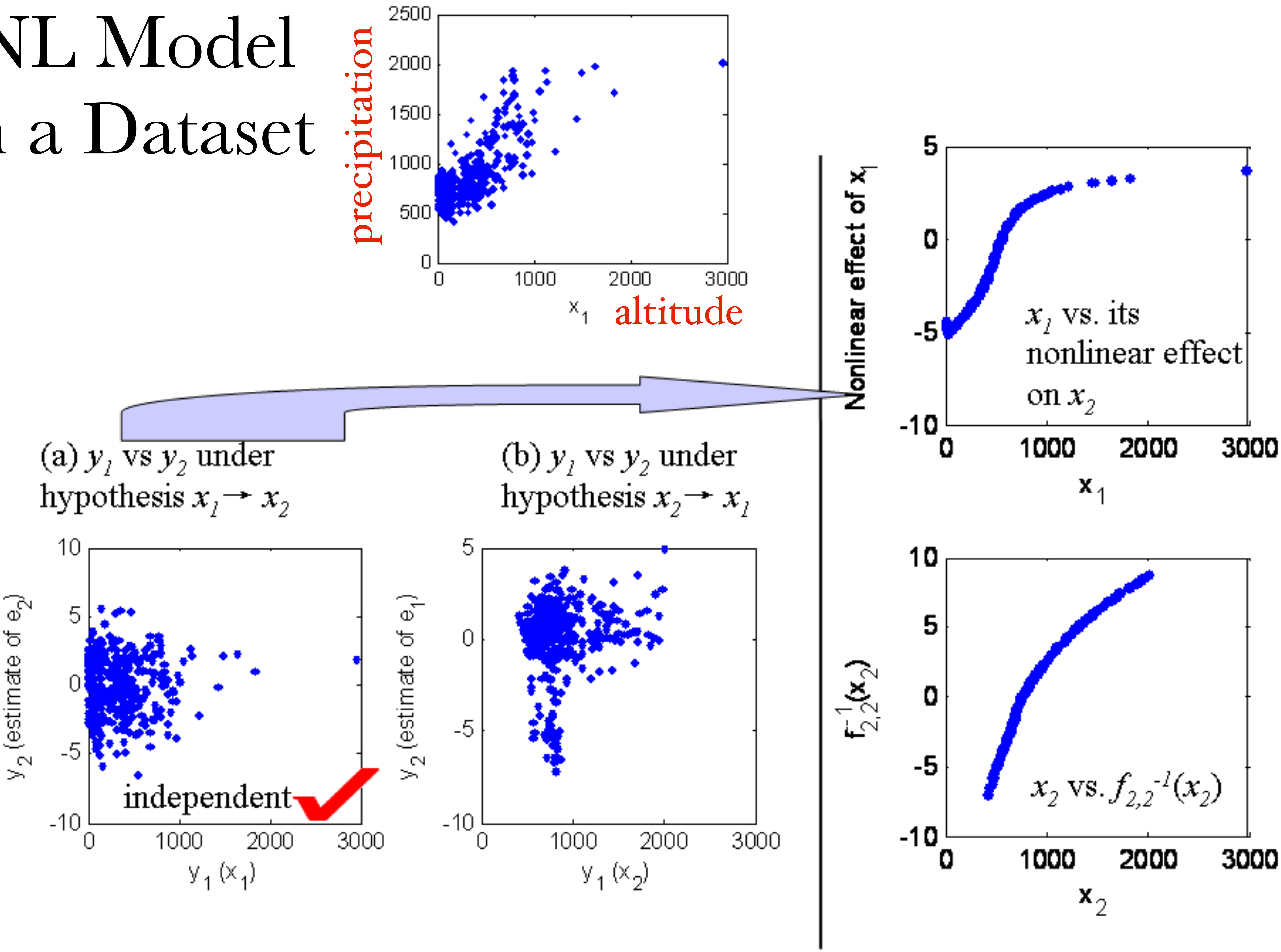


- Additive noise model (Hoyer et al, 2009)

$$Y = f(X) + E$$



PNL Model on a Dataset



A Problem in Psychology: Finding Underlying Mental Conditions?

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

- 50 questions for big 5 personality test

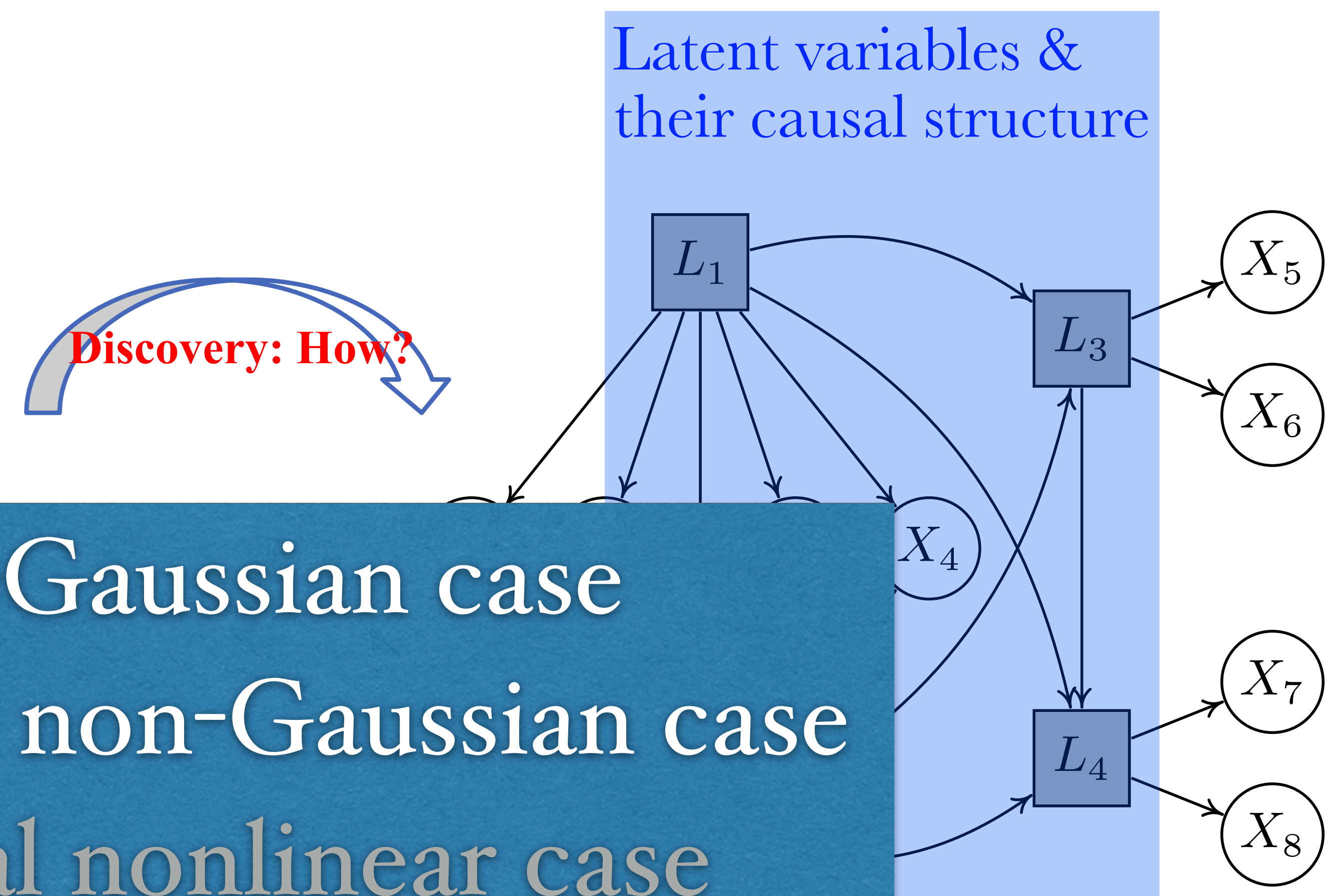
race	age	engnat	gender	hand	source	country	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	A1	A2	A3	A4	A5
3	53	1	1	1	1	US	4	2	5	2	5	1	4	3	5	1	1	5	2	5	1	1	1	1	1	1	1	5	1	5	2
13	46	1	2	1	1	US	2	2	3	3	3	3	1	5	1	5	2	3	4	2	3	4	3	2	2	4	1	3	3	4	4
1	14	2	2	1	1	PK	5	1	1	4	5	1	1	5	5	1	5	1	5	5	5	5	5	5	5	5	5	1	5	5	1
3	19	2	2	1	1	RO	2	5	2	4	3	4	3	4	4	5	5	4	4	2	4	5	5	5	4	5	2	5	4	4	3
11	25	2	2	1	2	US	3	1	3	3	3	1	3	1	3	5	3	3	3	4	3	3	3	3	3	4	5	5	3	5	1
13	31	1	2	1	2	US	1	5	2	4	1	3	2	4	1	5	1	5	4	5	1	4	4	1	5	2	2	2	3	4	3
5	20	1	2	1	5	US	5	1	5	1	5	1	5	4	4	1	2	4	2	4	2	2	3	2	2	2	5	5	1	5	1
4	23	2	1	1	2	IN	4	3	5	3	5	1	4	3	4	3	1	4	4	4	1	1	1	1	1	1	2	5	1	4	3
5	39	1	2	3	4	US	3	1	5	1	5	1	5	2	5	3	2	4	5	3	3	5	5	4	3	3	1	5	1	5	1
3	18	1	2	1	5	US	1	4	2	5	2	4	1	4	1	5	5	2	5	2	3	4	3	2	3	4	2	3	1	4	2
3	17	2	2	1	1	IT	1	5	2	5	1	4	1	4	1	5	5	3	5	3	2	5	3	3	4	3	2	4	2	4	1
13	15	2	1	1	1	IN	3	3	5	3	3	3	2	4	3	3	1	5	3	3	2	3	2	3	2	4	4	4	2	2	5
13	22	1	2	1	2	US	3	3	4	2	4	2	2	3	4	3	3	3	3	3	2	2	4	4	2	3	1	4	1	5	1
3	21	1	2	1	5	US	1	3	2	5	1	1	1	5	1	5	5	3	5	2	5	5	3	2	5	3	1	1	1	4	2
3	28	2	2	1	2	US	3	3	3	4	3	2	2	4	3	5	2	4	4	4	4	4	2	2	3	2	1	4	2	4	2
3	21	1	1	1	5	US	2	3	2	3	3	1	1	3	4	4	2	4	2	4	1	2	2	2	2	2	4	2	4	2	5
13	19	1	2	1	2	FR	1	3	2	4	2	4	1	4	3	4	4	2	3	2	1	3	1	2	2	3	4	2	3	1	4
3	21	1	2	1	5	US	4	1	5	2	5	1	5	3	5	1	5	2	5	2	3	3	3	3	4	2	1	5	2	5	2

