

Causality: the evitable C-word?

Jianle Sun

Department of Bioinformatics & Biostatistics,
Shanghai Jiao Tong University

Motivation

The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data

Causal inference is a core task of science. However, authors and editors often refrain from explicitly acknowledging the causal goal of research projects; they refer to causal effect estimates as associational estimates.

This commentary argues that using the term “causal” is necessary to improve the quality of

Miguel A. Hernán, MD, DrPH



See also Galea and Vaughan, p. 602; Begg and March, p. 620; Ahern, p. 621; Chiolero, p. 622; Glymour and Hamad, p. 623; Jones and Schooling, p. 624; and Hernán, p. 625.

You know the story:

Dear author: Your observational study cannot prove causation. Please replace all references to causal effects by references to associations.

Confusion then ensues at the most basic levels of the scientific process and, inevitably, errors are made.

We need to stop treating “causal” as a dirty word that

glass of red wine per day versus no alcohol drinking. For simplicity, disregard measurement error and random variability—that is, suppose the 0.8 comes from a very large population so that the 95%



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



Outline

- Why causality
- A brief history on causality studies
- Two causal models
- Three domains
- Causality in the age of artificial intelligence

Why causality?



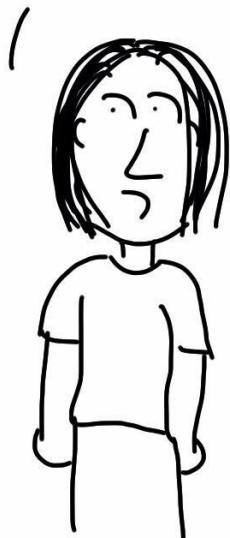
上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



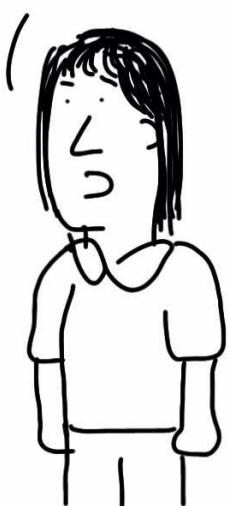
A scientific cliché

- Correlation ≠ dependence ≠ causality

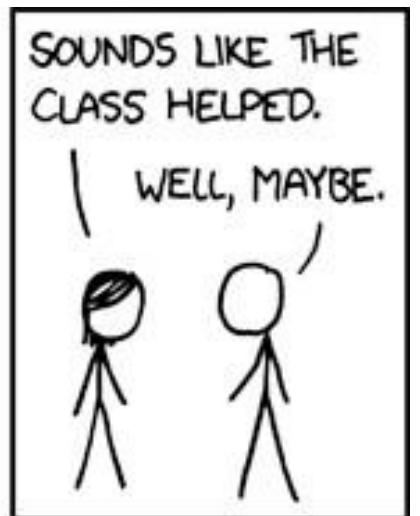
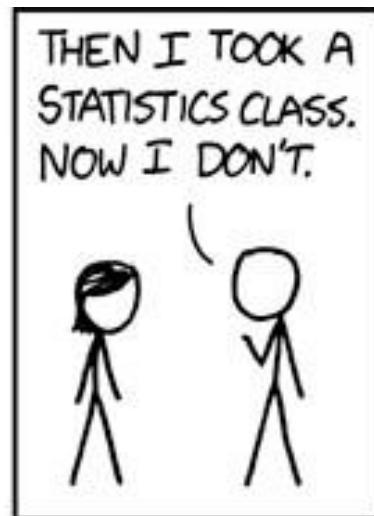
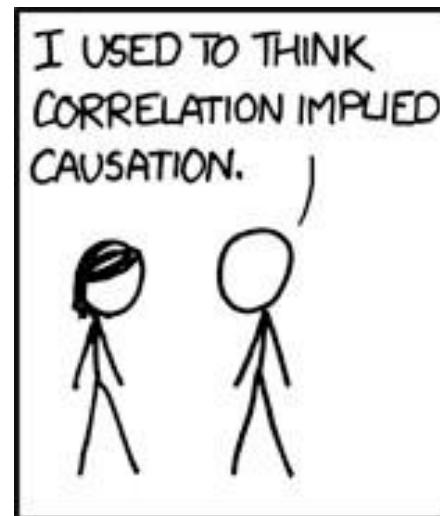
Everyone knows,
Correlation is not Causation



Sure, that's why
we have causal inference



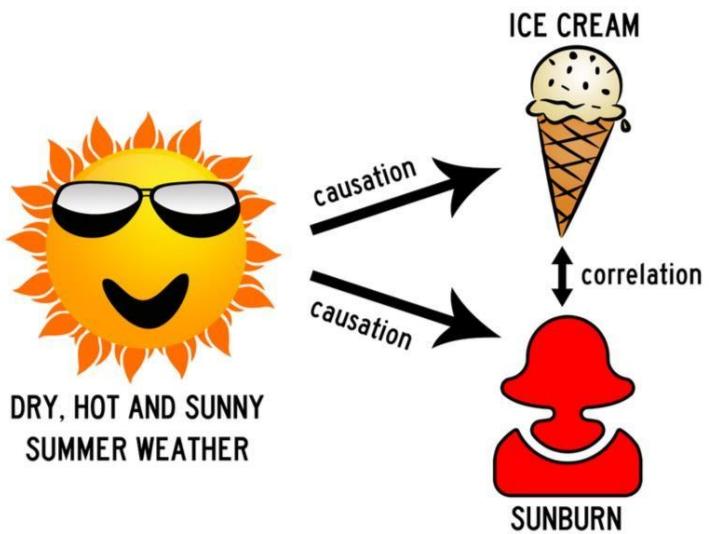
freshspectrum.com



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

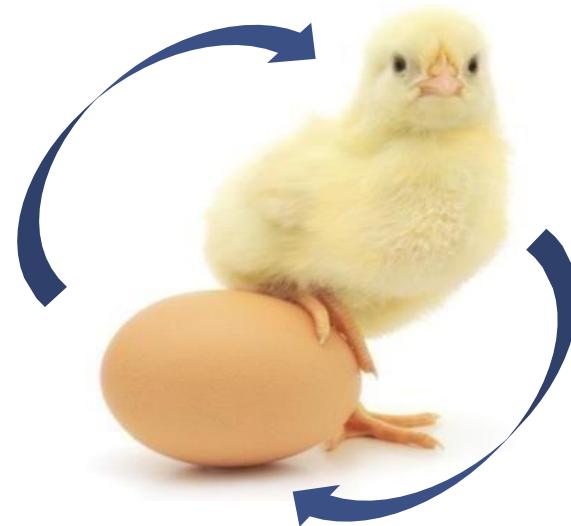


Causal inference



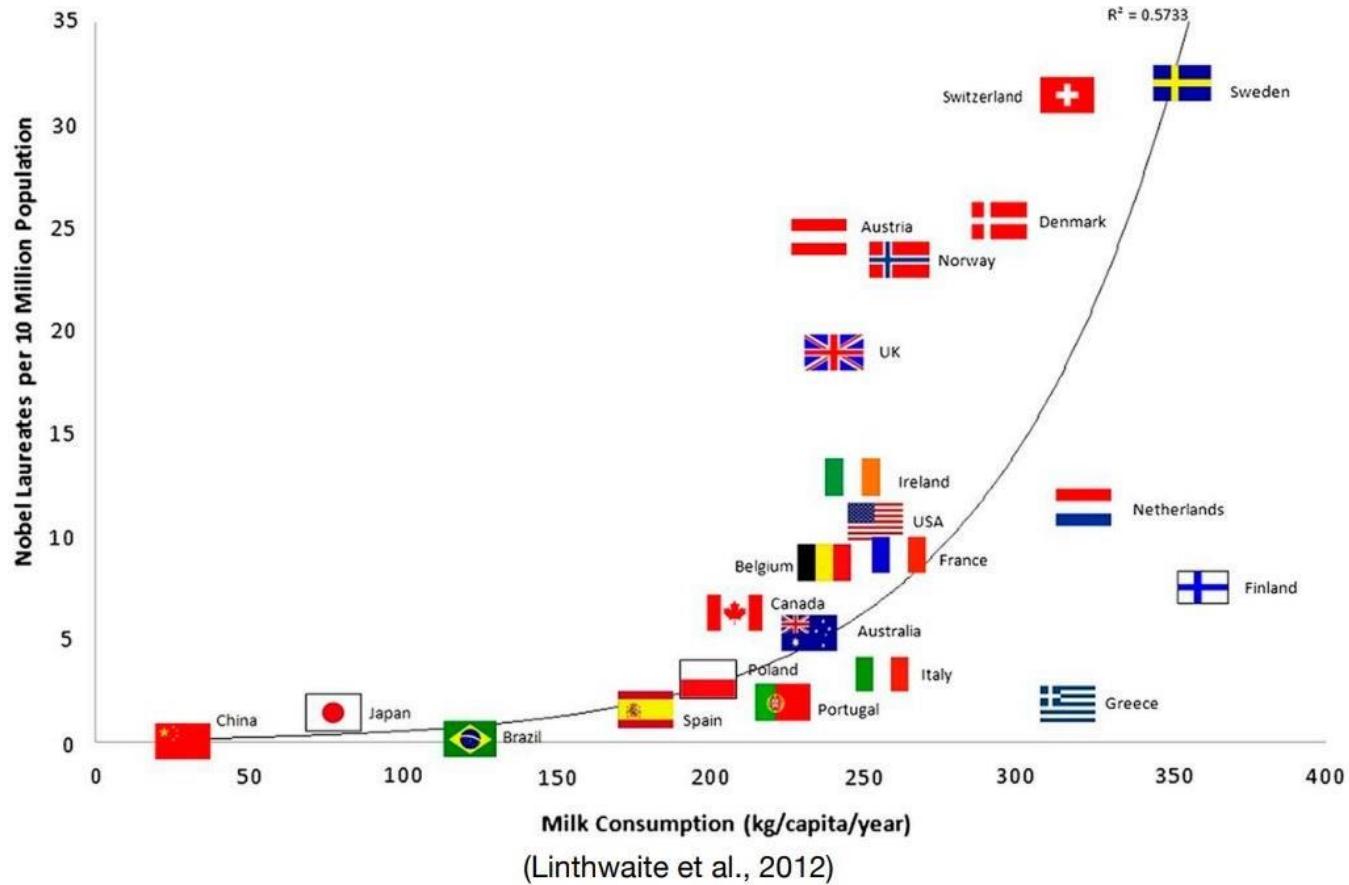
Confounding

Reverse causality



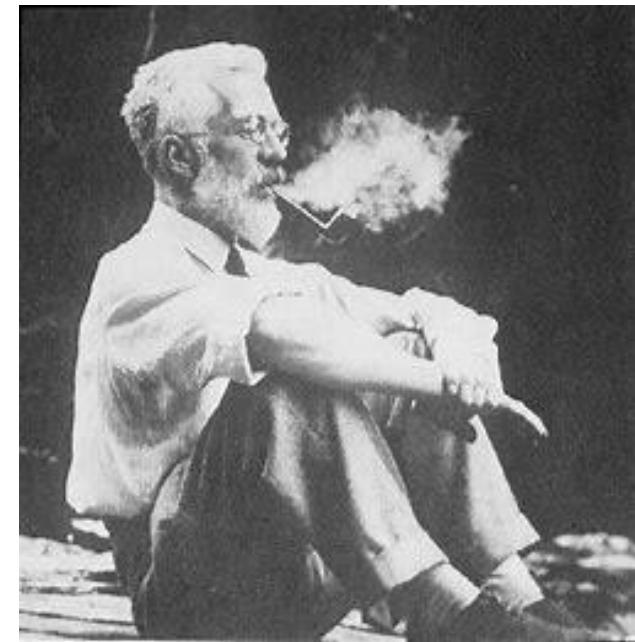
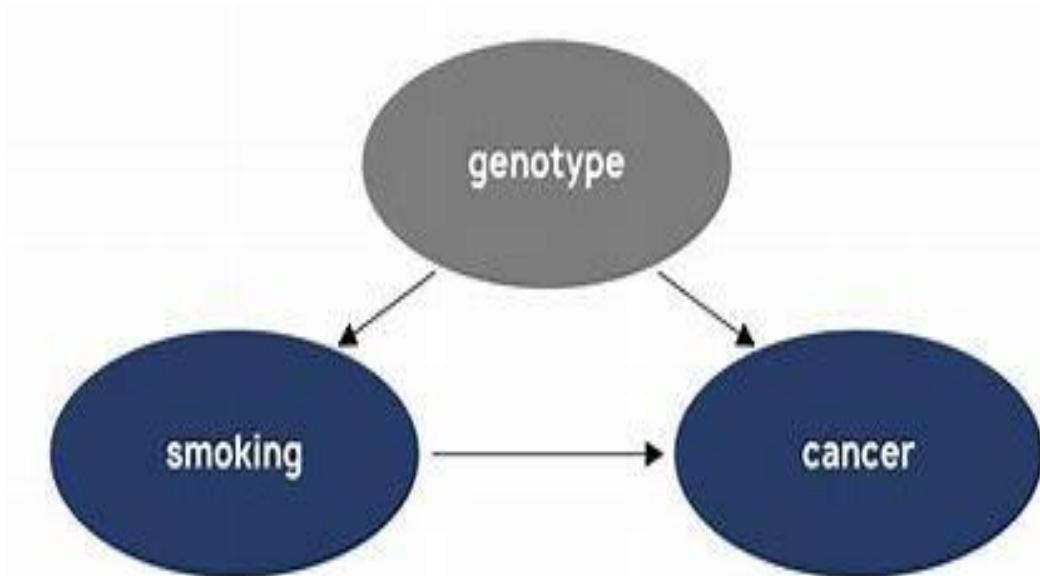
Bias