Learning Hidden Variables & Their Relations

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

- Measured variables (e.g., answer scores in psychometric questionnaires) were generated by causally related latent variables

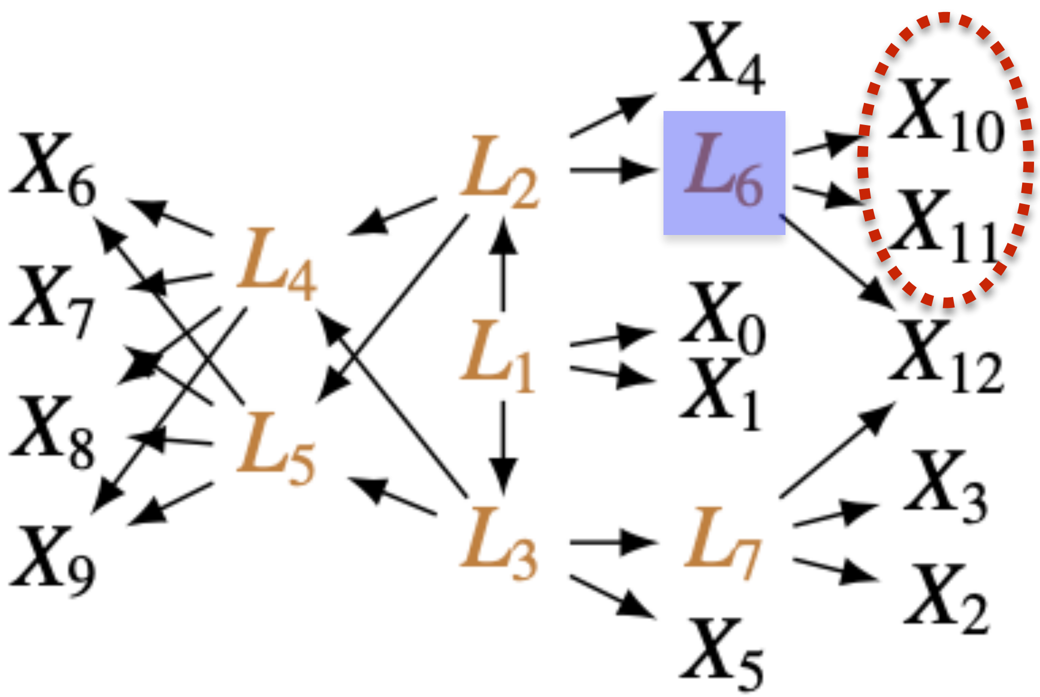
X1	X2	X3	X4	X5	X6	X7	X8
4.2	3.6	6.5	6.8	9.6	7.6	2.7	4.8
3.8	1.9	6.5	7.3	8.9	6.9	1.1	4.6
4.2	3.4	6.5	6.9	9.5	7.4	2.5	4.6
4.2	2.2	6.2	6.9	9.6	7.2	1.9	4.8
3.9	1.9	6.5	6.8	9.0	6.8	1.7	4.4
4.0	2.0	6.4	7.2	9			
3.8	1.7	6.4	7.3	9			
4.1	2.8	6.5	6.9	9			
...	...	...	...				



- Linear-Gaussian case
- Linear, non-Gaussian case
- General nonlinear case

- Find latent variables L_i and their causal relations from measured variables X_i ?

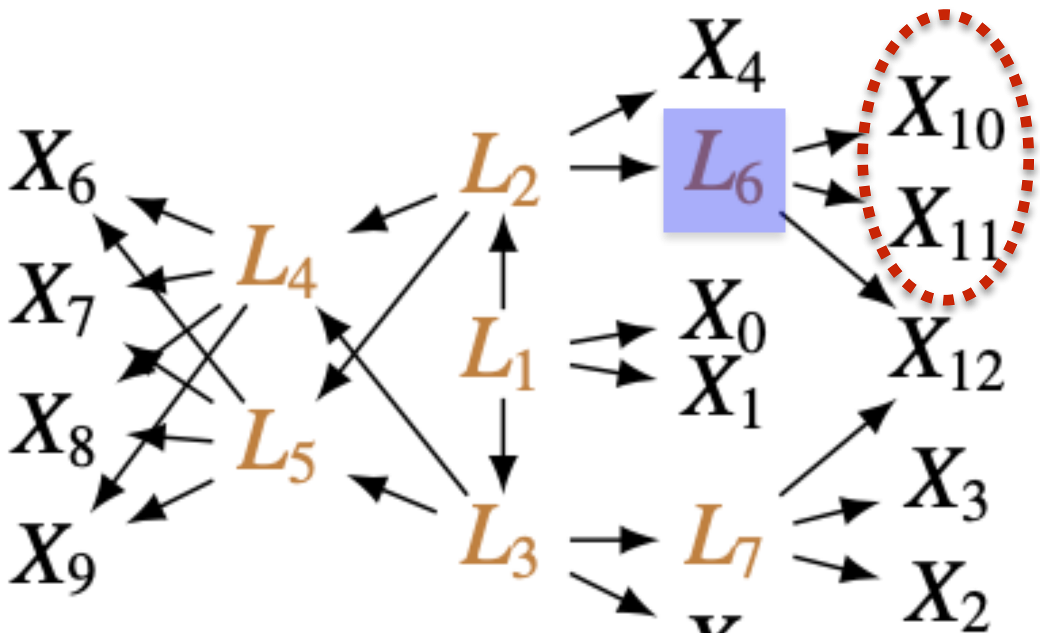
Linear, Gaussian Case: With Rank Deficiency Constraints



- Can we find L_6?
  - $\Sigma_{(X_{10},X_{11}),\ \mathbf{X}\backslash\{X_{10},X_{11}\}} = 1$
- Recovering the equivalence class
  - With rank deficiency of cross-covariance matrices
  - recursively and cleverly

- *Huang, Low, Xie, Glymour, Zhang, "Latent Hierarchical Causal Structure Discovery with Rank Constraints," NeurIPS 2022*

Linear, Gaussian Case: With Rank Deficiency Constraints



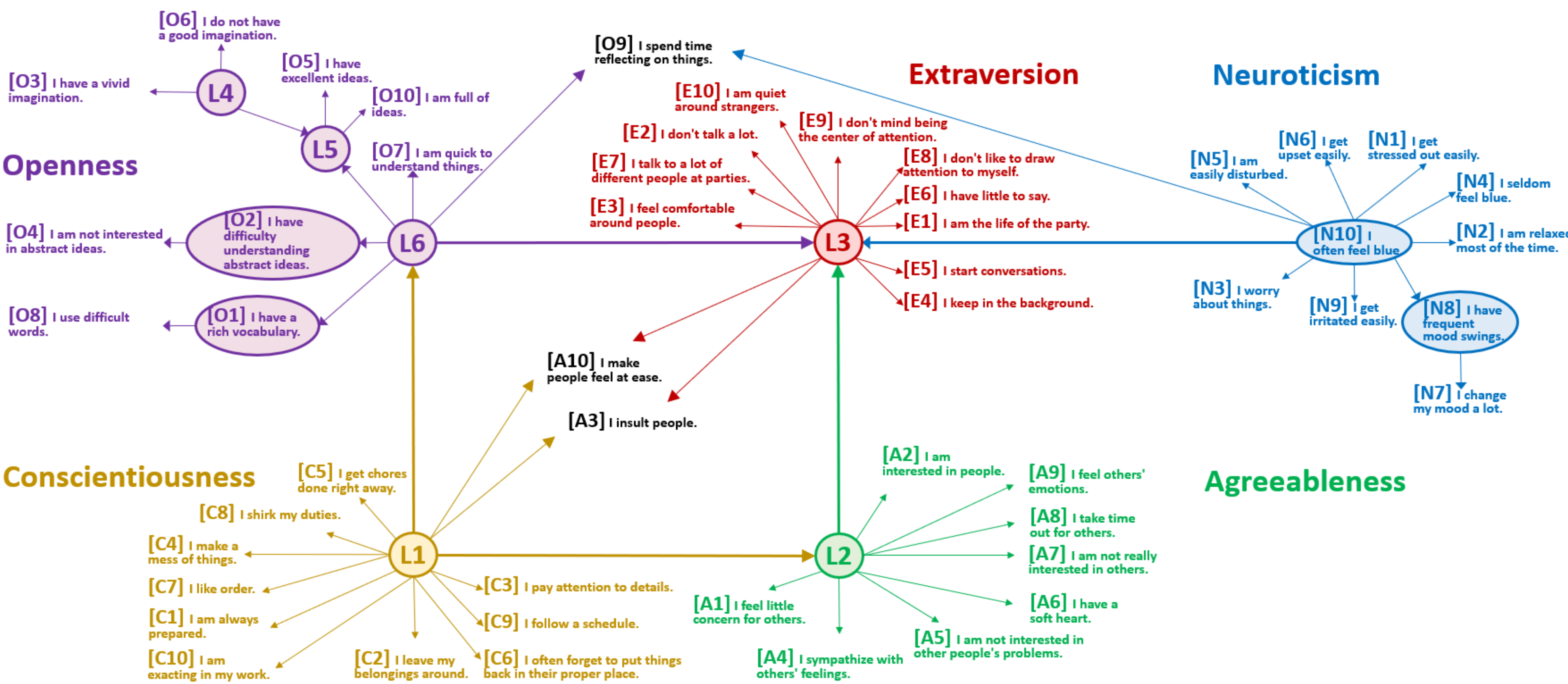
- Can we find L_6?
  - $\Sigma_{(X_{10},X_{11}),\,\mathbf{X}\setminus\{X_{10},X_{11}\}} = 1$
- Recovering the equivalence class
  - With rank deficiency of cross-covariance matrices