Example: confounding



Example: confounding

- Fisher's debate

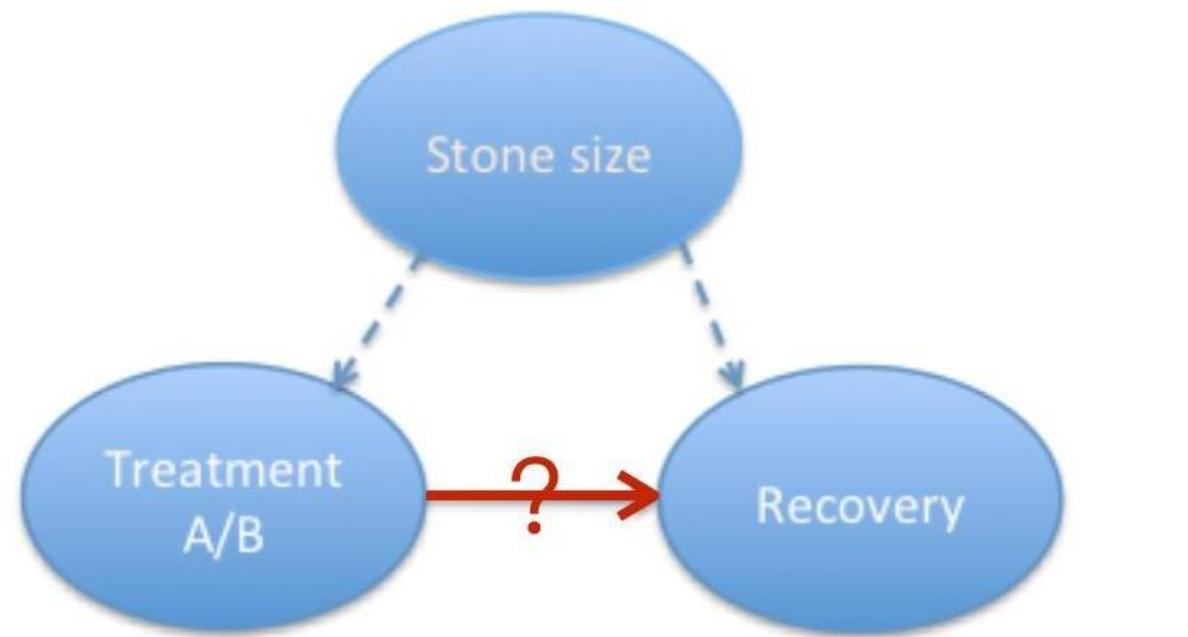


Sir R. A. Fisher

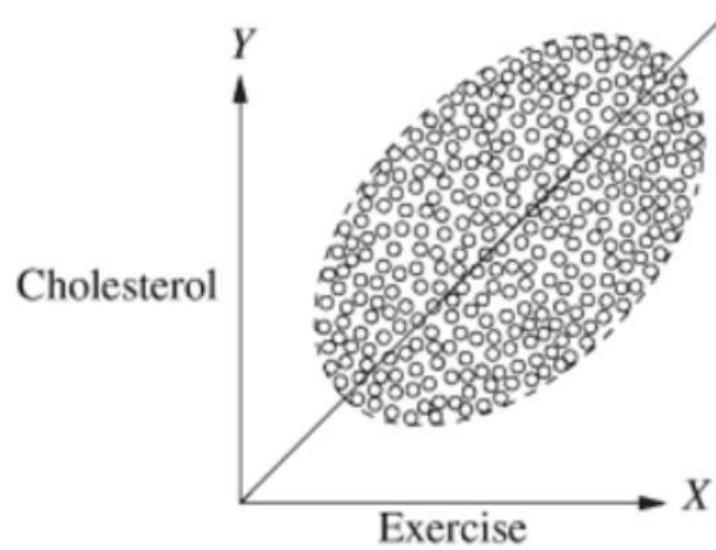
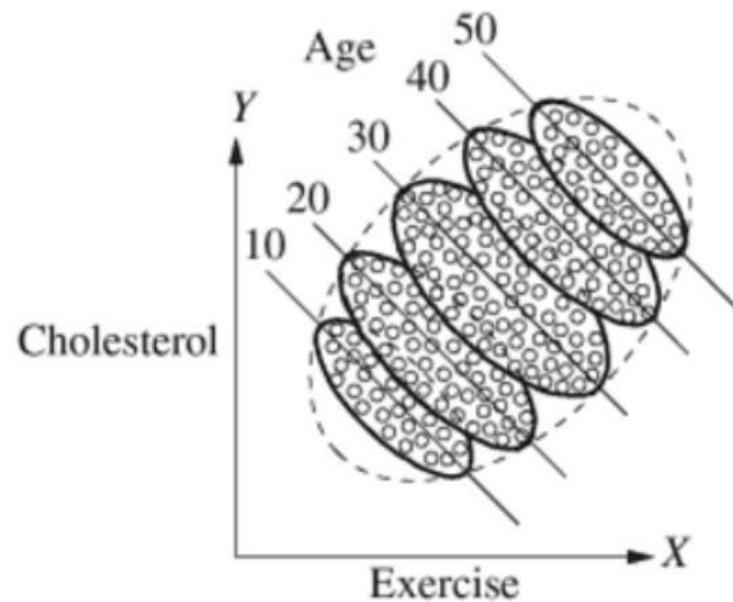
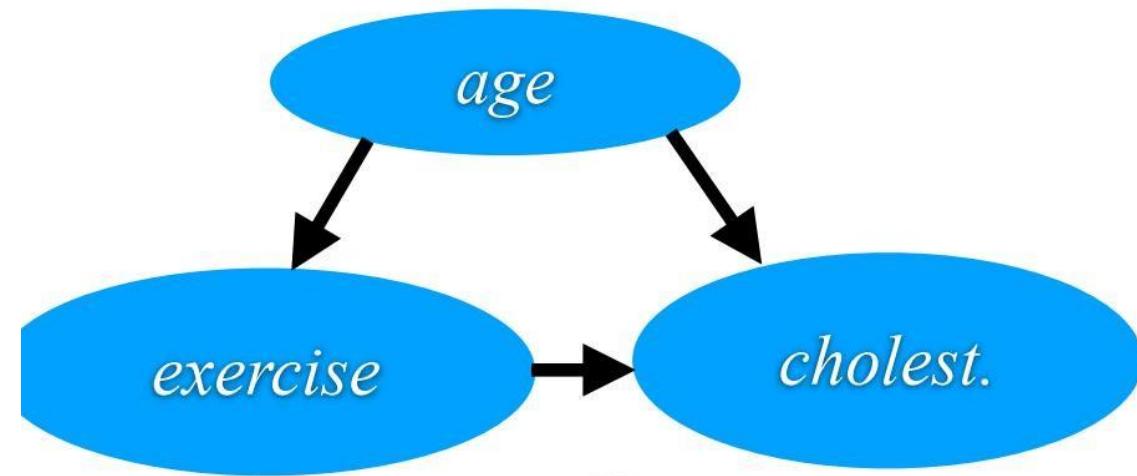
Examples: confounding

- Simpson's paradox

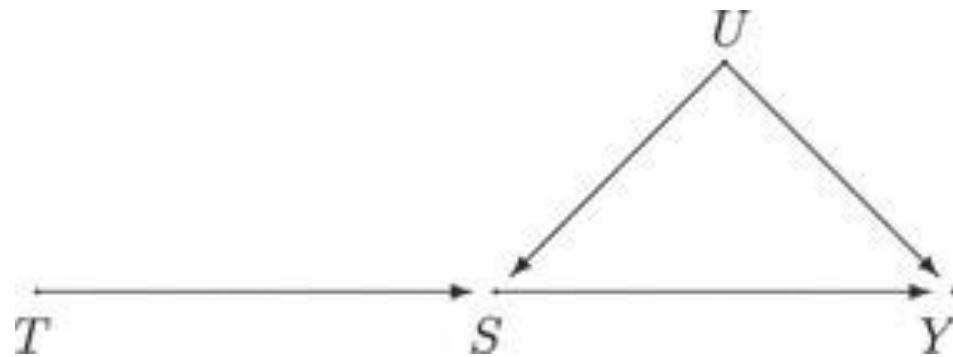
	Treatment A	Treatment B
Small Stones	<i>Group 1</i> 93% (81/87)	<i>Group 2</i> 87% (234/270)
Large Stones	<i>Group 3</i> 73% (192/263)	<i>Group 4</i> 69% (55/80)
Both	78% (273/350)	83% (289/350)



Example: confounding



Example: surrogate paradox



Not a pure mathematical game!

心律失常是猝死的危险因素，纠正心律失常能够预防猝死。FDA批准的一种药物有效地纠正了心律失常，却反而增加了猝死率



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

U	$p(S=1 u, t)$		$p(Y=1 u, s)$	
	T = 0	T=1	S=0	S=1
0	0.98	0.79	0.00	0.98
1	0.02	0.99	0.98	0.99

From the probabilities, we obtain two positive effects, $ACE\{T \rightarrow S | do(T=1), do(T=0)\}=0.6220$ and $ACE\{S \rightarrow Y | do(S=1), do(S=0)\}=0.3010$, but a negative effect $ACE\{T \rightarrow Y | do(T=1), do(T=0)\}=-0.0491$.

Chen, H., Geng, Z. and Jia, J. (2007). Criteria for surrogate end points. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*



Example: reverse causality

The Telegraph

Home News World Sport Finance Comment Culture Travel Life Women Fashion Luxe

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

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

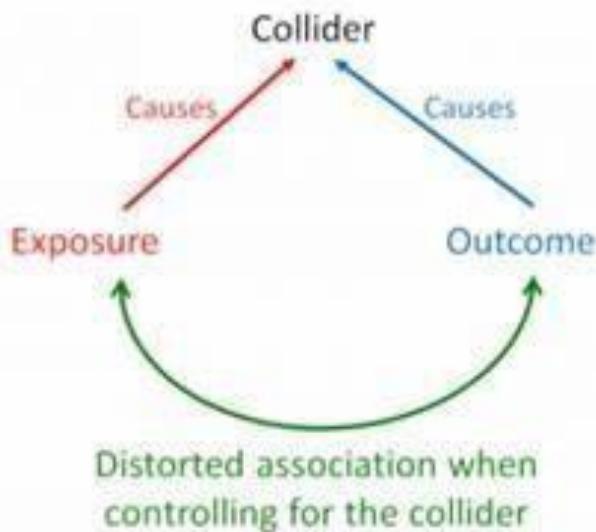
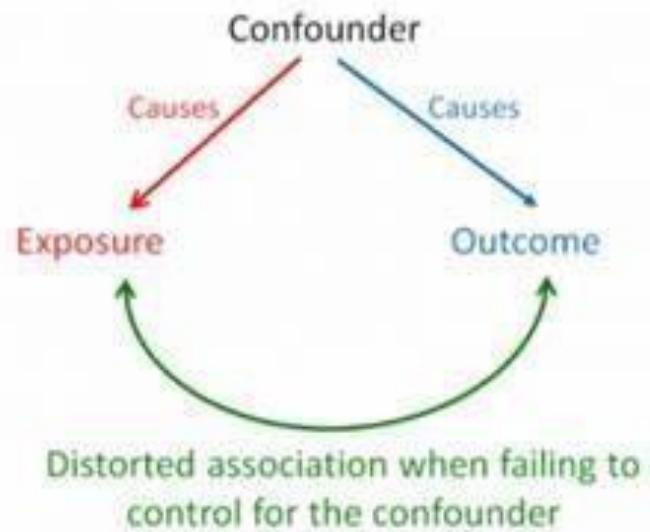
HOME » NEWS » WORLD NEWS » EUROPE

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

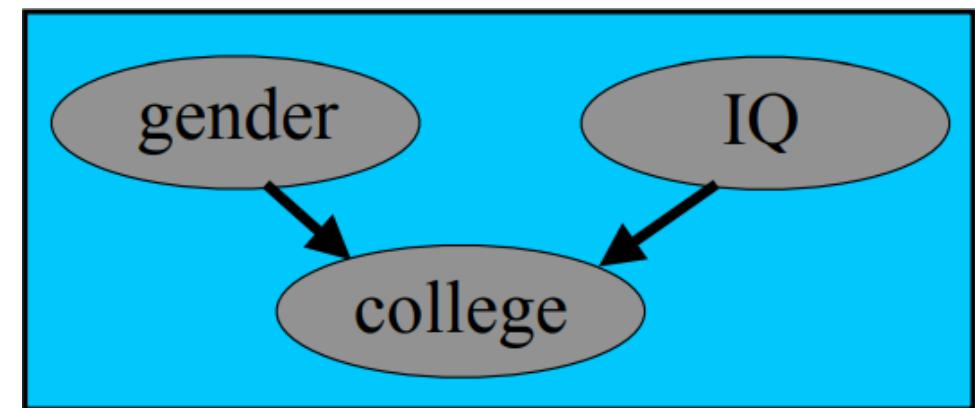
Divorce rates are far higher among “modern” couples who share the housework than in those where the woman does the lion’s share of the chores, a Norwegian study has found.



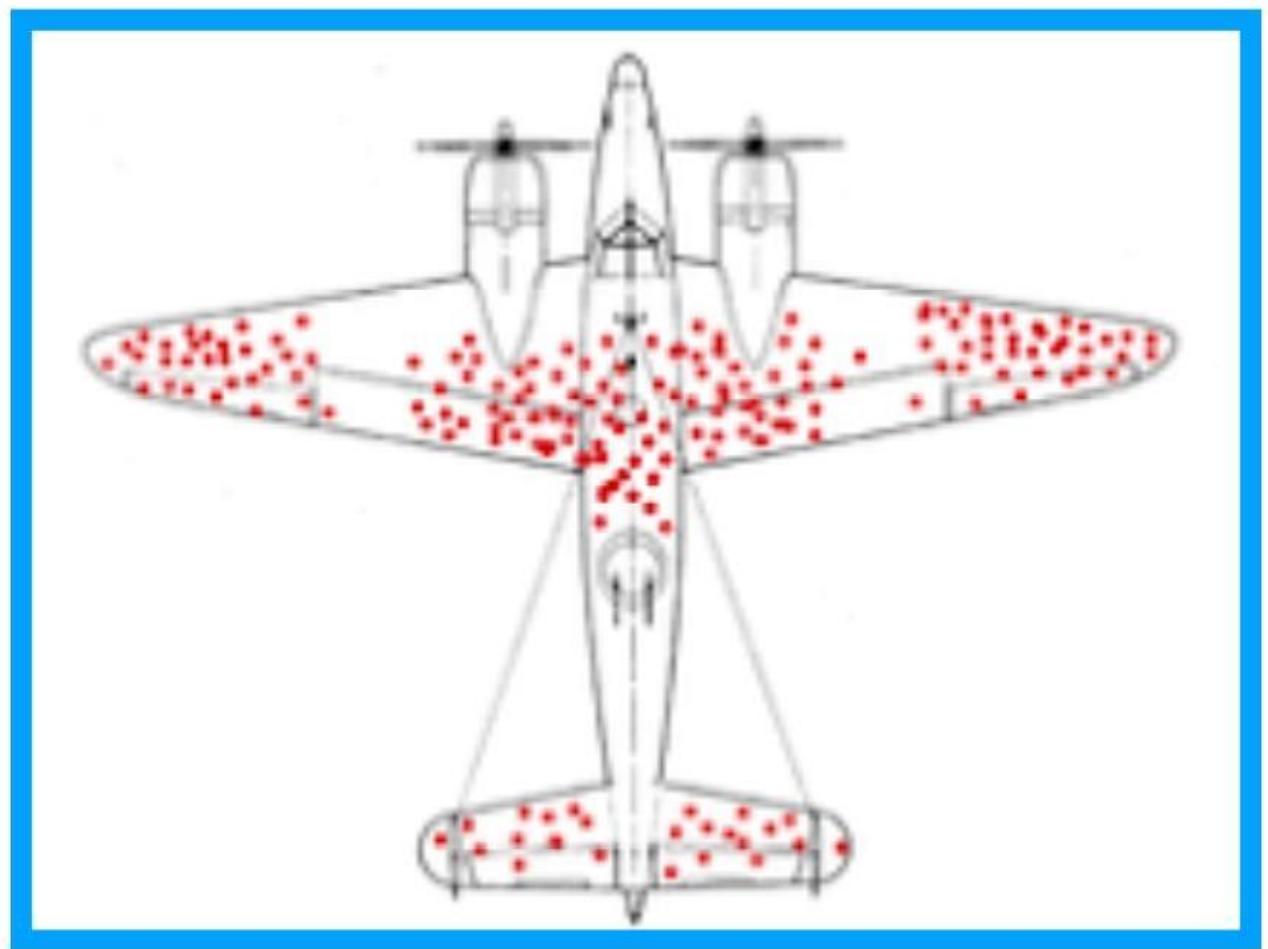
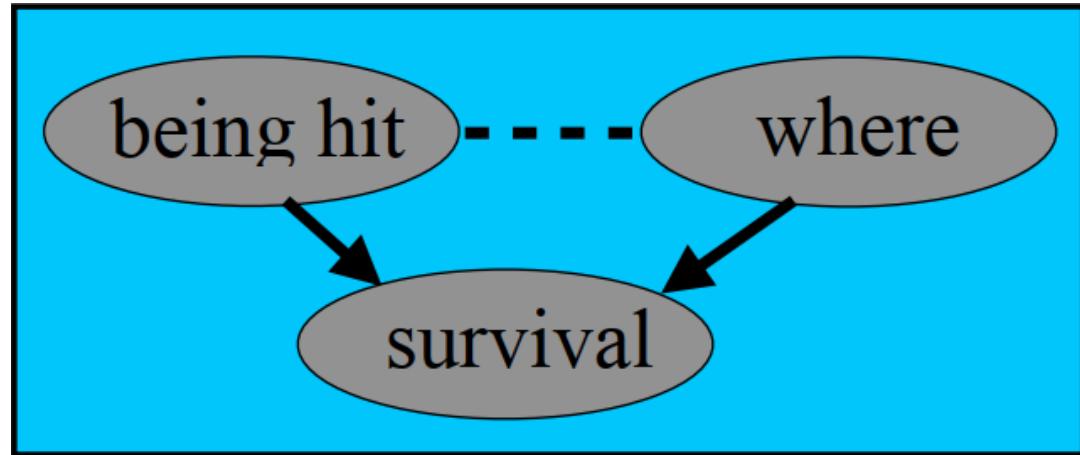
Example: Collider bias



Go back 50 years; in Western world, female college students were smarter than male ones on average. Why?