- Conditional independence is a special case
  - $\mathrm{rank}(\Sigma_{(X_1,\, X_2),\, (X_2,\, X_3)}) = 1 \Leftrightarrow X_1 \perp\!\!\!\perp X_3 \mid X_2$
- Unified causal discovery based on rank deficiency constraints

- *Dong, Huang, Ng, Song, Zheng, Jin, Legaspi, Spirtes, Zhang, "A Versatile Causal Discovery Framework to Allow Causally-Related Hidden Variables," ICLR 2024*

Example: Big 5 Questions Are Well Designed but…

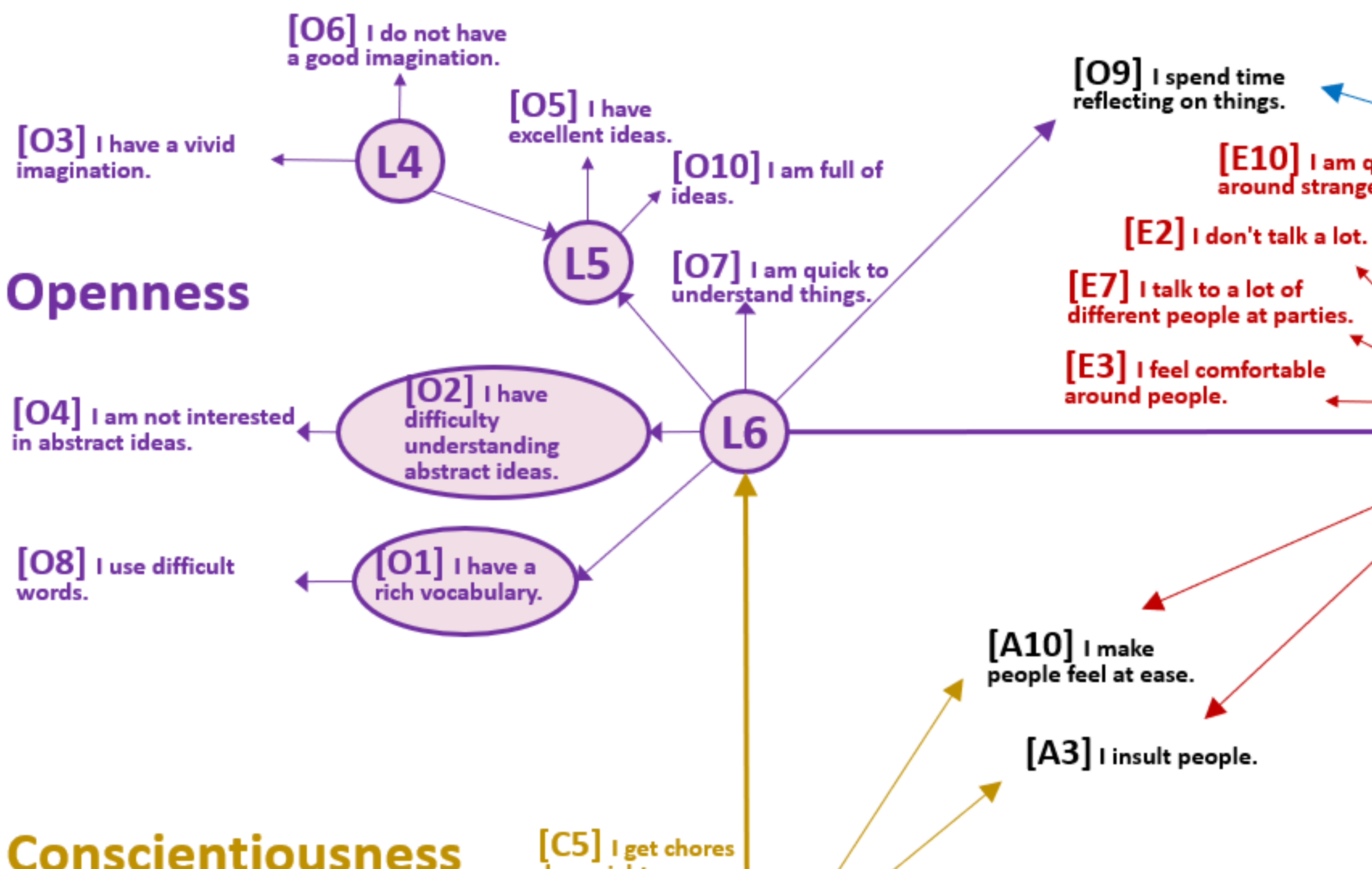
Big 5:
openness; conscientiousness; extraversion; agreeableness; neuroticism



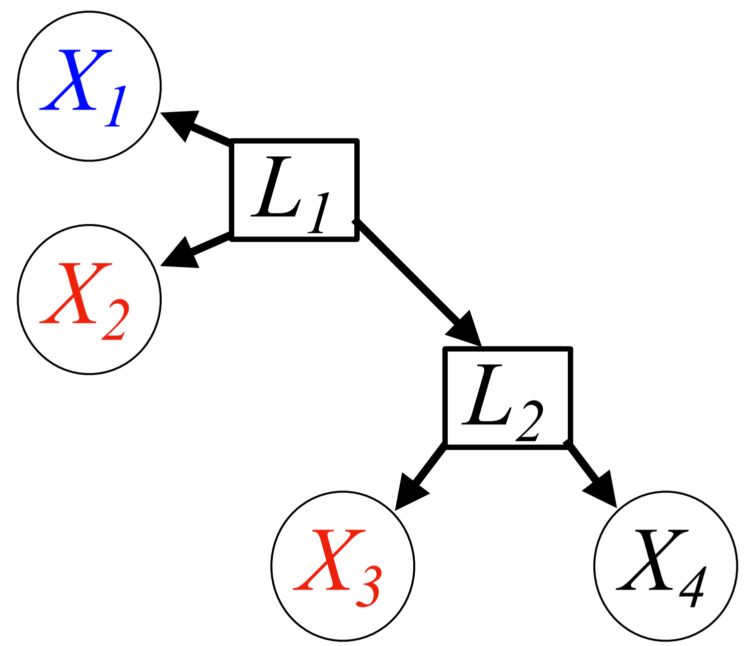
- *Dong, Huang, Ng, Song, Zheng, Jin, Legaspi, Spirtes, Zhang, "A Versatile Causal Discovery Framework to Allow Causally-Related Hidden Variables," ICLR 2024*

Example: Big 5 Questions Are Well Designed but…

Big 5:
openness; conscientiousness; extraversion; agreeableness; neuroticism



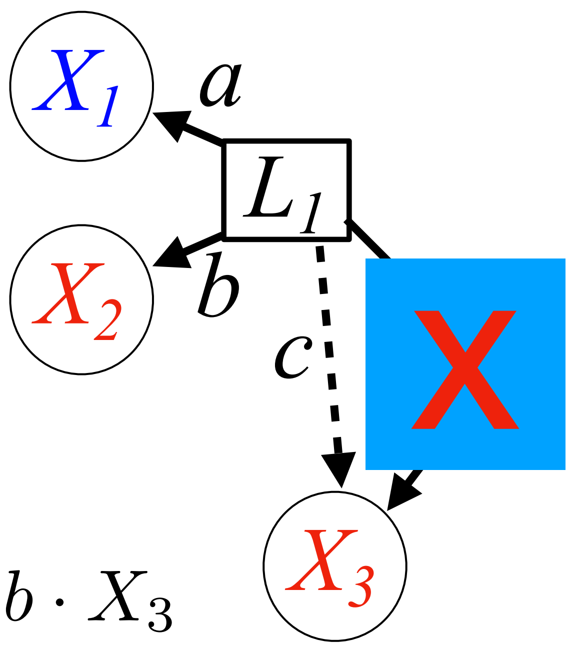
Linear, Non-Gaussian Case: Generalized Independent Noise Condition

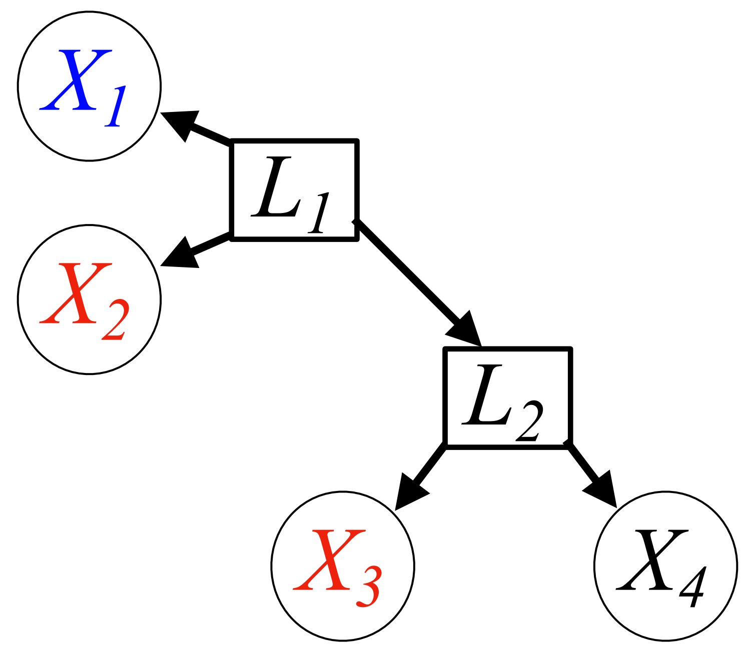


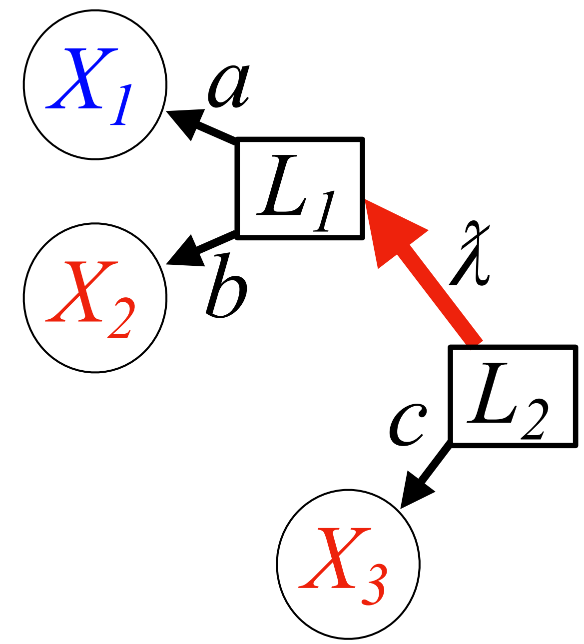
- Find direction between latent variables L_1 and L_2?

- *Xie, Cai, Huang, Glymour, Hao, Zhang, "Generalized Independent Noise Condition for Estimating Linear Non-Gaussian Latent Variable Causal Graphs," NeurIPS 2020*
- *Cai, Xie, Glymour, Hao, Zhang, "Triad Constraints for Learning Causal Structure of Latent Variables," NeurIPS 2019*

Linear, Non-Gaussian Case: Generalized Independent Noise Condition







$$\begin{aligned} & c \cdot X_2 - b \cdot X_3 \\ =& c(bL1 + E_2) - b(cL_1 + E_3) \\ =& cE_2 - bE_3, \end{aligned}$$

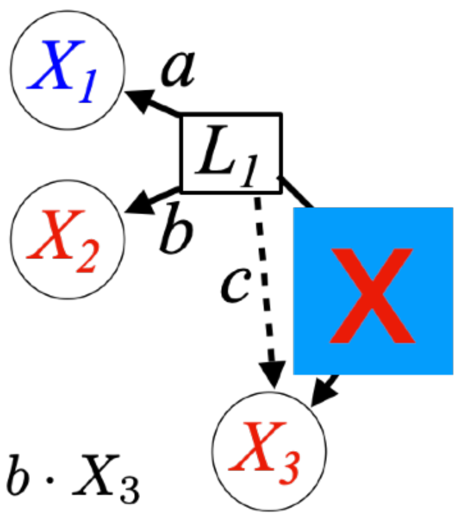
independent from L_1 and from X_1,

and we know $\frac{b}{c} = \frac{Cov(X_2, X_3)}{Cov(X_1, X_3)}$

Nontrivial linear combination of X_2 and X_3 will involve the noise term in L_1, hence dependent on X_1

- *Xie, et al., "Generalized Independent Noise Condition for Estimating Linear Non-Gaussian Latent Variable Causal Graphs," NeurIPS 2020*

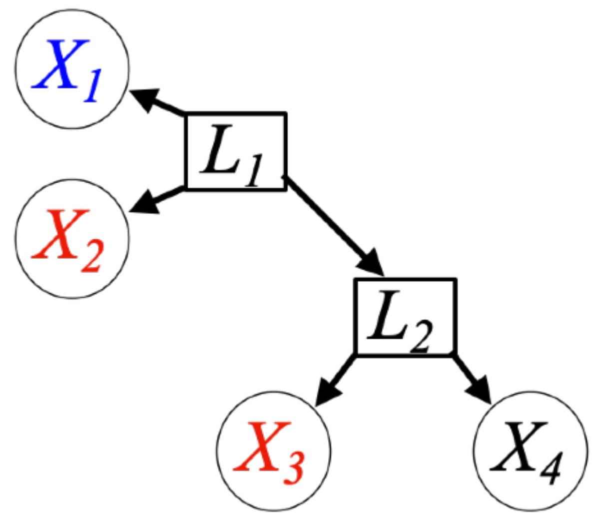
Linear, Non-Gaussian Case: Generalized Independent Noise Condition

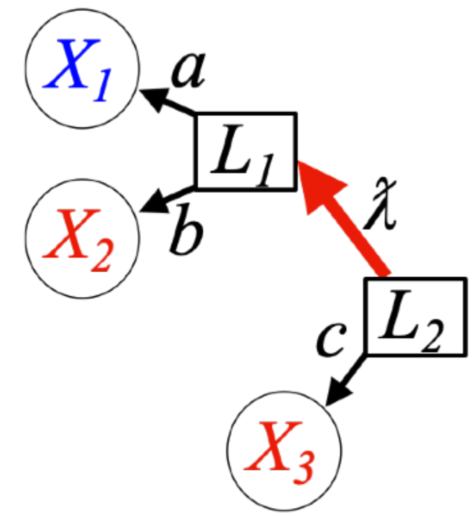


$c \cdot X_2 - b \cdot X_3$
$= c(bL1 + E_2) - b(cL_1 + E_3)$
$= cE_2 - bE_3,$

independent from L_1 and from X_1,

and we know $\frac{b}{c} = \frac{Cov(X_2, X_3)}{Cov(X_1, X_3)}$





Nontrivial linear combination of X_2 and X_3 will involve the noise term in L_1, hence dependent on X_1

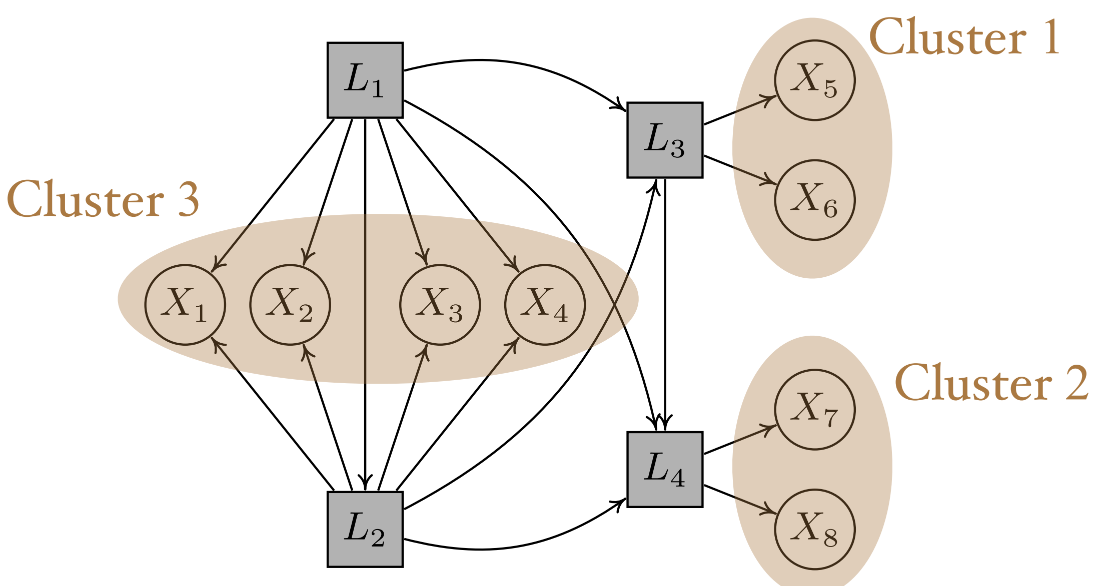
Let Z = {X_1} and Y = {X_2, X_3}, GIN!