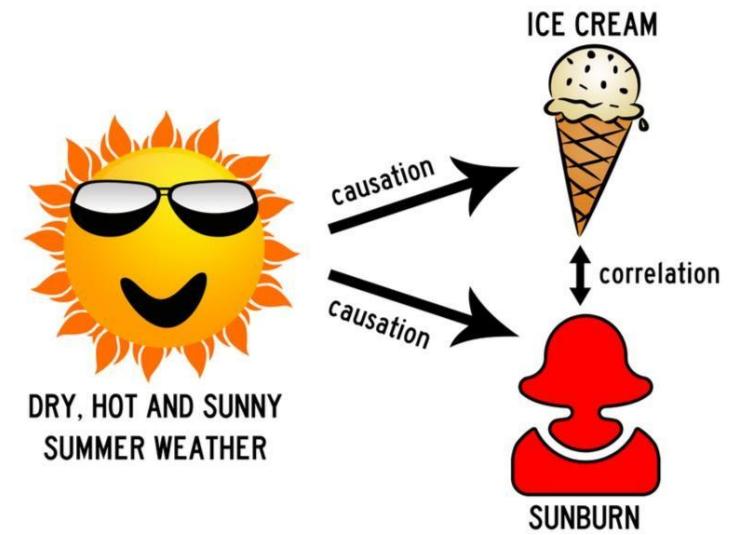
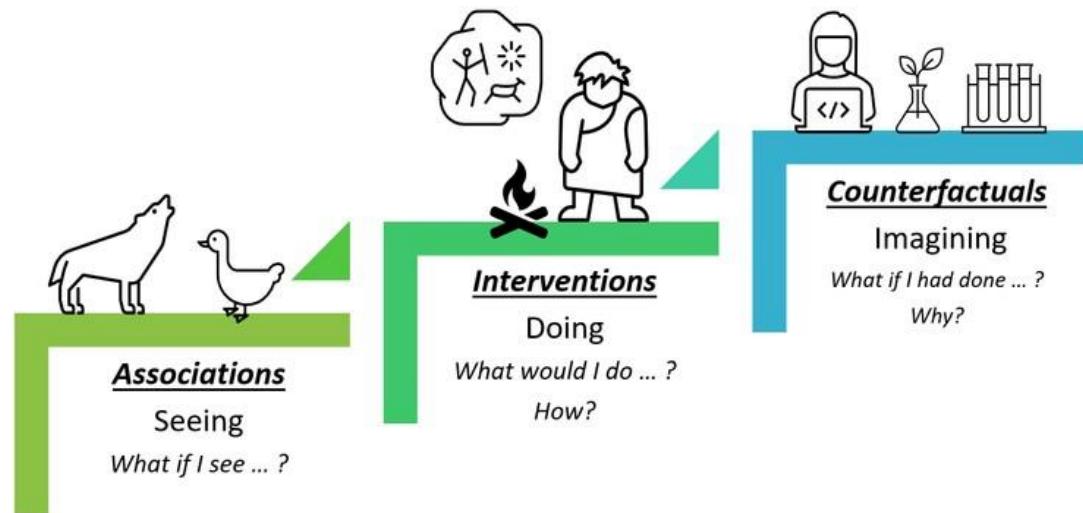


Example: Collider bias



Why causality so important

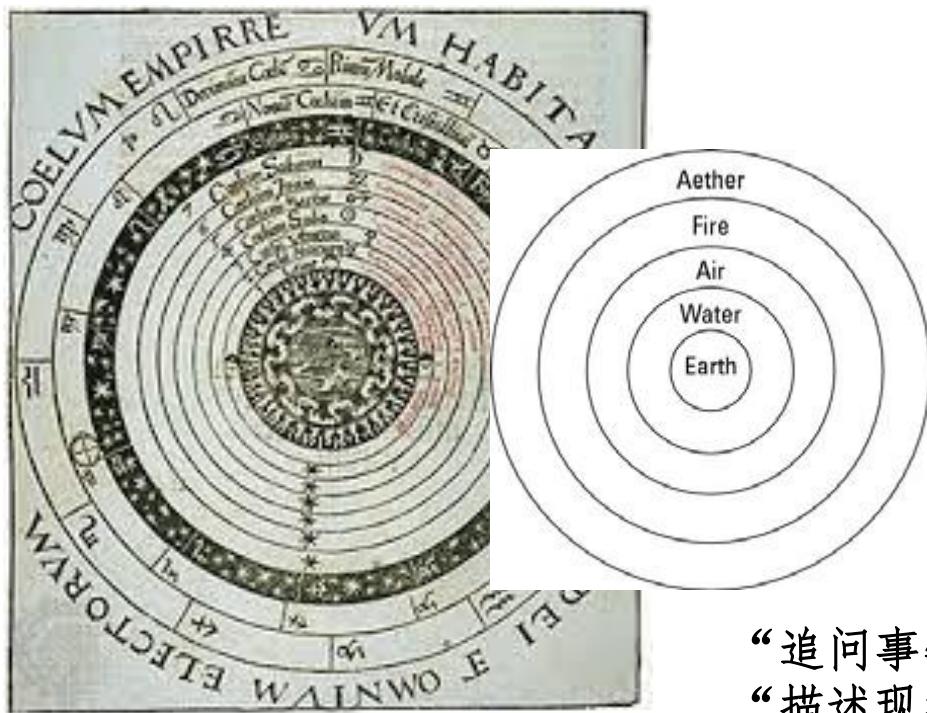
- Correlation → Prediction;
Causation → Decision



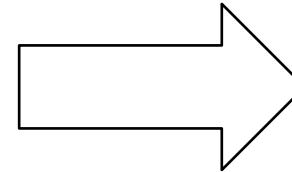
A brief history on causality studies

Hellenistic era

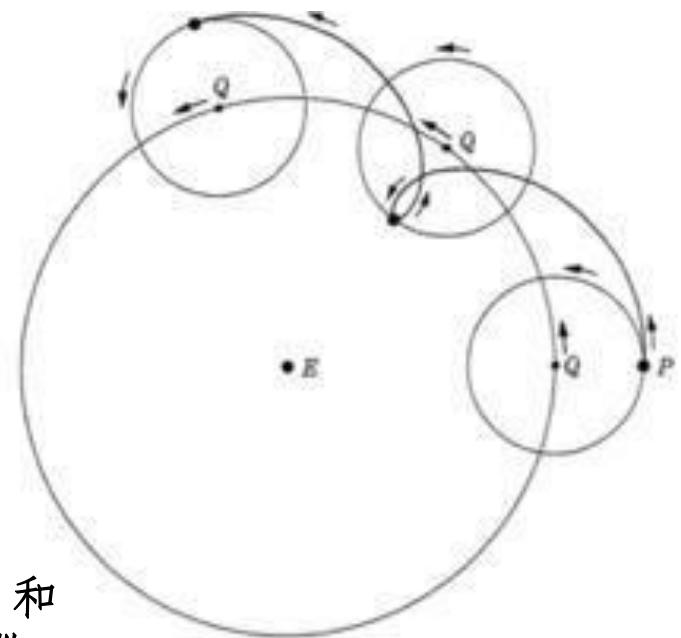
- Aristotle physics (*cosmos*)



Saving the phenomena



- Ptolemaic astronomy



“追问事物原因/解释”【物理学】和
“描述现象规律并作出预测”【(数理)天文学】分属于相对独立的两个
不同学科



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



De revolutionibus orbium, 1543

- *Neque enim necesse est, eas hypotheses esse veras, imo ne verisimiles quidem, sed sufficit hoc unum, si calculum observationibus congruentem exhibeant.*
- **It is indeed not necessary that these hypotheses are real or very deeply believable, but if these hypotheses would give a calculation method which is consistent to the observations, this one is sufficient.**

出自出版商加的序言，不完全是哥白尼本意

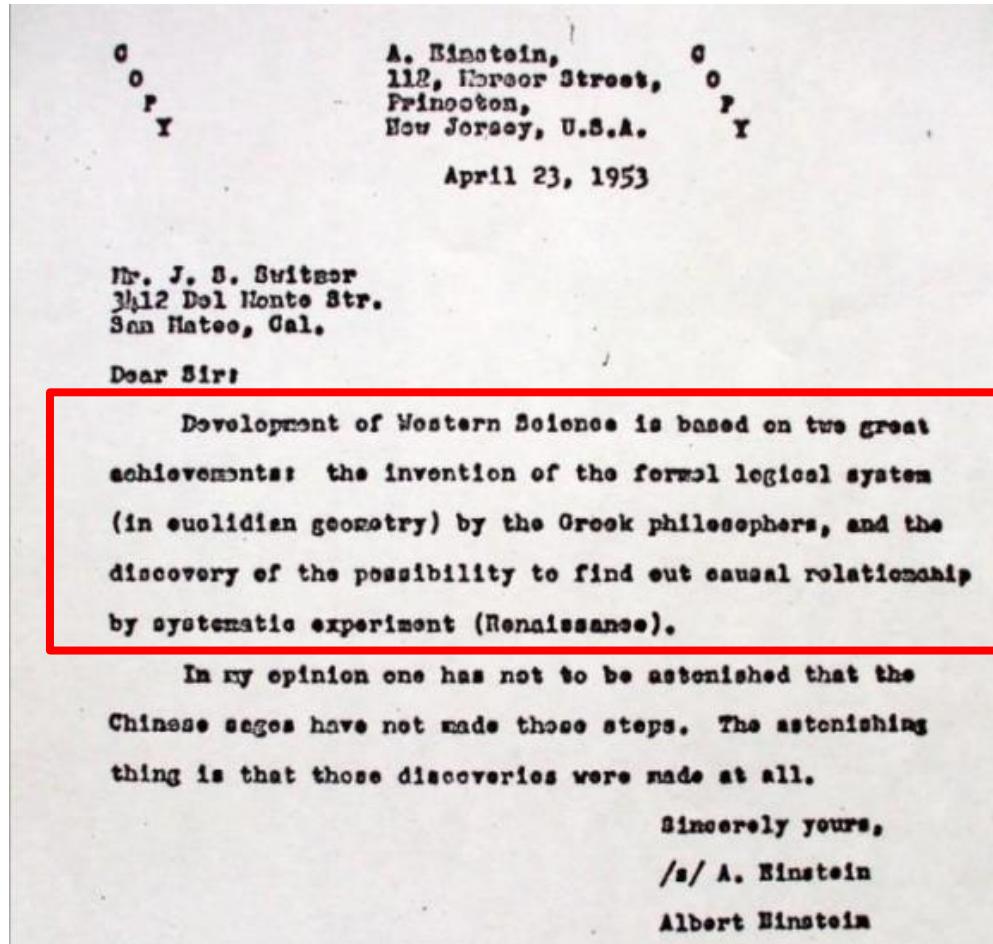
以亚里士多德物理学的原则反对托勒密天文学，又用托勒密天文学的方法摧毁了亚里士多德物理学



ON dubito, quin erudití quidam, uulgata iam de nouitate hypotheseson huius operis fama, quod ter ram mobilem, Solem uero in medio uniuersi immobilem constituit, uehementer sinit offensi, p̄tētq; disciplinas liberales recte iam olim constitutas, turbari nō oportere. Verum si rem exācte perpendere volent, inueniēt auctorem huius operis, nihil quod reprehendi mereatur cōmis̄sse. Est enim Astronomi proprium, historiam motuum cœlestium diligenti & artificiosa obseruatione colligere. Deinde causas earundem, seu hypotheses, cum ueras aſſequi nulla ratione possit, qualescunq; excogitare & configere, quibus suppositis, īdem motus, ex Geometriæ principijs, tam in futurū, quam in præteritū recte possint calculari. Horū autē utruncq; egregie præstítit hic artifex. Nec enim necesse est, eas hypotheses esse ueras, imo ne uerisimiles quidem, sed sufficit hoc unum, si calculum obseruationibus congruentem exhibeant. ni



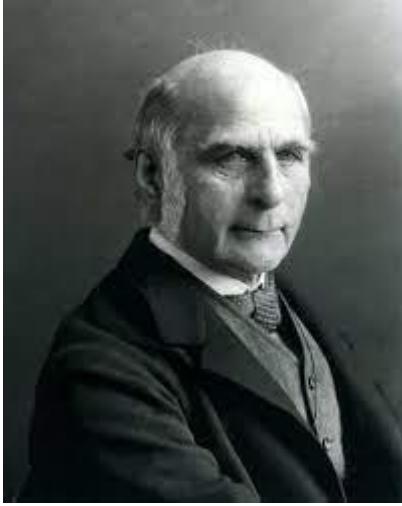
Renaissance & scientific revolution



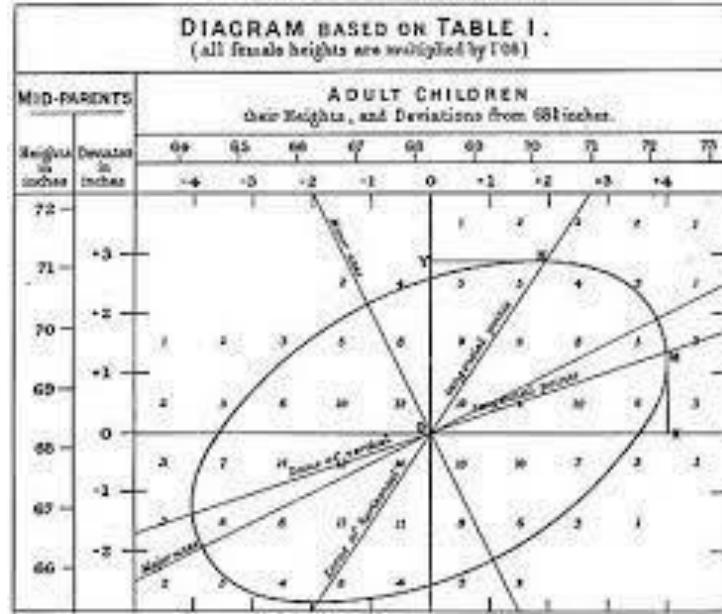
西方科学的发展是以两个伟大的成就为基础，那就是：希腊哲学家发明形式逻辑体系（在欧几里得几何学中），以及通过系统的实验发现有可能找出因果关系（在文艺复兴时期）。



Era of correlation



Sir Francis Galton



Regression to mean

“作为运动原因的力，与作为成长原因的树神完全一样。因果关系是现代科学高深奥秘中的迷信”

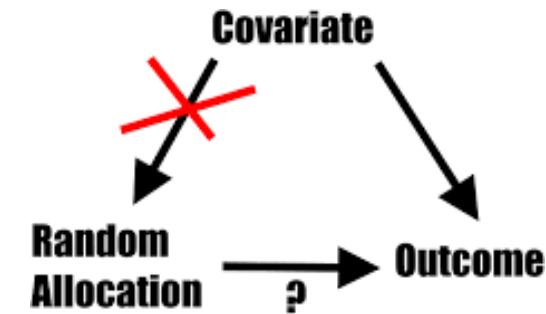


Karl Pearson

Episode: Debate between Mendelianist (molecular genetics)
and Biometrician (quantitative genetics)

R. A. Fisher: randomization

- Lady tasting tea



- “a genius who almost single-handedly created the foundations for modern statistical science”
- one to most comprehensively combine the ideas of Gregor Mendel and Charles Darwin.

Sir R. A. Fisher

Causal revolution



Statistics



*Computer
science*



Epidemiology



Economics

Causal revolution



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

因果推断研究获2021诺贝尔经济学奖，图灵奖得主Judea Pearl祝贺并反对

Judea Pearl ✅
@yudapearl ...

1/ Congratulations are due to our colleagues Joshua Angrist and Guido Imbens on receiving the 2020 Nobel Prize in Economics, thus drawing the limelight to the science of causal inference and to the new methodology which they have helped develop.