- GIN condition: (**Z**, **Y**) follows GIN $\Leftrightarrow w^\intercal \mathbf{Y} \perp\!\!\!\perp \mathbf{Z}$ for nonzero w
  - has graphical implications

GIN for Estimating Linear, Non-Gaussian LV Model

- A two-step algorithm to identify the latent variable graph
  - By testing for GIN conditions over the input $X_1, \cdots, X_8$

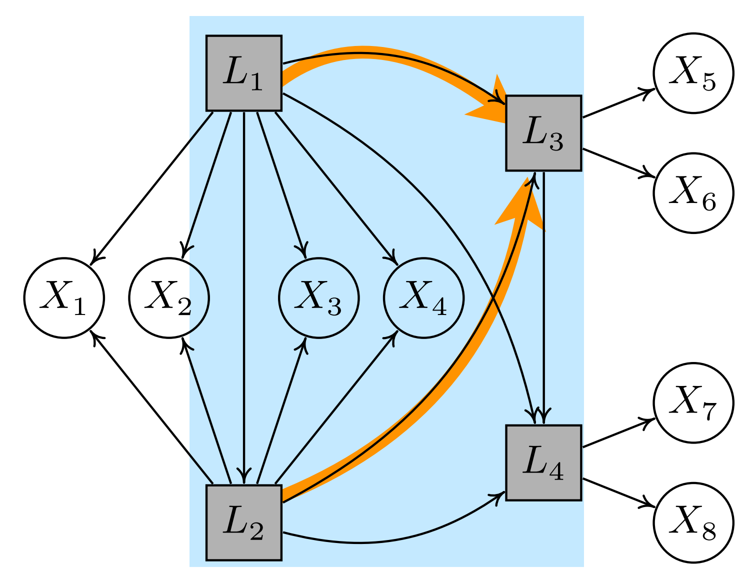
Step 1: find ***causal clusters***



$$(\overbrace{\{X_1, \cdots, X_4, X_7, X_8\}}^{\color{blue}{\mathbf{Z}}}, \overbrace{\{X_5, X_6\}}^{\color{red}{\mathbf{Y}}})$$

satisfies GIN condition

Step 2: determine ***causal structure*** of the latent variables



$$(\overbrace{\{X_3, X_4\}}^{\color{blue}{\mathbf{Z}}}, \overbrace{\{X_1, X_2, X_5\}}^{\color{red}{\mathbf{Y}}})$$

Cluster 3 Cluster 1 & 3

satisfies GIN condition

GIN-Based Method: Application to Teacher's Burnout Data

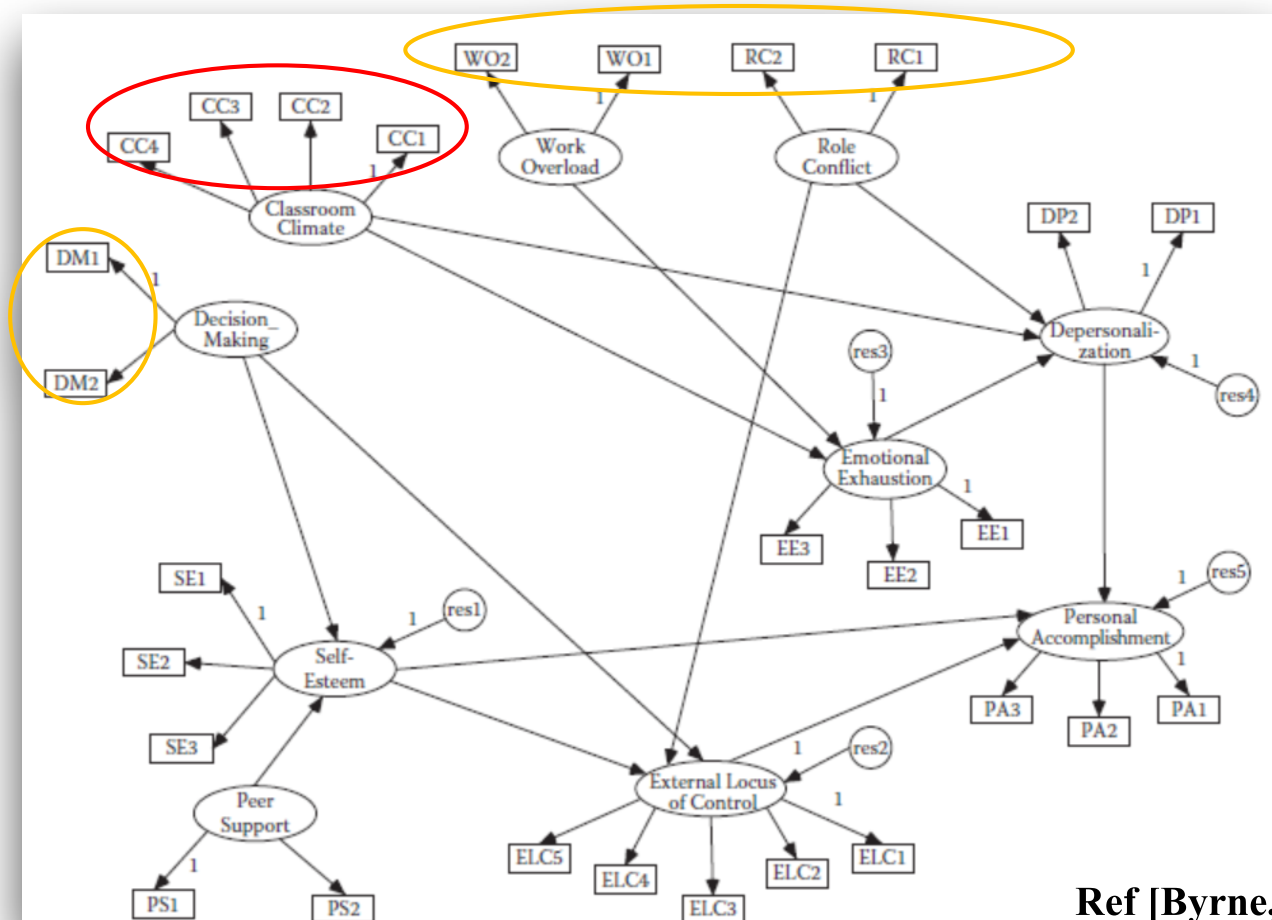
- Contains 28 measured variables
- Discovered clusters and causal order of the latent variables:

Causal Clusters	Observed variables
$\mathcal{S}_1$ (1)	$RC_1, RC_2, WO_1, WO_2, DM_1, DM_2$
$\mathcal{S}_2$ (1)	CC_1, CC_2, CC_3, CC_4
$\mathcal{S}_3$ (1)	PS_1, PS_2
$\mathcal{S}_4$ (1)	$ELC_1, ELC_2, ELC_3, ELC_4, ELC_5$
$\mathcal{S}_5$ (2)	$SE_1, SE_2, SE_3, EE_1, EE_2, EE_3, DP_1, PA_3$
$\mathcal{S}_6$ (3)	DP_2, PA_1, PA_2

$$L(\mathcal{S}_1) \succ L(\mathcal{S}_2) \succ L(\mathcal{S}_3) \succ L(\mathcal{S}_5) \succ L(\mathcal{S}_4) \succ L(\mathcal{S}_6).$$

(from root to leaf)

Hypothesized model by experts



Ref [Byrne, 2010]

- Consistent with the hypothesized model

- *Xie, Cai, Huang, Glymour, Hao, Zhang, "Generalized Independent Noise Condition for Estimating Linear Non-Gaussian Latent Variable Causal Graphs," NeurIPS 2020*
- *Cai, Xie, Glymour, Hao, Zhang, "Triad Constraints for Learning Causal Structure of Latent Variables," NeurIPS 2019*

Where Are We?

<table>
<tr><th>i.i.d. data?</th><th>Parametric constraint?</th><th>Latent confounders?</th><th>What can we get?</th></tr>
<tr><td rowspan="4">Yes</td><td rowspan="2">No</td><td>No</td><td rowspan="2">(Different types of) equivalence class</td></tr>
<tr><td>Yes</td></tr>
<tr><td rowspan="2">Yes</td><td>No</td><td rowspan="2">Unique identifiability (under structural conditions)</td></tr>
<tr><td>Yes</td></tr>
<tr><td rowspan="2">Non-I, but I.D.</td><td rowspan="2">No/Yes</td><td>No</td><td rowspan="6">?</td></tr>
<tr><td>Yes</td></tr>
<tr><td rowspan="4">I., but non-I.D.</td><td>No</td><td rowspan="2">No</td></tr>
<tr><td>Yes</td></tr>
<tr><td>No</td><td rowspan="2">Yes</td></tr>
<tr><td>Yes</td></tr>
</table>

Learning Latent Causal Dynamics

i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes

"Time-delayed" influence renders latent processes & their relations identifiable

Time-series Inputs $\{x_t\}_{t=0}^{T}$

Unsupervised Representation Learning

$$\mathbf{x}_t = \mathbf{g}(\mathbf{z}_t)$$

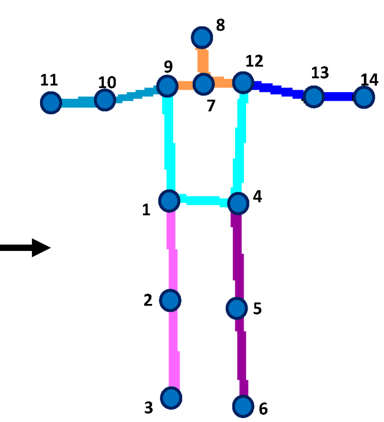
Latent processes

Temporal VAE with causal prior

Latent temporal causal processes z_{it} follow

- completely **nonparametric** model; or furthermore,
- **non-stationary** noise or causal influence, or
- **Parametric** constraints

Causal Skeleton Recovery

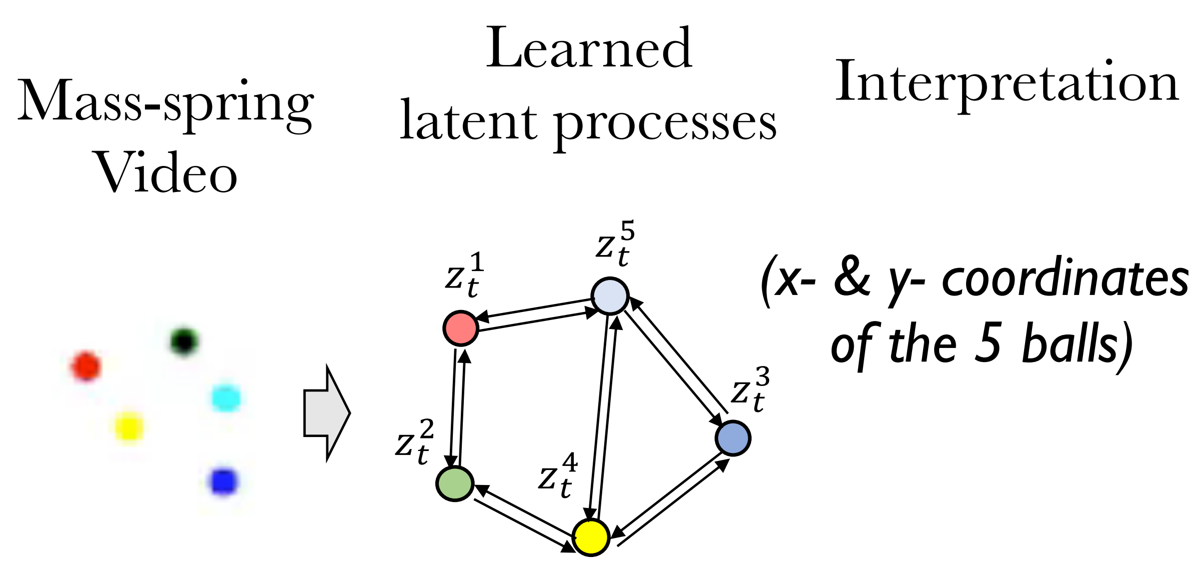


Recovered latent processes

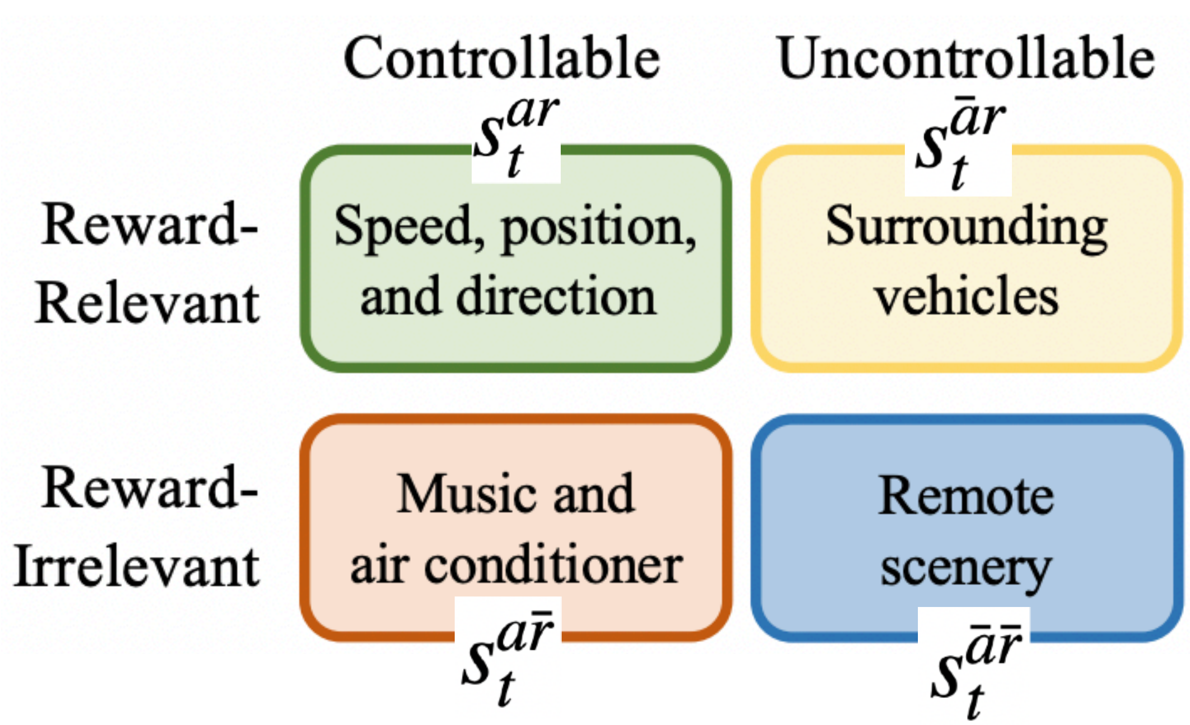
- *Yao, Chen, Zhang, "Causal Disentanglement for Time Series," NeurIPS 2022*
- *Yao, Sun, Ho, Sun, Zhang, "Learning Temporally causal latent processes from general temporal data," ICLR 2022*

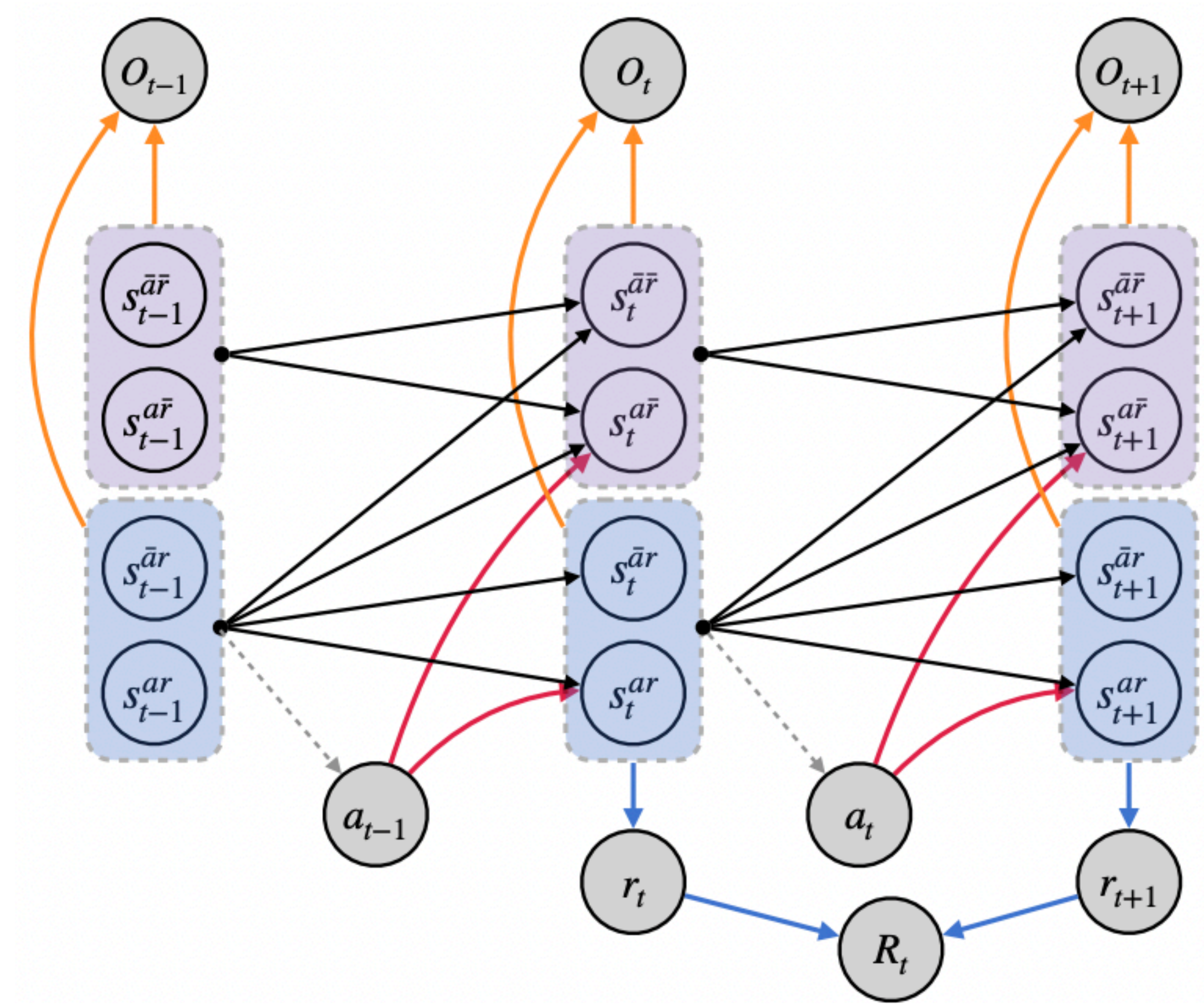
Results on Simple Video Data

- For easy interpretation, consider a simple video data set
  - Mass-spring system: a video dataset with ball movement and invisible springs



Extension: Four Categories of State Representations in RL





Each category is identifiable!

- Liu*, Huang*, Zhu, Tian, Gong, Yu, Zhang. Learning world models with identifiable factorization. NeurIPS 2023

Where Are We?