It is no secret that I have been

2/ (and still am) a staunch opponent of Angrist and Imben's methodology, primarily for its overlooking the two Fundamental Laws of causal inference.

Nevertheless, their method has caused a shakeup in economics and a greater appreciation for the general problem of drawing causal

翻译推文

量子位

量子位



Two (modern) causal models

Potential outcome/Counterfactuals

- Independently invented by Jerzy Neyman (1921) and Donald Rubin (1976); now most commonly used in statistics and econometrics
- Notation:
 - Y – outcome;
 - A – treatment / exposure / intervention / policy / assignment
- In observed data, we have the following table

	是否用药	是否痊愈
id	A	Y
1	0	0
2	0	0
3	1	0
4	1	1
5	0	1
...

Potential outcome/Counterfactuals

- How can we quantify the causal effect of A on Y?

id	A	Y	A	Y
1	0	0	1	1
2	0	0	1	1
3	1	0	0	0
4	1	1	0	0
5	0	1	1	0
...		

Red columns: hypothetical,
not really observed in data



Will I be healthy if I
do take vitamin C?
 C_1

Will I be healthy if I
don't take vitamin C?
 C_0

Potential outcome/Counterfactuals

- How can we quantify the causal effect of A on Y?

id	A	Y	A	Y
1	0	0	1	1
2	0	0	1	1
3	1	0	0	0
4	1	1	0	0
5	0	1	1	0
...

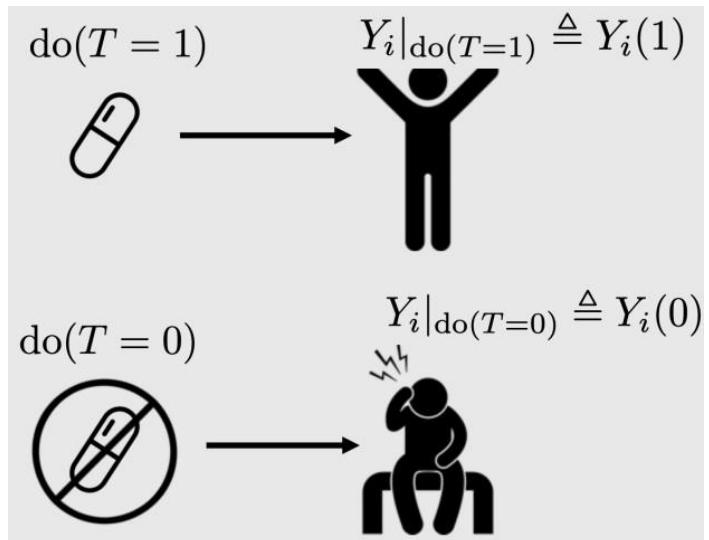
Potential outcome



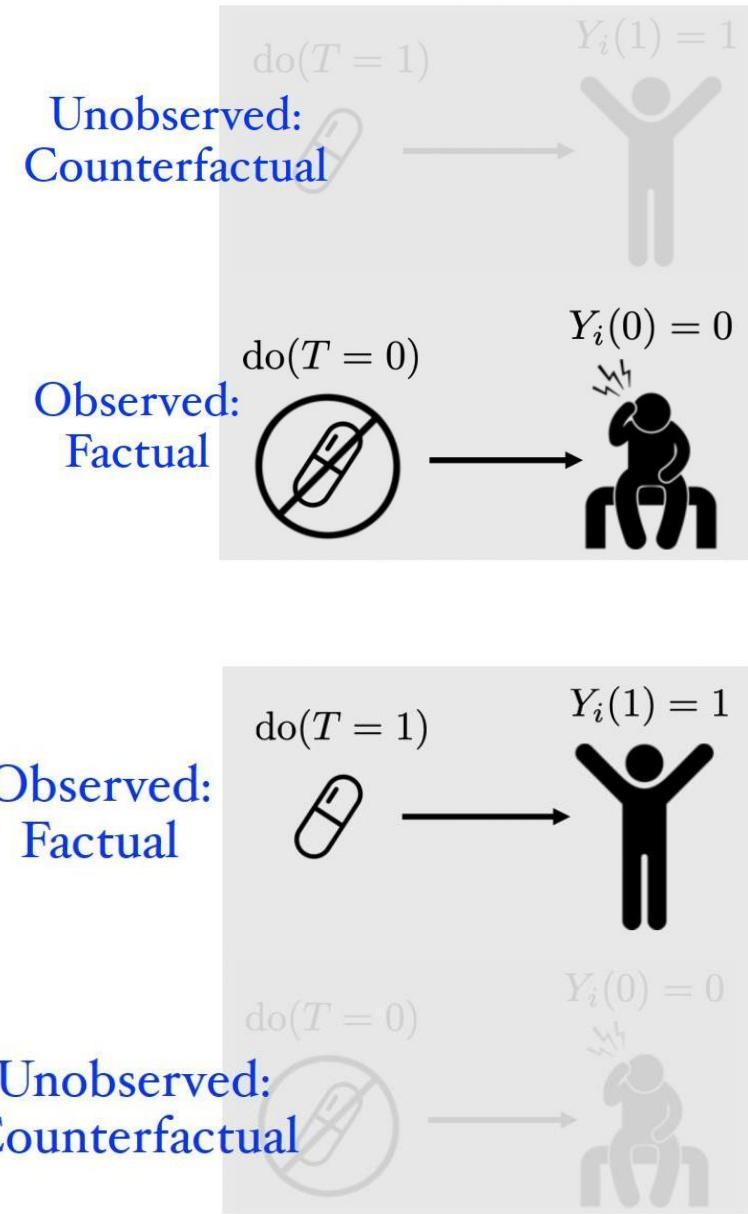
id	A	Y(1)	Y(0)
1	0	1	0
2	0	1	0
3	1	0	0
4	1	1	0
5	0	0	1
...

$Y(a)$: the outcome were A assigned value a
(subjunctive mood)

Missing data issue



Causal effect:
 $Y_i(1) - Y_i(0)$

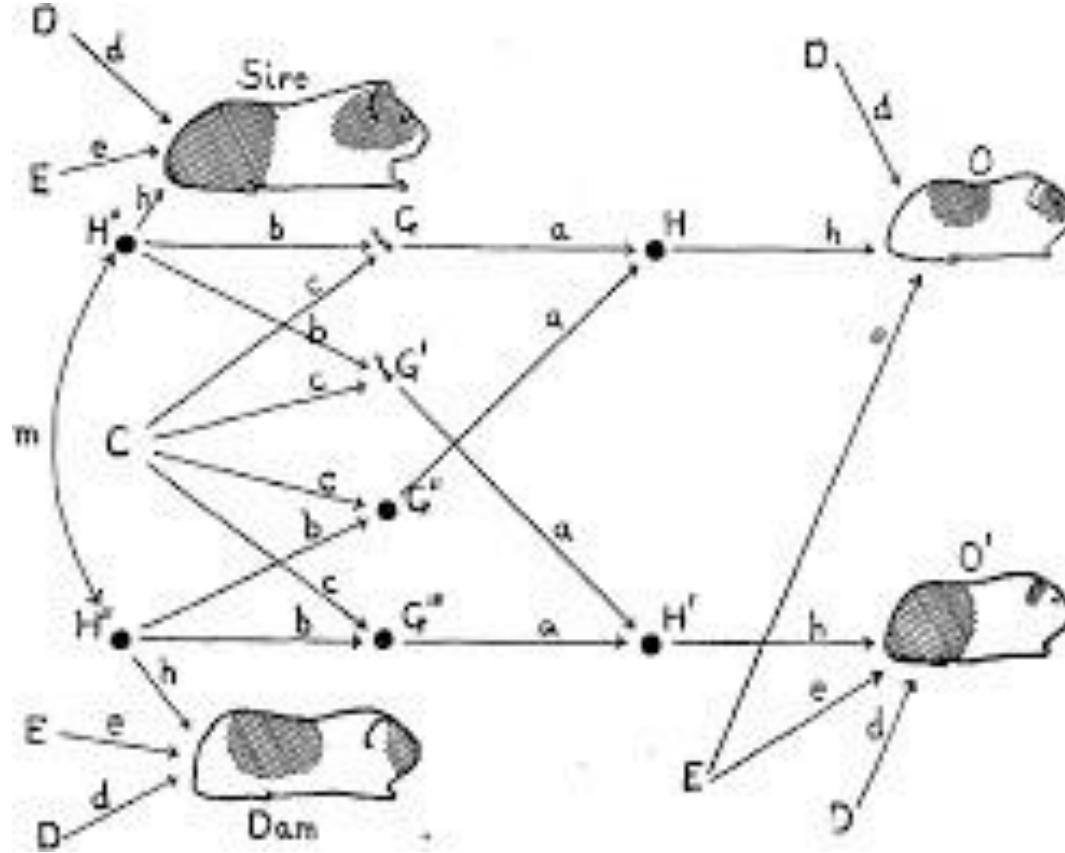


Average treatment
effect (ATE)
 $\tau := \mathbb{E}\{Y(1) - Y(0)\}$

Wright's path analysis

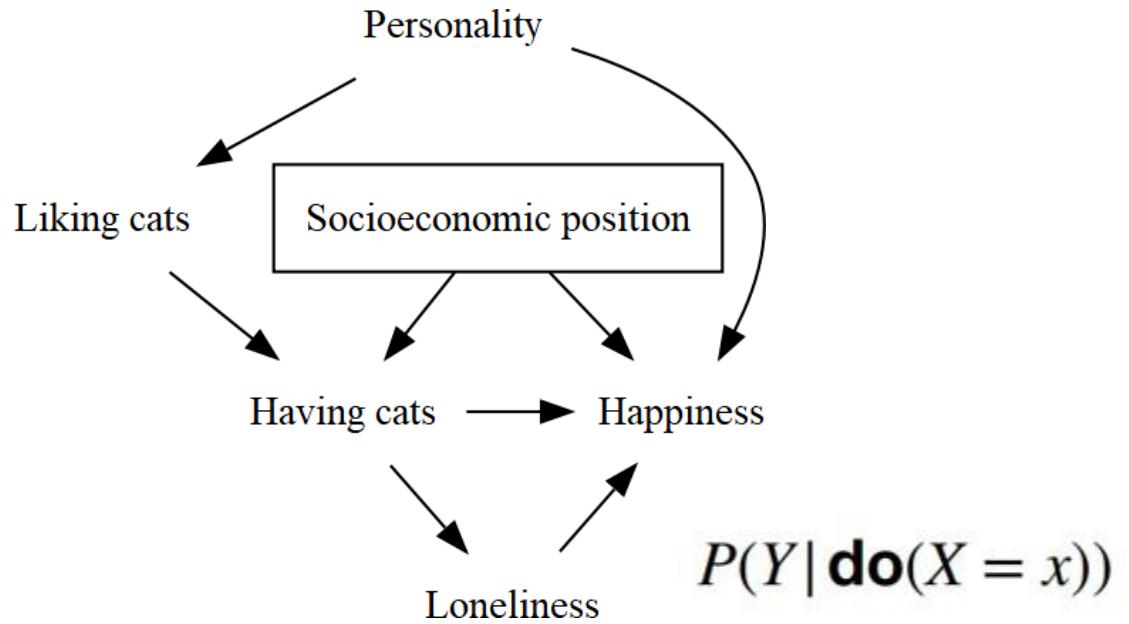


Sewall Wright



Structural causal model

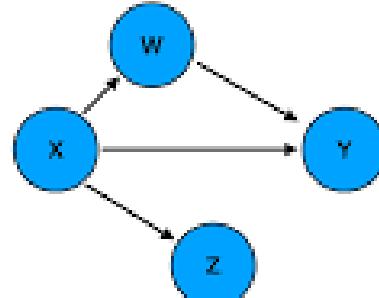
- Causal graph: DAG



- Structural equation model

$$X_i = f_i(PA_i, E_i), i = 1, \dots, n$$

Directed Acyclic Graphs (DAGs)



Structural Equation Models (SEMs)

$$W := f_1(X)$$

$$Z := f_2(X)$$

$$Y := f_3(X, W)$$

Three domains



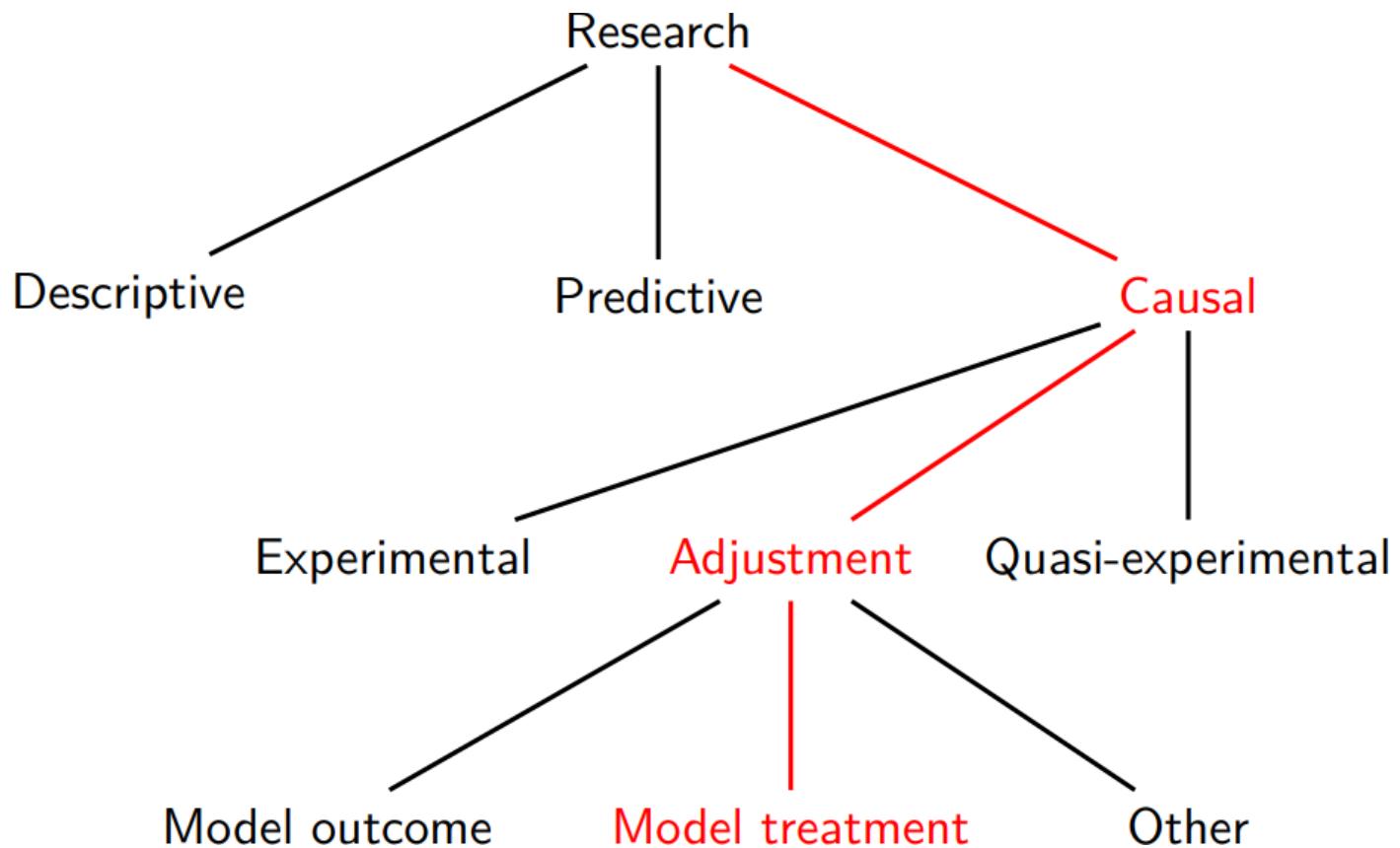
上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



Three domains

- **Causal inference (causal effect estimation) – known variables, known graph**
 - Experimental: RCT or A/B test
 - Observed confounding: adjustment
 - Unobserved confounding: quasi-experimental
- **Causal discovery (causal structure learning) – known variables, unknown graph**
 - BN structure learning: Constraint-based & score-based
 - Functional Causal Models
- **Causal representation learning – unknown variables, unknown graph**

Causal effect estimation



Two perspectives

- In Pearl's framework, the investigator starts from an assumption that the data was generated by a particular directed acyclic graph, and then proves that if he is able to block all backdoor paths between the exposure and the outcome by conditioning on certain covariates, then controlling for those variables is sufficient to eliminate confounding.
- In Rubin's framework, the investigator starts from the assumption that treatment is assigned by a "treatment assignment mechanism", and then proves that if conditional on certain covariates this mechanism does not depend on the counterfactual outcome , then controlling for those covariates is sufficient to eliminate confounding.



Assumptions

- Ignorability: $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp T_i$

Potential outcome $\perp\!\!\!\perp$ assignment mechanism

Conditional ignorability: $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp T_i | X_i$

- Positivity: $0 < P(T = 1 | X = x) < 1$

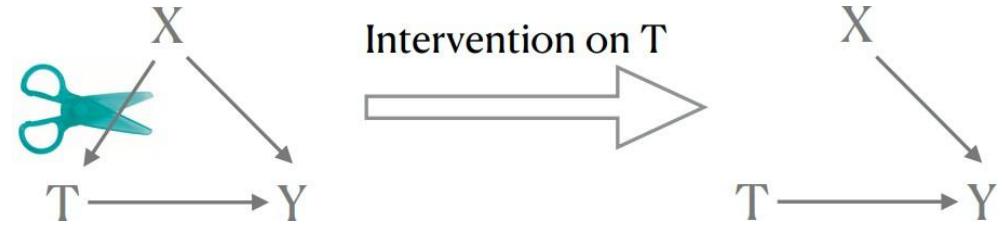
- No interference: $Y_i(t_1, \dots, t_{i-1}, t_i, t_{i+1}, \dots, t_n) = Y_i(t_i)$

} Stable Treatment Value
Assumption (SUTVA)

- Consistency: $T = t \implies Y = Y(t)$



RCT or A/B test



- average causal effect (ACE) / average treatment effect (ATE):

$$\tau := \mathbb{E}\{Y(1) - Y(0)\}$$

- g-formula

$$\tau = \mathbb{E}[Y(1) - Y(0)]$$

$= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$ by linearity of expectation

$= \mathbb{E}[Y(1)|A = 1] - \mathbb{E}[Y(0)|A = 0]$ by randomization $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp A_i$

$= \mathbb{E}[Y|A = 1] - \mathbb{E}[Y|A = 0]$ by consistency

: mean difference between each treatment group

Confounding

$$CATE = E[Y(1) - Y(0) | X]$$



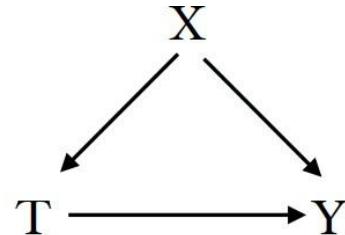
$$= E[Y(1) | X] - E[Y(0) | X]$$

$$= E[Y(1) | T = 1, X] - E[Y(0) | T = 0, X] \quad (\text{Conditional ignorability})$$

$$= [E[Y | T = 1, X] - E[Y | T = 0, X]] \quad (\text{Consistency})$$

Only contains observable moments

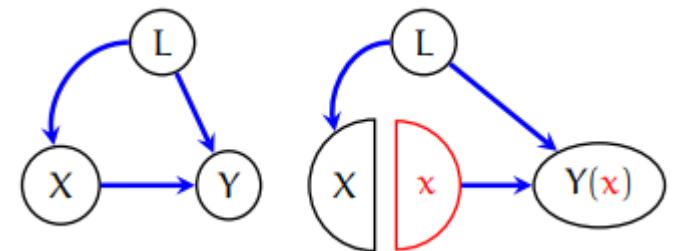
$$ATE = E[Y_i(1) - Y_i(0)]$$



$$= E_X E[Y_i(1) - Y_i(0) | X_i]$$

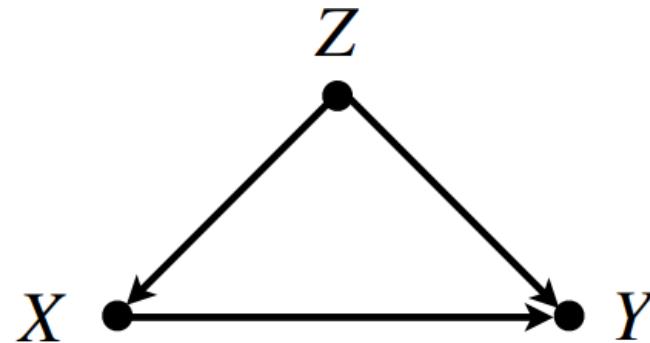
$$= E_X [E[Y_i | T_i = 1, X_i] - E[Y_i | T_i = 0, X_i]]$$

Single World Intervention
Graphs (SWIGs)

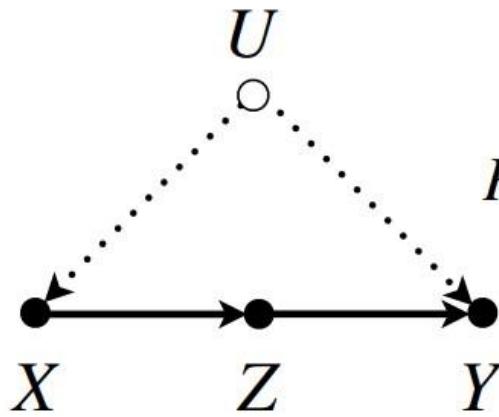


Graph perspective

- Back door
- Front door



$$P(Y \mid \text{do}(X = x)) = \sum_z P(Y \mid X = x, Z = z)P(Z = z)$$



$$P(Y \mid \text{do}(X = x)) = \sum_z P(z \mid x) \sum_{x'} P(Y \mid X = x', Z = z)P(X = x')$$

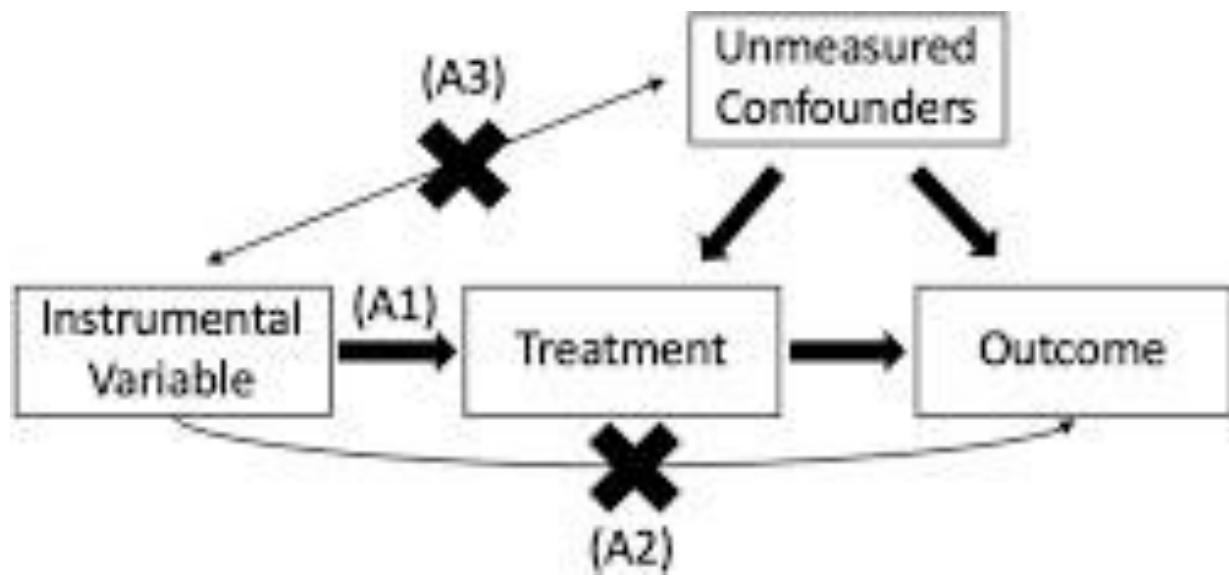


Adjustment

- Observed confounding
 - Regression adjustment
 - Matching
 - Propensity score
 - Inverse probability weighting
 - Doubly robust estimation
 - G-estimation
 - Double Machine Learning
- Unobserved confounding
 - Instrumental variable
 - Difference-in-Difference
 - Regression Discontinuity
 - Sensitivity analysis

Examples

- quasi-experimental:
Instrumental variable



The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables With Moments From Two Samples

JOSHUA D. ANGRIST and ALAN B. KRUEGER*

Economic Shocks and Civil Conflict: An Instrumental Variables Approach

Edward Miguel

University of California, Berkeley and National Bureau of Economic Research

Shanker Satyanath and Ernest Sergenti

New York University



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

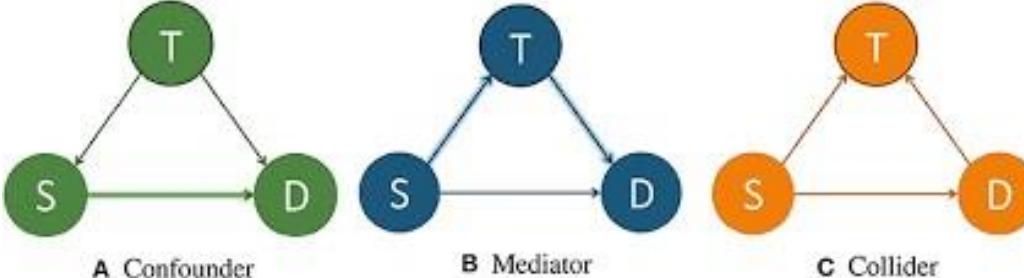


Three domains

- **Causal inference (causal effect estimation) – known variables, known graph**
 - Experimental: RCT or A/B test
 - Observed confounding: adjustment
 - Unobserved confounding: quasi-experimental
- **Causal discovery (causal structure learning) – known variables, unknown graph**
 - BN structure learning: Constraint-based & score-based
 - Functional Causal Models
- **Causal representation learning – unknown variables, unknown graph**

Causal discovery

- BN structure learning



Causal Markov condition: each variable is ind. of its non-descendants (**non-effects**) conditional on its parents (**direct causes**)

causal structure
(causal graph)
 $Y \rightarrow X \rightarrow Z$
 $Y \dashv\vdash X \dashv\vdash Z ?$

Statistical
independence(s)
 $Y \perp\!\!\!\perp Z | X$

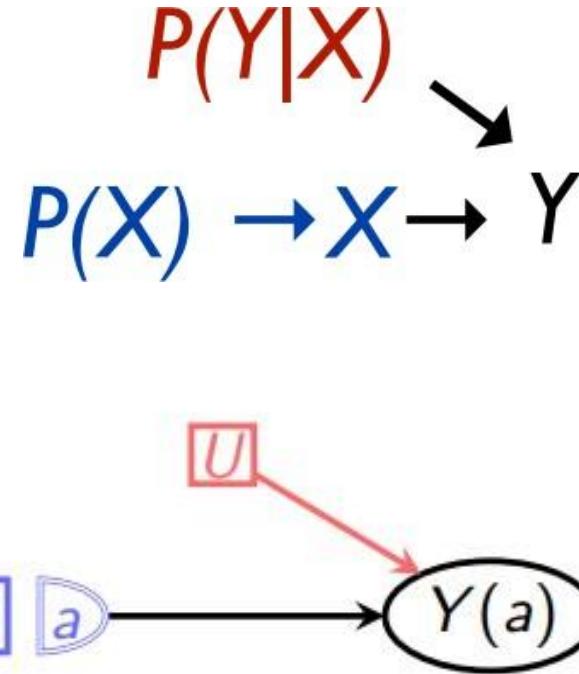
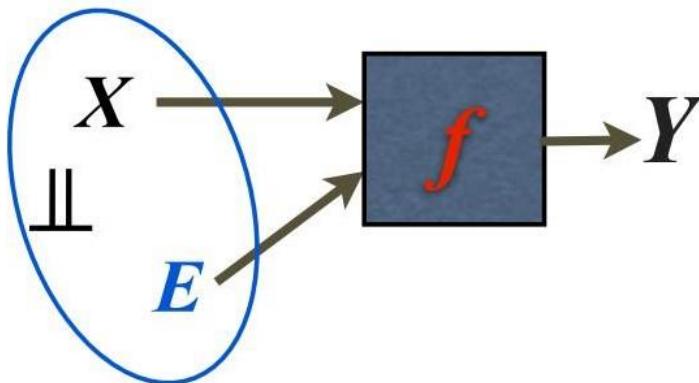
Faithfulness: all observed (conditional) independencies
are entailed by **Markov condition** in the causal graph



Causal discovery: Functional Causal Models

- independent noise assumption

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



5

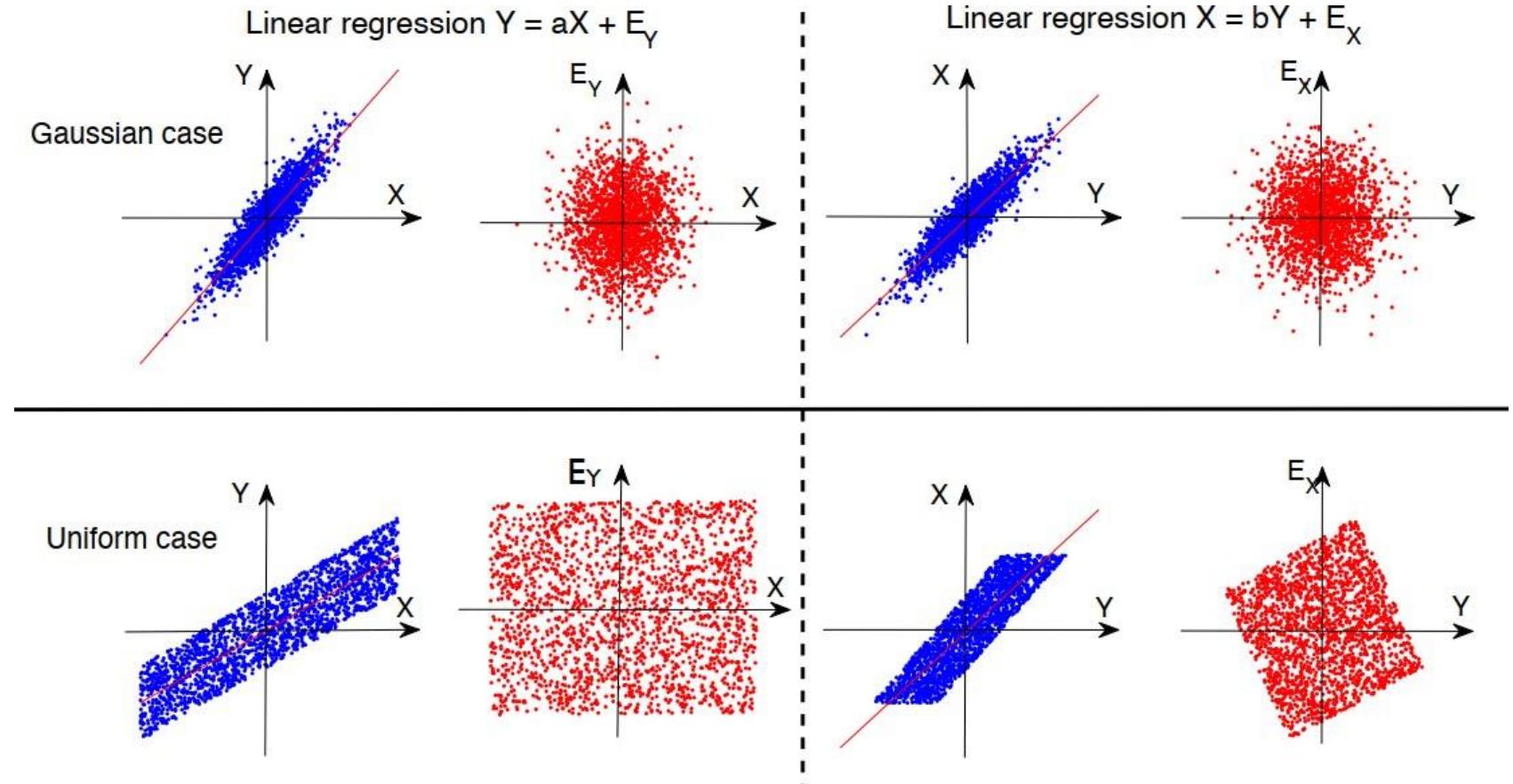
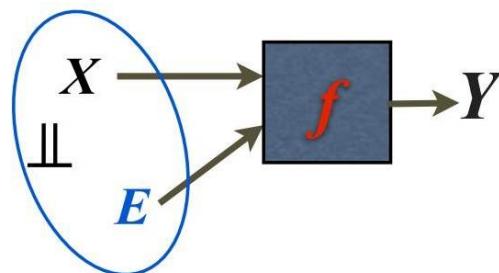
5