<table>
<tr><th>i.i.d. data?</th><th>Parametric constraints?</th><th>Latent confounders?</th><th>What can we get?</th></tr>
<tr><td rowspan="4">Yes</td><td rowspan="2">No</td><td>No</td><td rowspan="2">(Different types of) equivalence class</td></tr>
<tr><td>Yes</td></tr>
<tr><td rowspan="2">Yes</td><td>No</td><td rowspan="2">Unique identifiability (under structural conditions)</td></tr>
<tr><td>Yes</td></tr>
<tr><td rowspan="2">Non-I, but I.D.</td><td rowspan="2">No/Yes</td><td>No</td><td>(Extended) regression</td></tr>
<tr><td>Yes</td><td>Latent temporal causal processes identifiable!</td></tr>
<tr><td rowspan="4">I., but non-I.D.</td><td>No</td><td rowspan="2">No</td><td rowspan="4">?</td></tr>
<tr><td>Yes</td></tr>
<tr><td>No</td><td rowspan="2">Yes</td></tr>
<tr><td>Yes</td></tr>
</table>

(Automated) Scientific Discovery: A Story

- Consider puerperal fever in the mid-19th century
- Two clinics used almost the same techniques but had very different mortality rates
- Semmelweis discovered the only major difference was the individuals who worked there
  - *Hypothesis*: Unknown "cadaverous material" caused puerperal fever
  - Proposed *intervention*: washing hands
  - *Conflicted* with the established scientific and medical opinions of the time
  - Rejected by the medical community until years after his death, when Louis Pasteur *confirmed* the germ theory

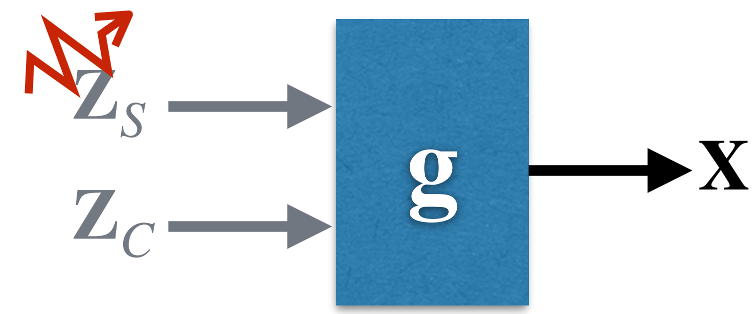
https://amol-kulkarni.com/project/semmelweis/

Ignaz Semmelweis

Semmelweis, aged 42 in 1860, photograph by Borsos and Doctor

Finding Changing Hidden Variables for Transfer Learning

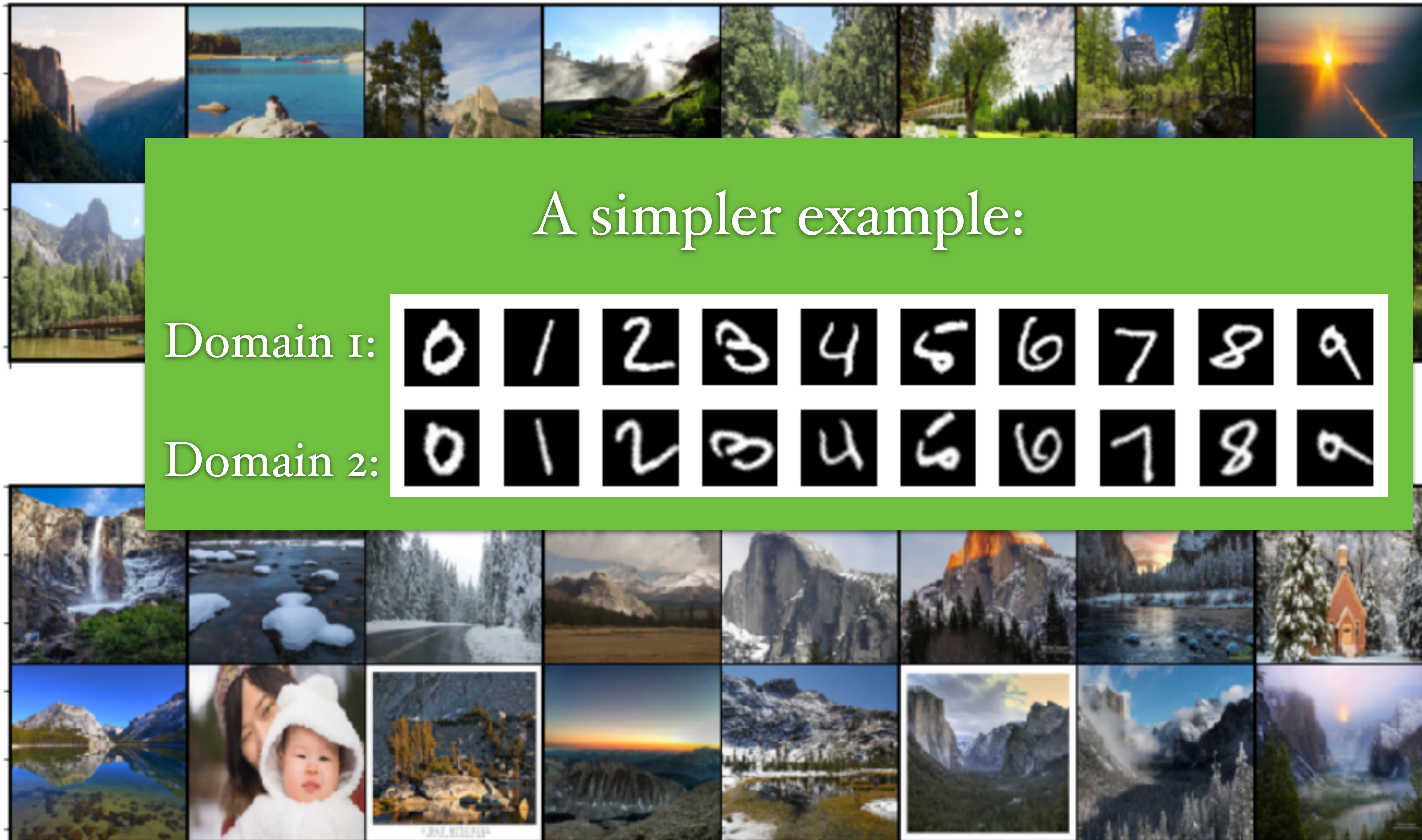
i.i.d. data?	Parametric constraints?	Latent confounders?
Yes	No	No
No	Yes	Yes



- Underlying components $\mathbf{Z}_S$ may change across domains
- Changing components $\mathbf{Z}_S$ are identifiable; invariant part $\mathbf{Z}_C$ is identifiable up to its subspace
- Using invariant part $\mathbf{Z}_C$ and transformed changing part $\tilde{\mathbf{Z}}_S$ for transfer learning

- *Kong, Xie, Yao, Zheng, Chen, Stojanov, Akinwande, Zhang, Partial disentanglement for domain adaptation, ICML 2022*

Image Translation: How to Learn 'Style'?



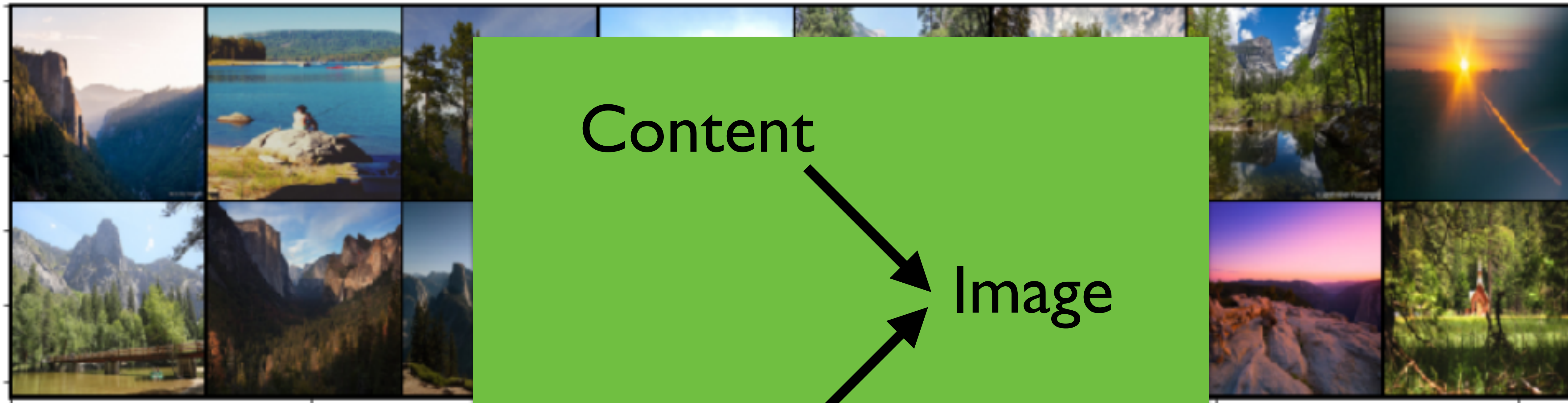
A simpler example:

Domain 1:

Domain 2:

Images from the winter season domain.

Image Translation Based on Minimal Changes

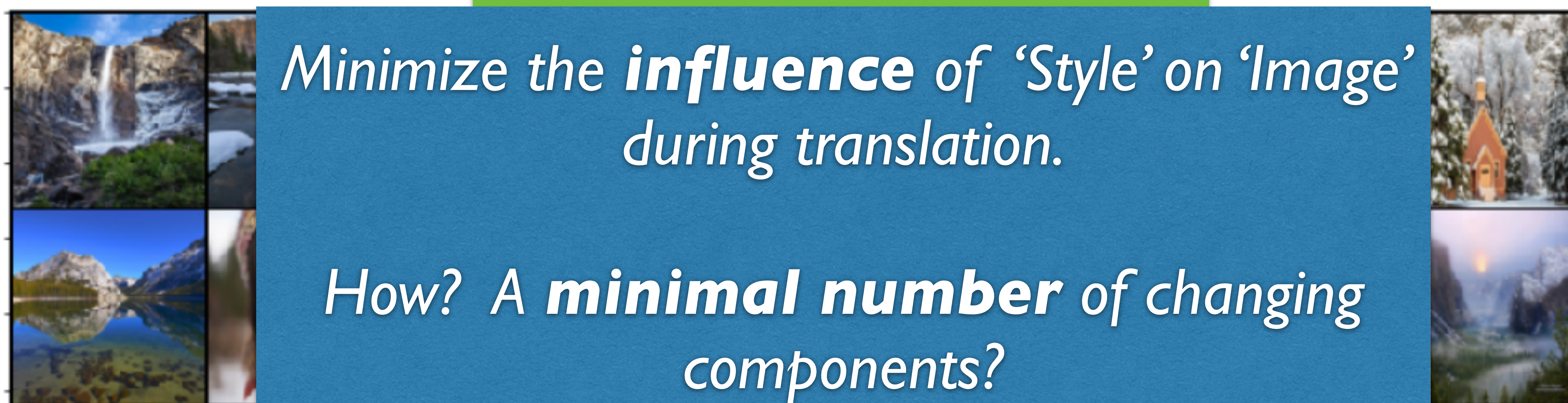


Content

Image

Style

Im ain.