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



Asymmetry between cause and effect



E_Y 与 Y 独立

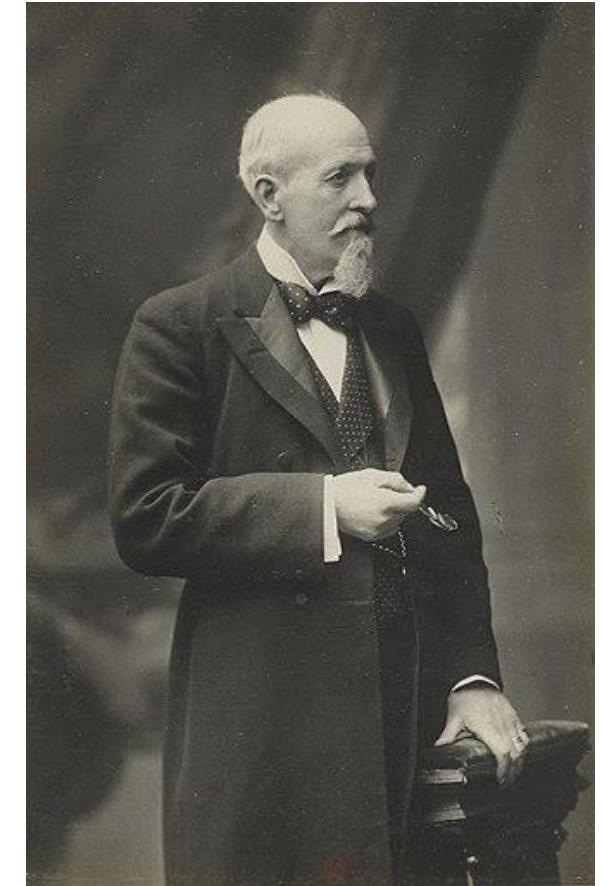
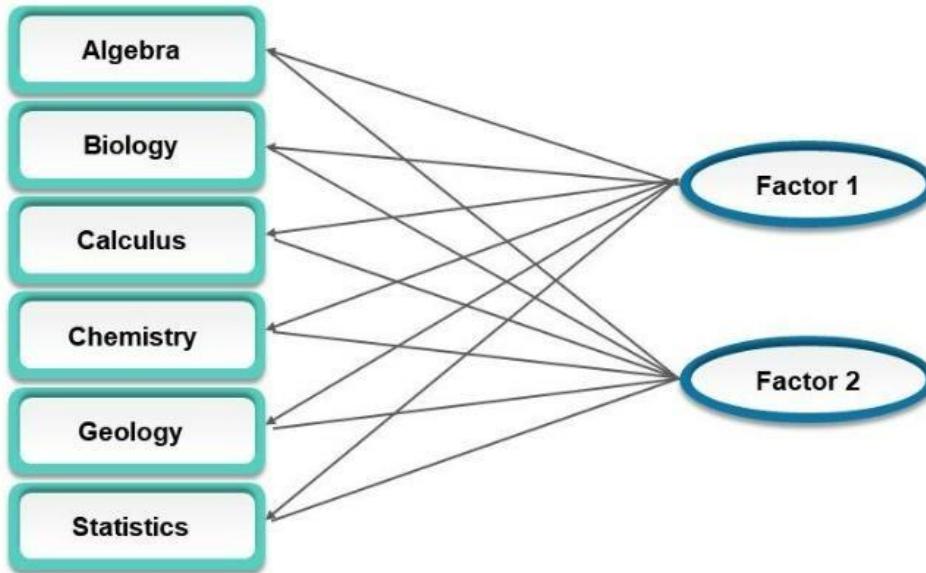
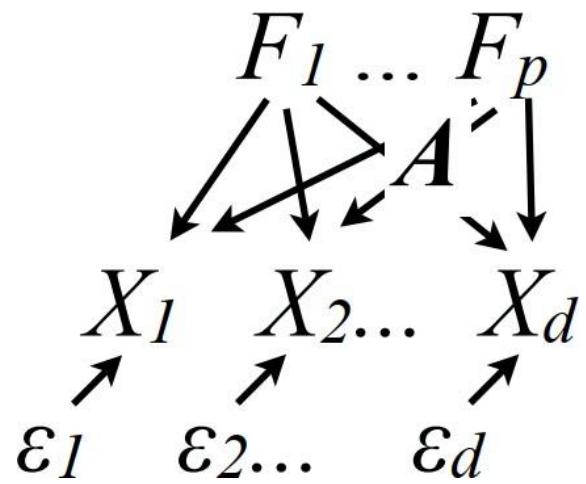
E_Y 与 Y 不相关
(未必独立)



Three domains

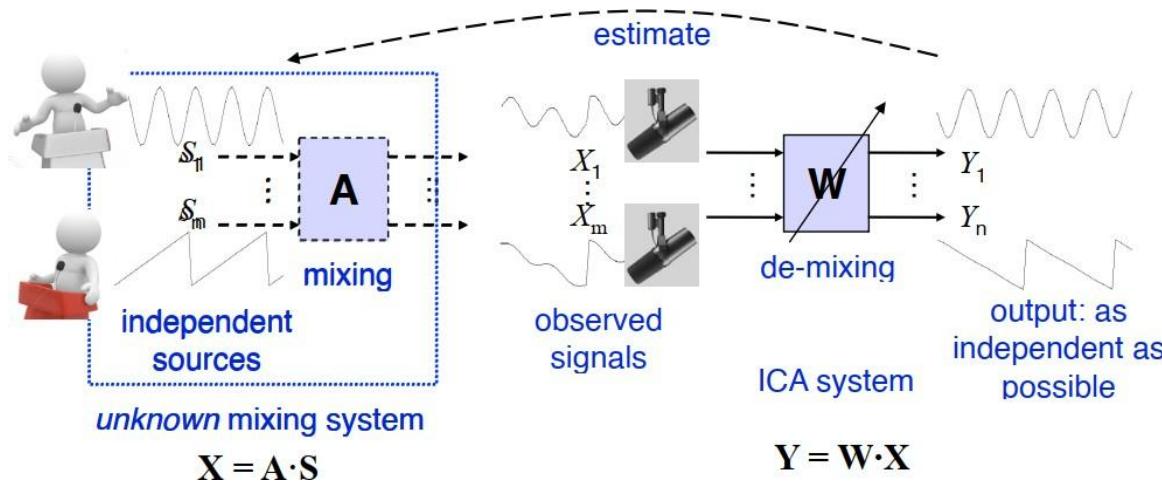
- **Causal inference (causal effect estimation) – known variables, known graph**
 - Experimental: RCT or A/B test
 - Observed confounding: adjustment
 - Unobserved confounding: quasi-experimental
- **Causal discovery (causal structure learning) – known variables, unknown graph**
 - BN structure learning: Constraint-based & score-based
 - Functional Causal Models
- **Causal representation learning – unknown variables, unknown graph**

Factor Analysis



Independent causal mechanism

- Independent component analysis

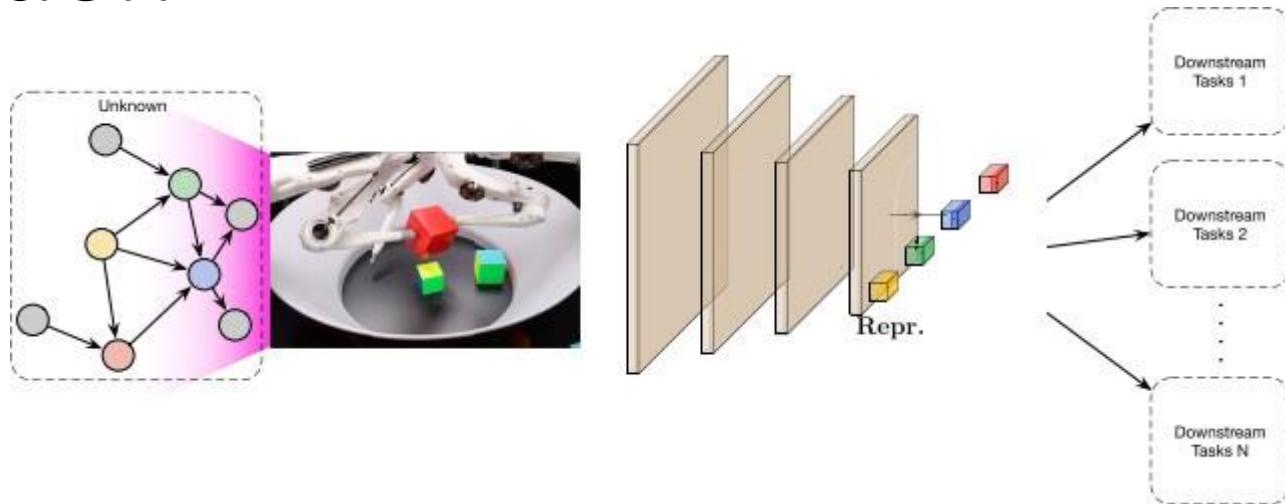


Independent Causal Mechanisms (ICM) Principle.
The causal generative process of a system's variables is composed of autonomous modules that do not inform or influence each other. In the probabilistic case, this means that the conditional distribution of each variable given its causes (i.e., its mechanism) does not inform or influence the other mechanisms.

- **INFLUENCE:** changing (or performing an intervention upon) one mechanism $P(X_i | \mathbf{PA}_i)$ does not change any of the other mechanisms $P(X_j | \mathbf{PA}_j)$ ($i \neq j$) .
- **INFORMATION:** knowing some other mechanisms $P(X_j | \mathbf{PA}_j)$ ($i \neq j$) does not give us information about a mechanism $P(X_i | \mathbf{PA}_i)$.

(Bernhard Schölkopf. Causality for Machine Learning. 2019)

Causal representation learning



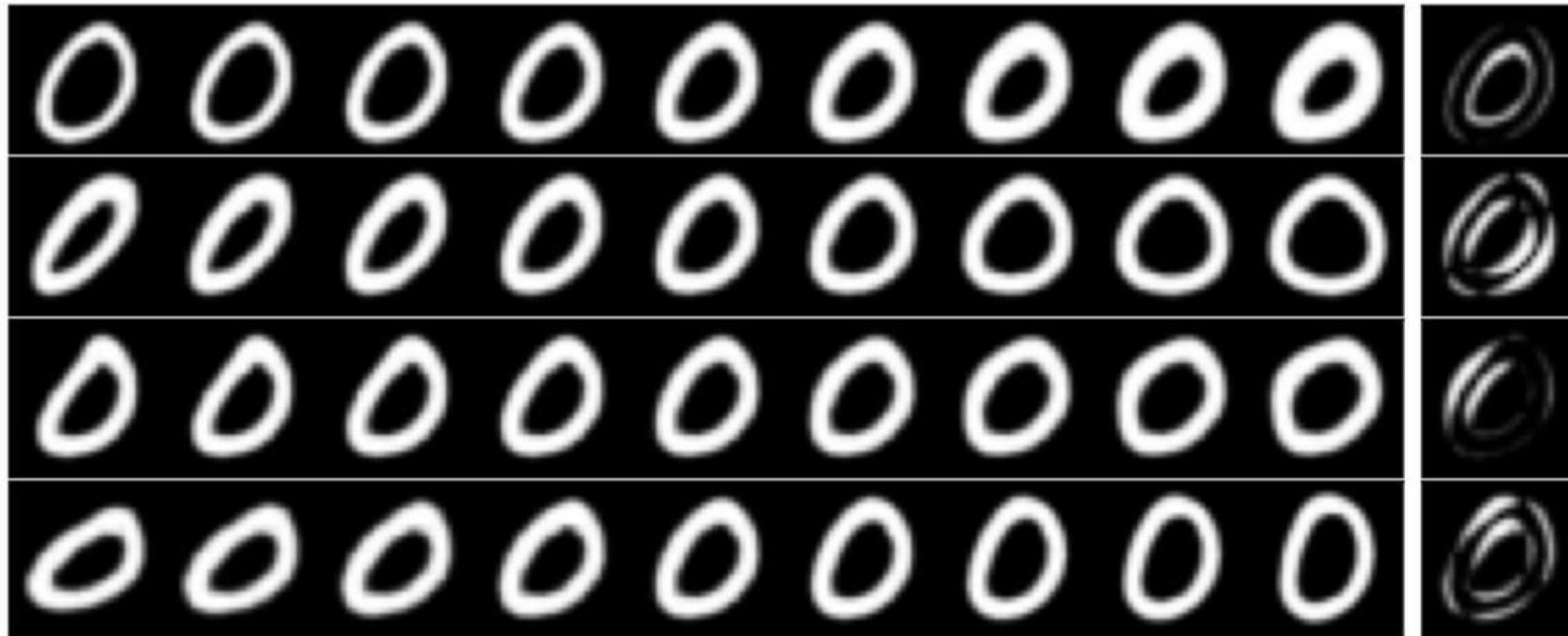
Toward Causal Representation Learning

This article reviews fundamental concepts of causal inference and relates them to crucial open problems of machine learning, including transfer learning and generalization, thereby assaying how causality can contribute to modern machine learning research.

By BERNHARD SCHÖLKOPF^{ID}, FRANCESCO LOCATELLO^{ID}, STEFAN BAUER^{ID}, NAN ROSEMARY KE,
NAL KALCHBRENNER, ANIRUDH GOYAL, AND YOSHUA BENGIO^{ID}

Example

Line thickness



Angle

Upper width

Height

Zheng et al., On the Identifiability of Nonlinear ICA: Sparsity and Beyond, NeurIPS 2022



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



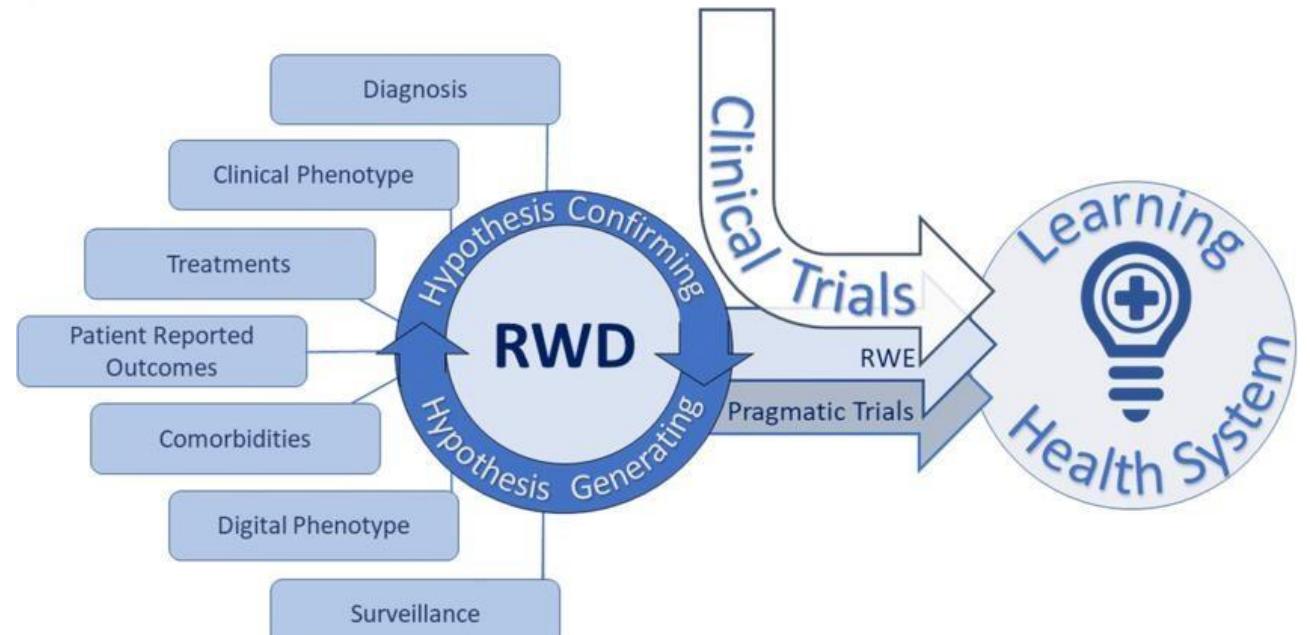
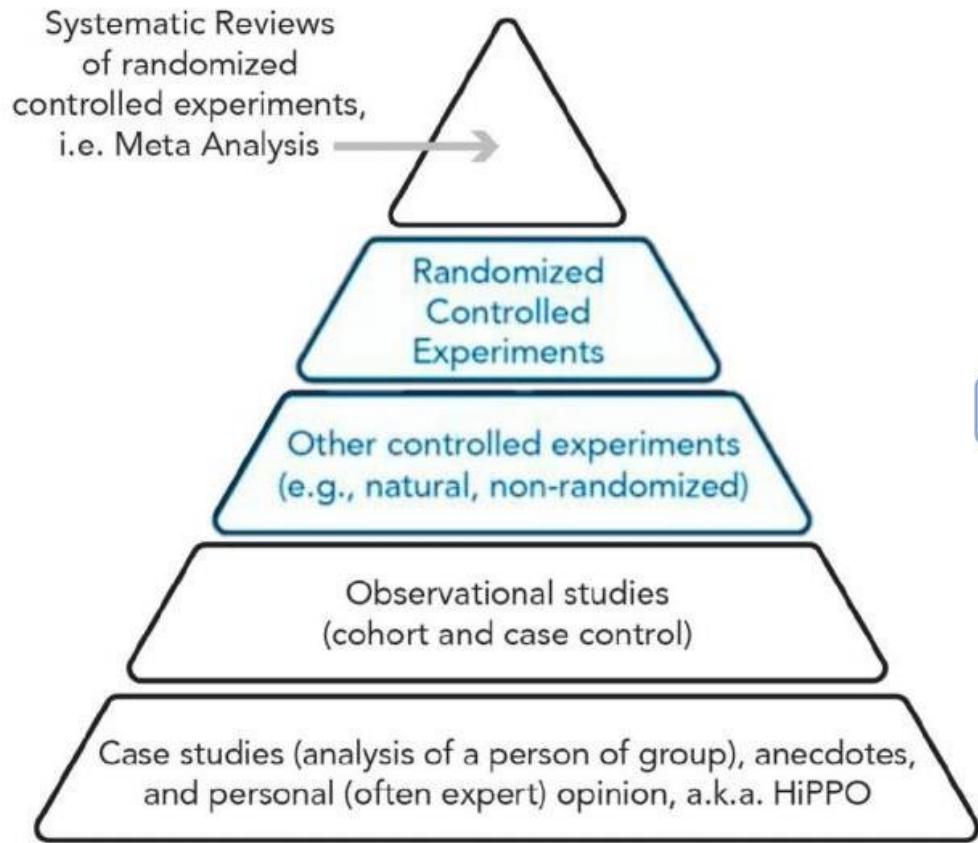
Causality in the age of artificial intelligence



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

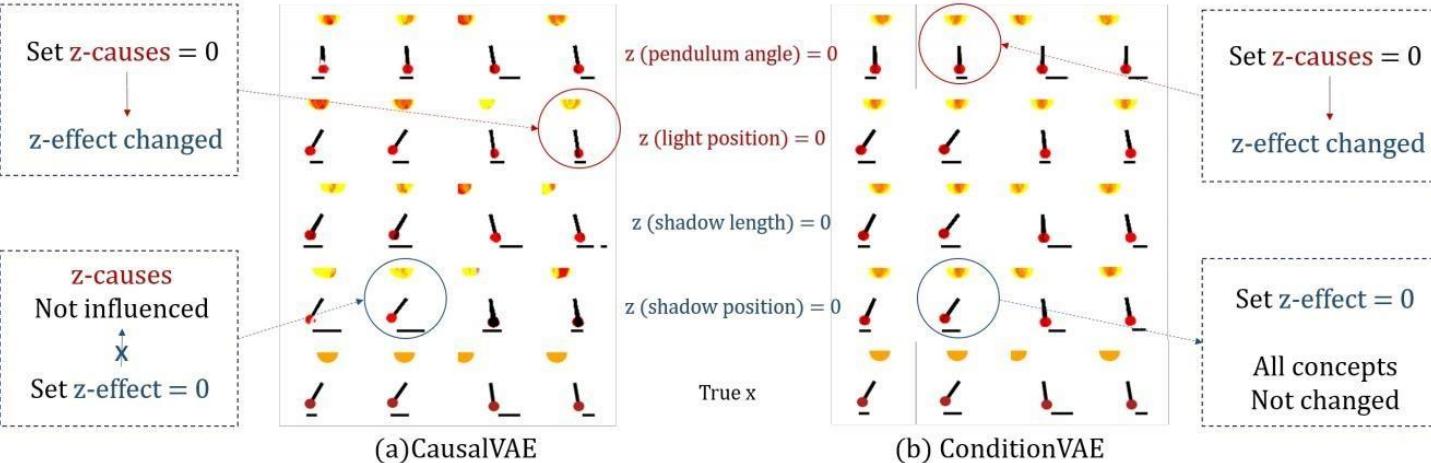


Pharmaceutical industry

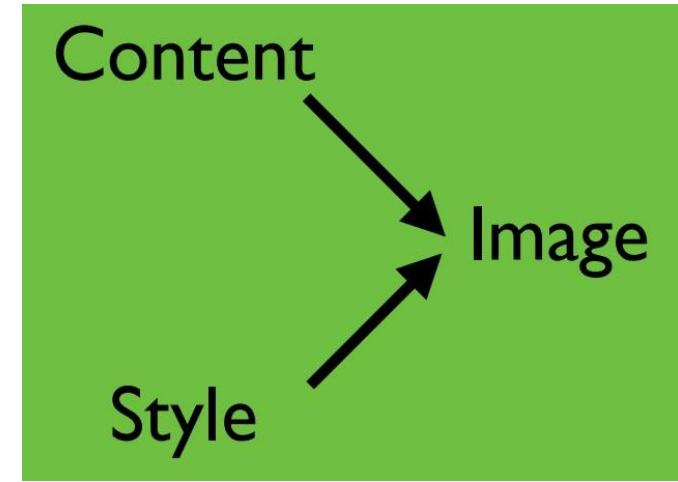


Real-world evidence

Generative AI

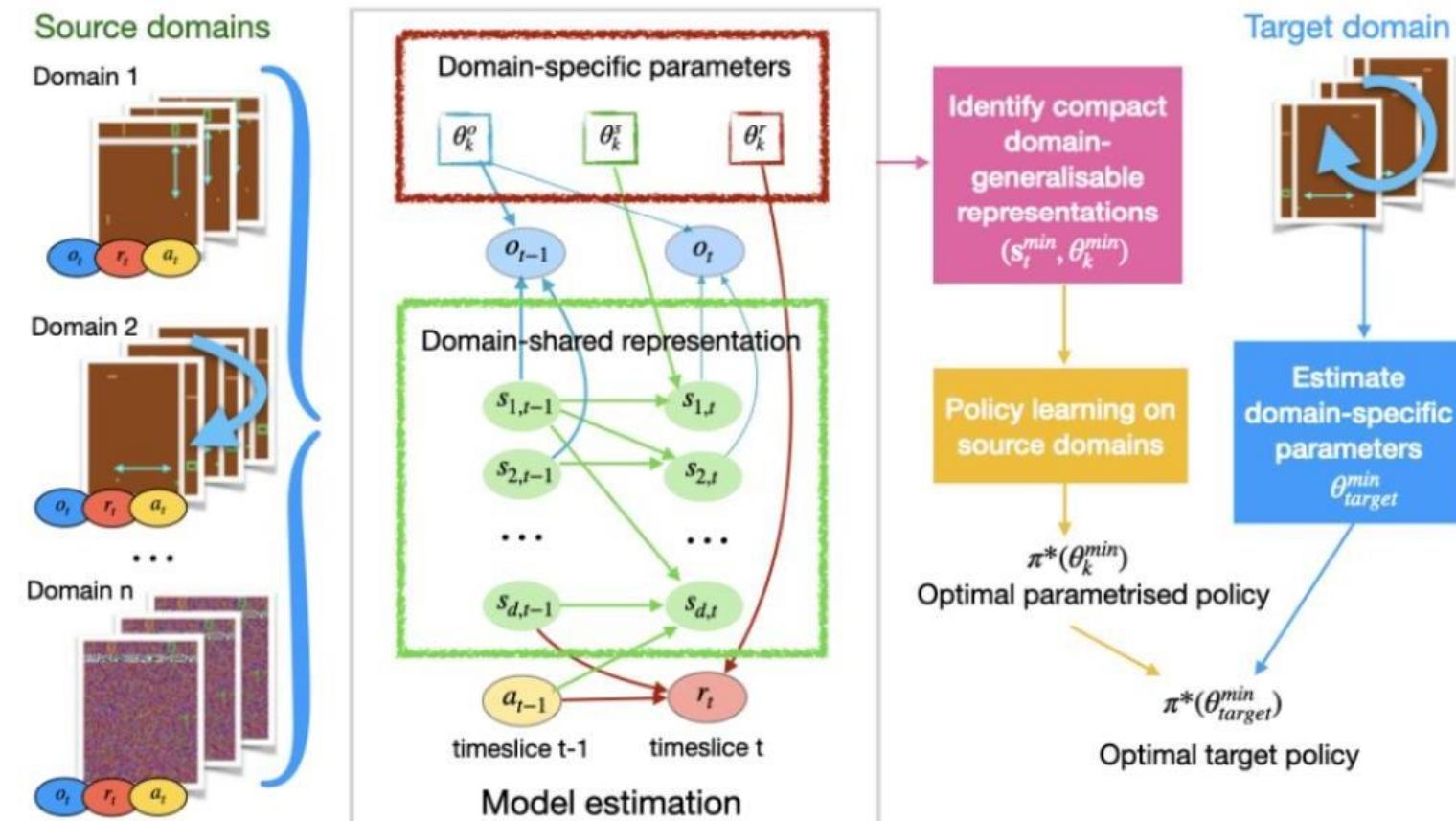


- minimal change

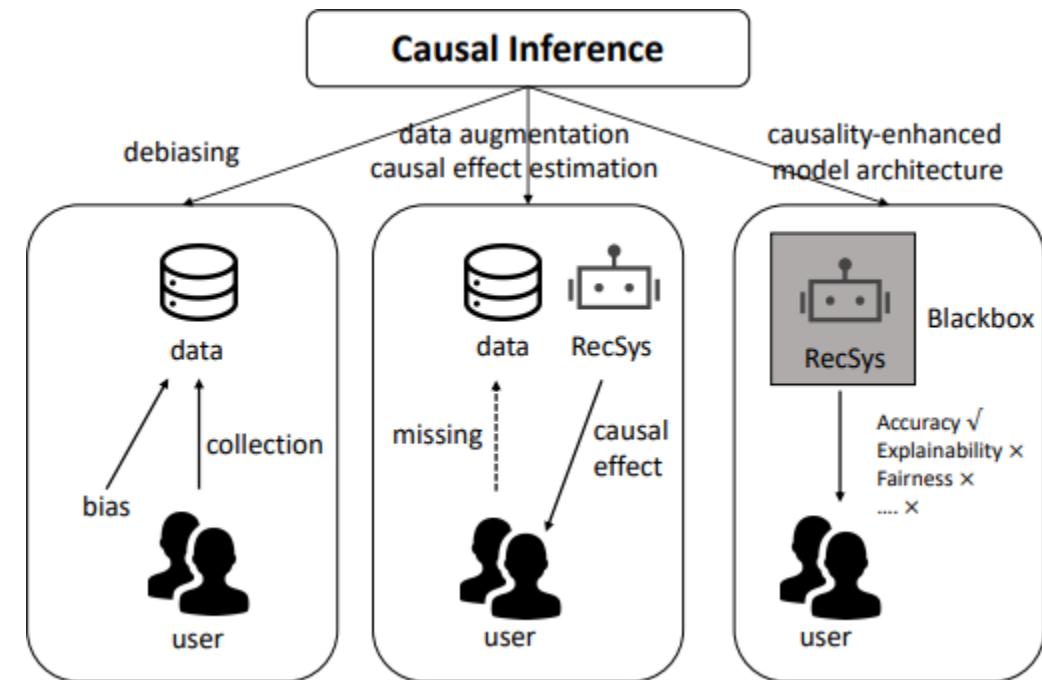
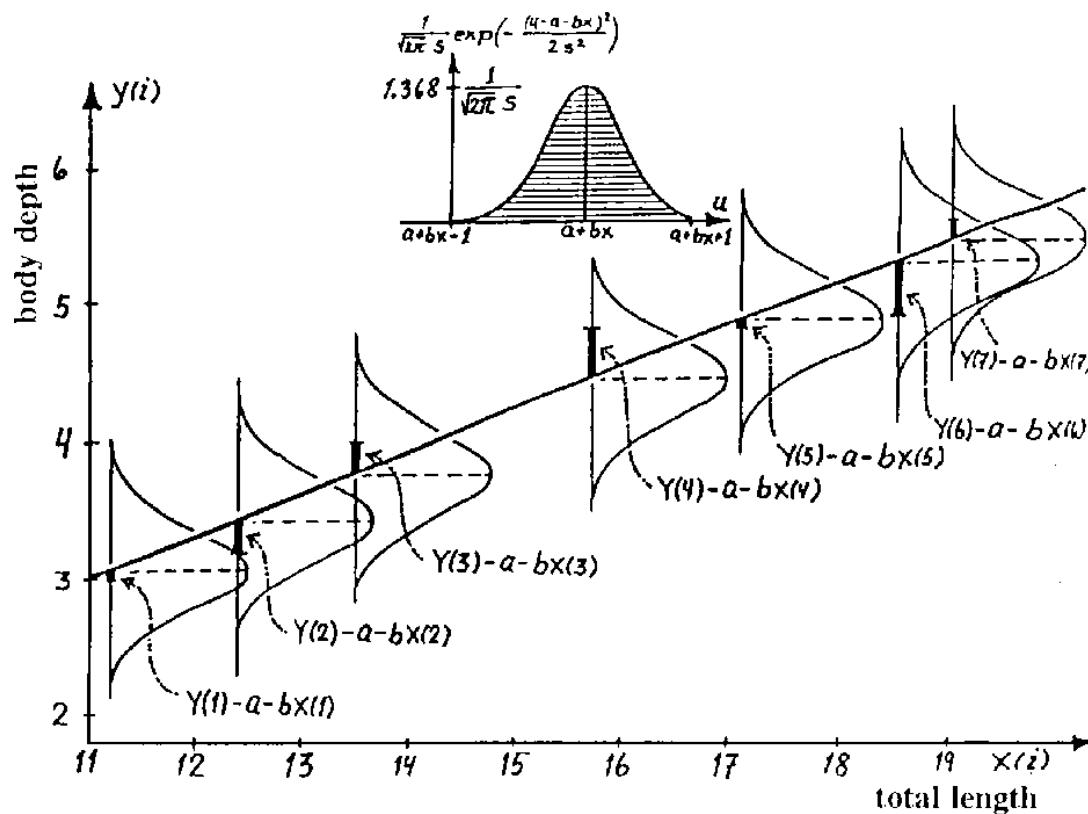