*Minimize the **influence** of ‘Style’ on ‘Image’ during translation.*

*How? A **minimal number** of changing components?*

Images from the winter season domain.

Multi-domain Image Generation & Translation with Identifiability Guarantees

- Idea: Matching the distributions across domains **with a minimal number of changing components**
- Correspondence info (joint distribution) identifiable under mild assumptions
- Example: Generating female & male images with the same "content"

Ours

StyleGAN2-ADA

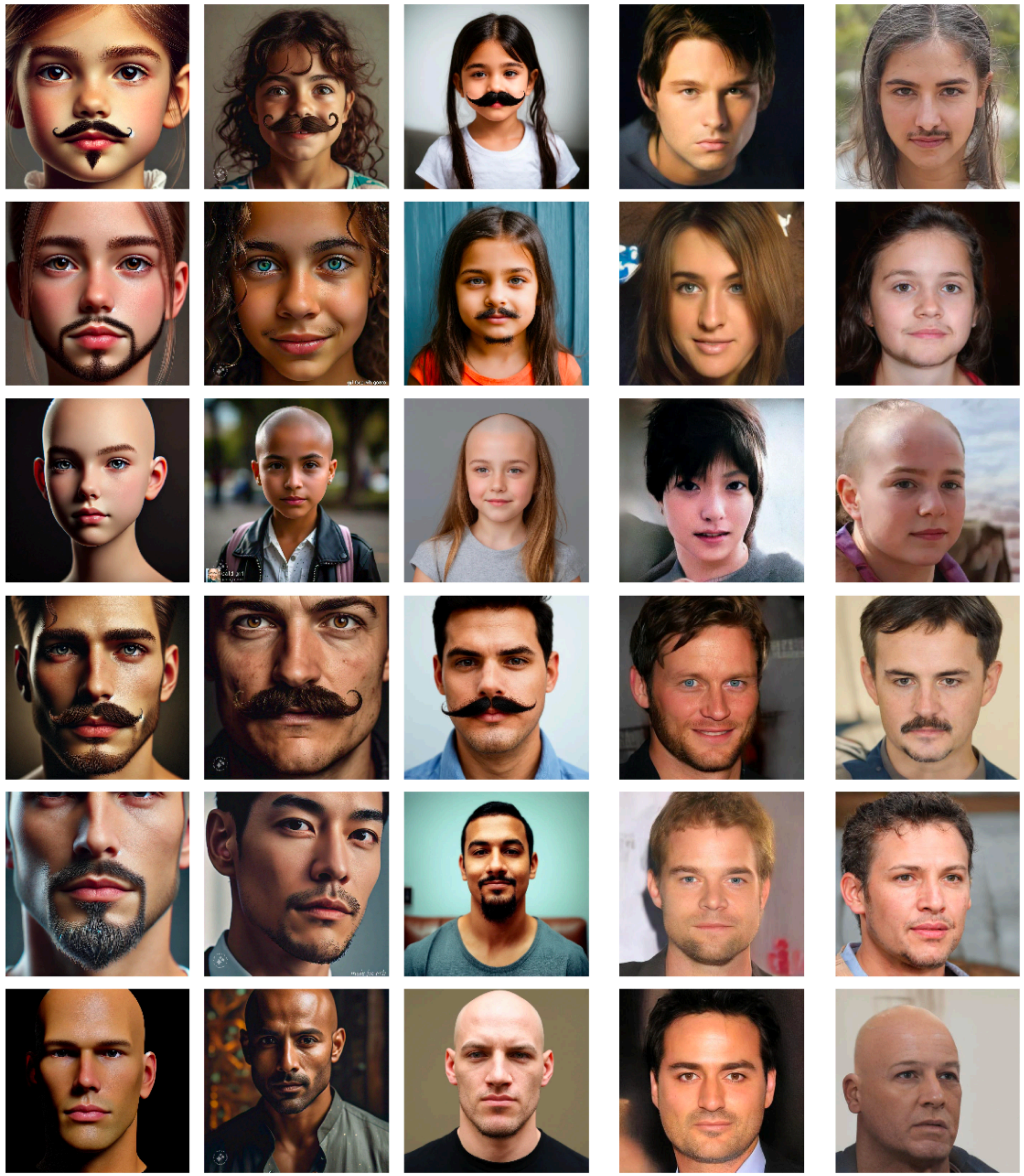
TGAN

- *Xie, Kong, Gong, Zhang, "Multi-domain image generation and translation with identifiability guarantees", ICLR 2023*
- *Yan, Kong, Gui, Chi, Xing, He, Zhang, Counterfactual Generation with Identifiability Guarantee, NeurIPS 2023*
- *Kong, Xie, Yao, Zheng, Chen, Stojanov, Akinwande, Zhang, Partial disentanglement for domain adaptation, ICML 2022*

Minimal Changes + *Causal Modeling* for Generation

GPT-4o | Meta AI | Stable 3 | FaceDiffusion | Ours

Corresponding to prompt:



girl with mustache,

girl with goatee,

bald girl,

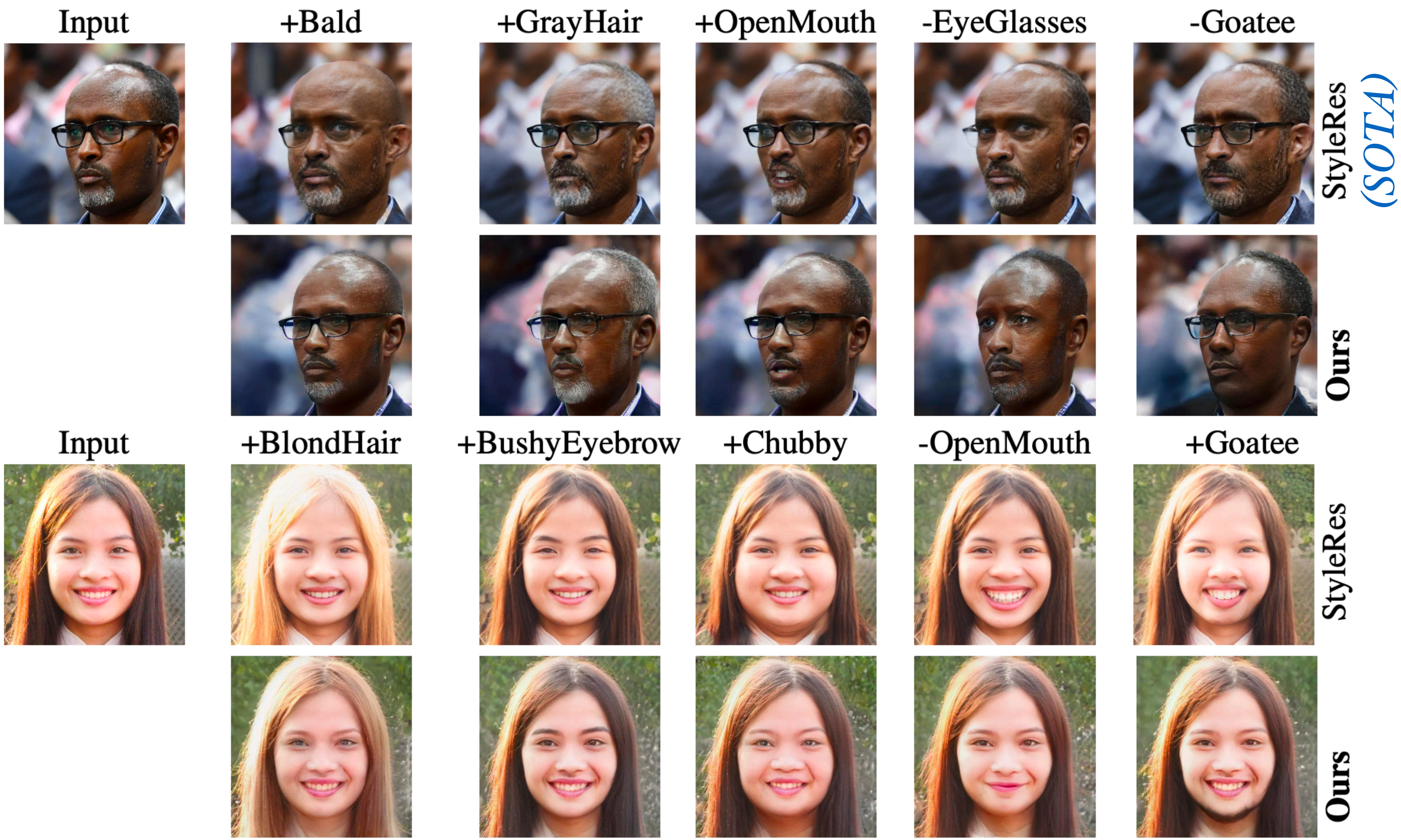
male with mustache,

male with goatee, and

bald male

- S. Xie, Y. Zheng, I. Ng, K. Zhang, Causal Compositional Image Generation with Minimal Change, under submission

For Real Image Editing



Summary

- Various tasks involve suitable (causal) representations of data
  - Scientific discovery, domain generalization/adaptation, trustworthy AI, explainable AI, fairness…
- Causal representations can be recovered under appropriate assumptions
  - Technically operational causal principles
  - Identifiability!
    - Strong identifiability results in non-IID cases
    - Benefit from parametric constraints in the IID case
- In the era of large models: Is causality essential?

HOW?