Domain adaptation

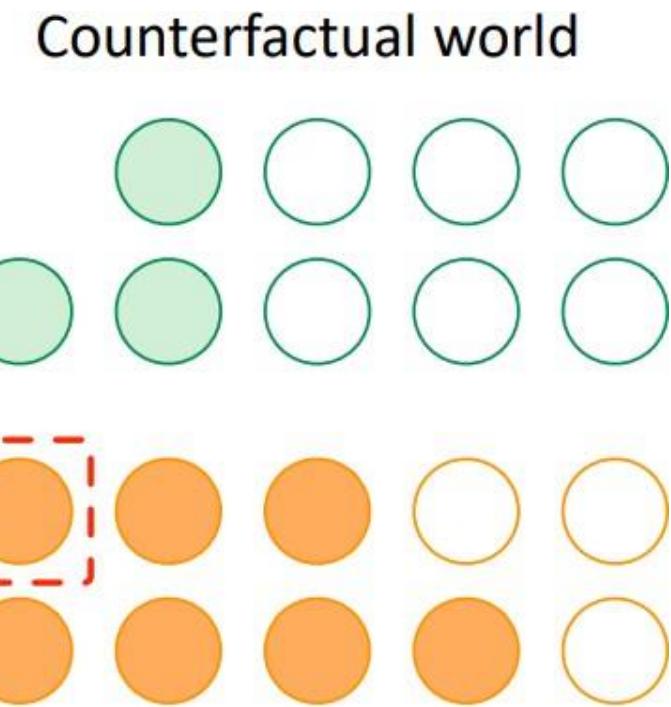
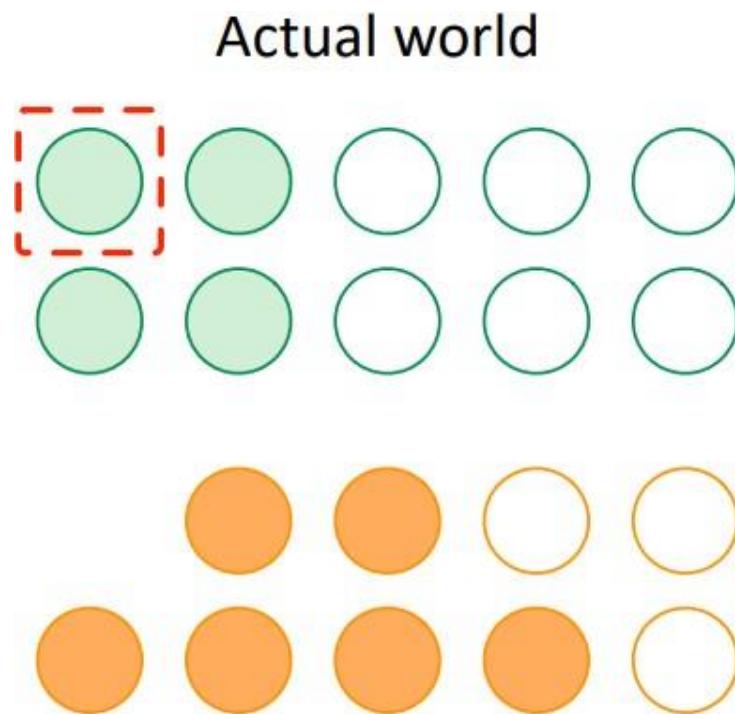
- causal invariance



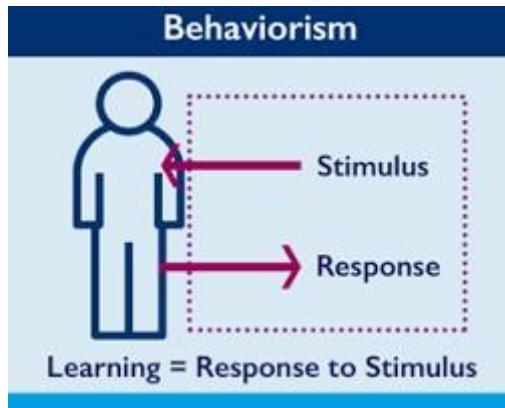
Individualized recommendation



Fairness



Causal AI

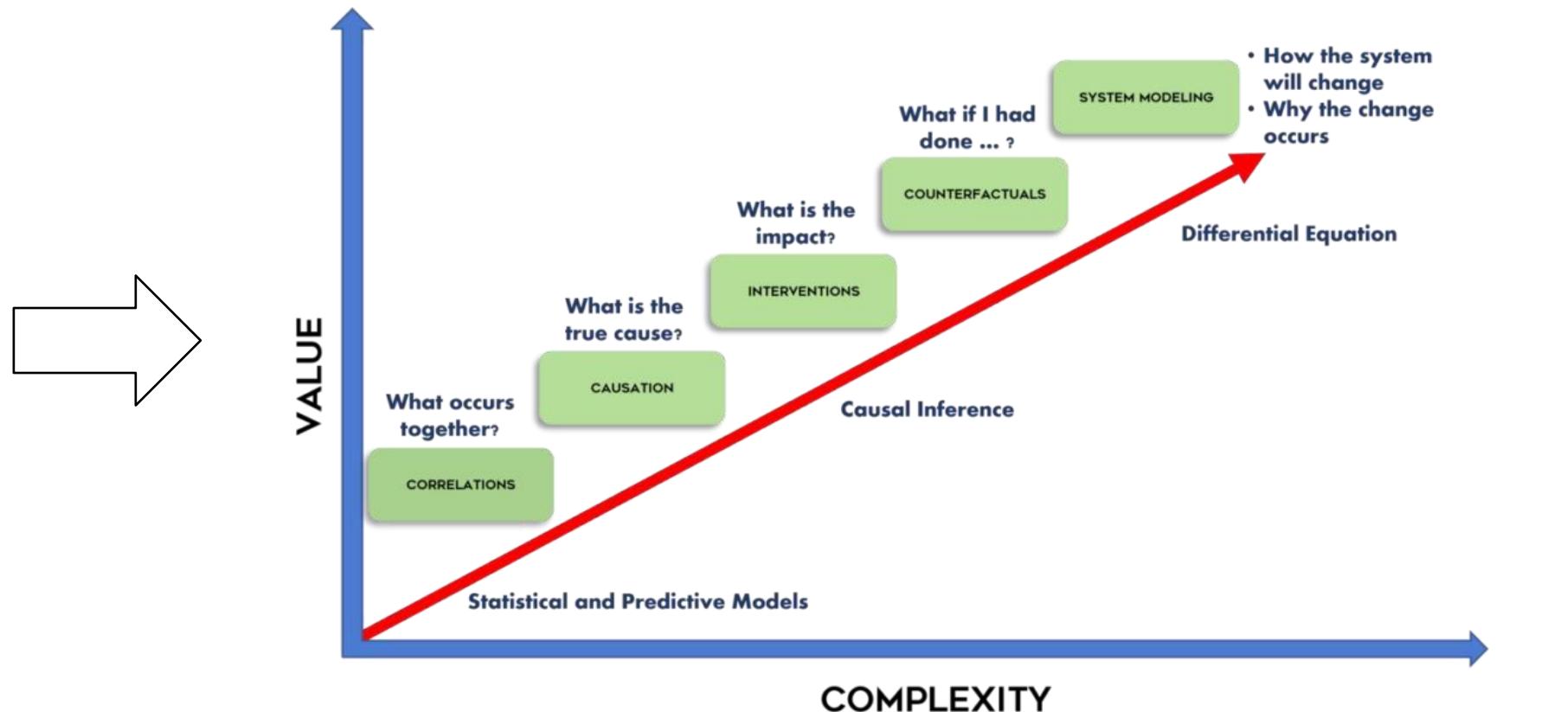


Deep Learning for Causality

- Modeling functional relationships
- Learning distributions over graphs
- Representations as rich compositions of learned features
- Latent causal variables

Causality for Deep Learning

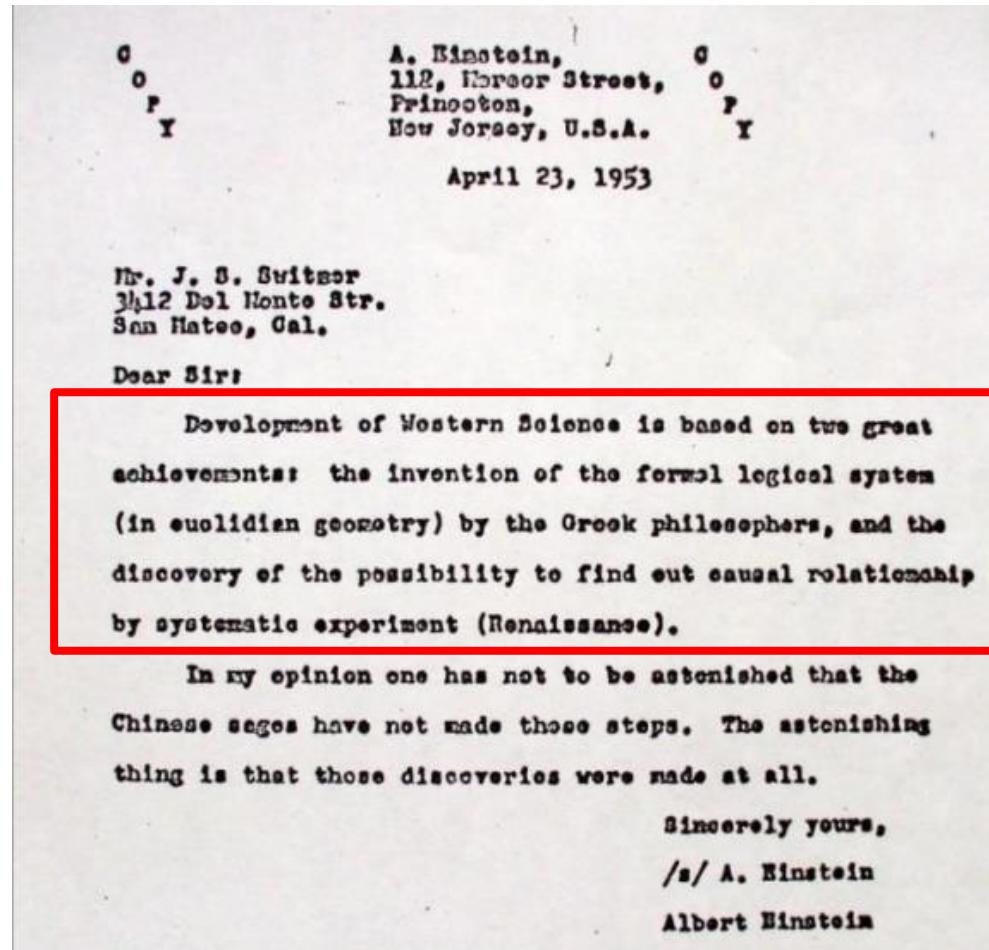
- Why causality for DL
- Benchmarks for causal learning in DL
- Objectives & architectures for causal learning in DL
- Using notions of causality to help DL



Causal thinking

Foreword to the *I Ching* 易经 by Carl Gustav Jung

HTML Edition by Dan Baruth



ised upon the principle of causality, and causality is considered to be an axiomatic truth. But a great change in o causal in 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 are not causally 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 this 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

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 it 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 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 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 able of natural laws constituting empirical reality holds more significance for him than a causal explanation of events that, moreover, must usually be separated

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 tter 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 lly sifts, weighs, selects, classifies, isolates, the Chinese picture of the moment encompasses everything down to the minutest nonsensical detail, because all of

ints 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 l. 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 inevitable rgument 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 ho 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 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 -lasting traces.

d 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 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

ave termed synchronicity,¹² a concept that formulates a point of view diametrically opposed to that of causality. Since the latter is a merely statistical truth and s 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 lives as well as with the subjective (psychic) states of the observer or observers.

y comparable to that of the modern physicist, who cannot deny that his model of the world is a decidedly psychophysical structure. The microphysical event 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 e 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 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 its *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. Transcendental situations are unique and cannot be repeated. Experimentation with synchronicity seems to be impossible under ordinary



Plato's Allegory of the Cave



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



Cassandra's curse

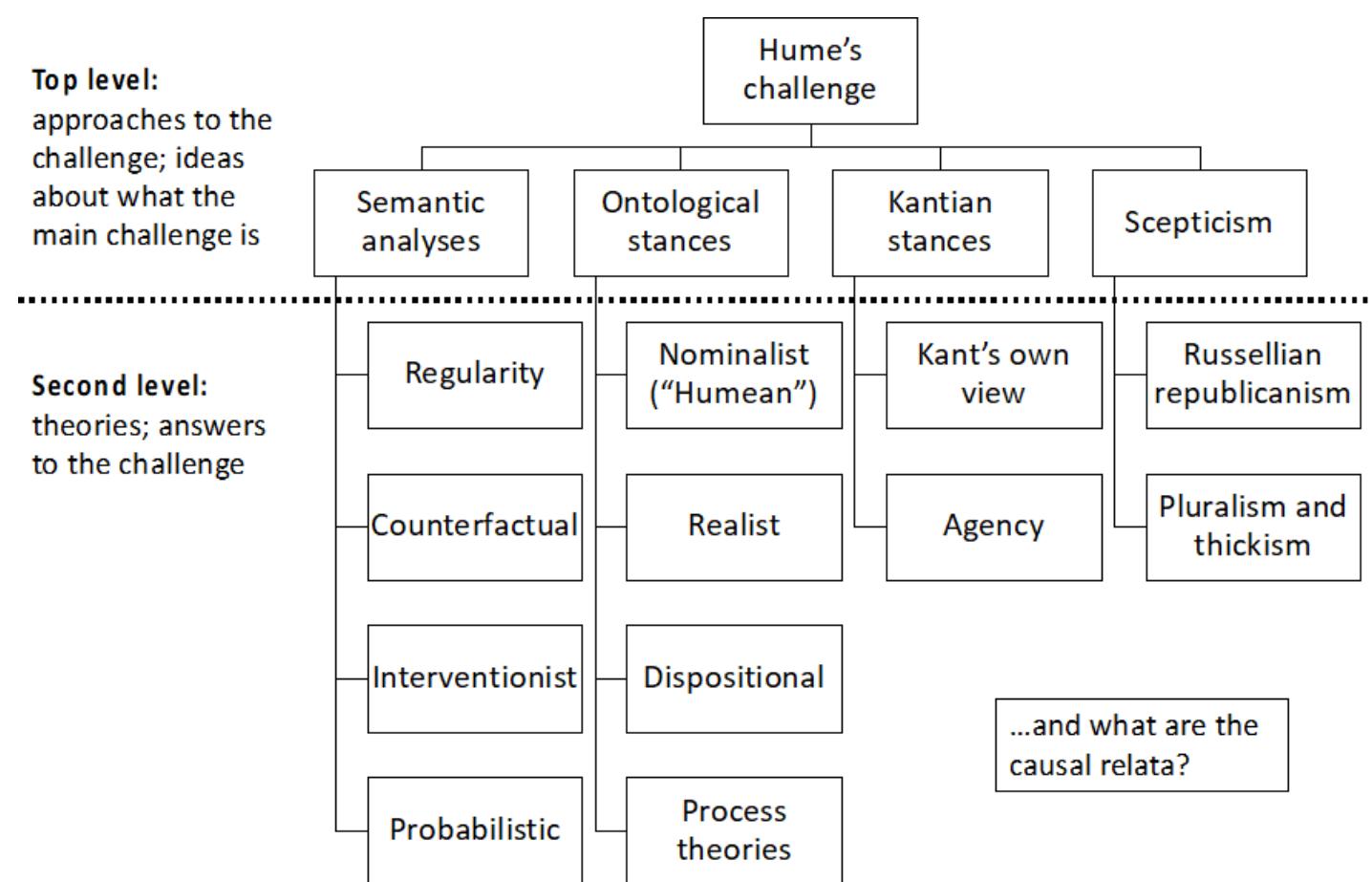


Causality & Freedom

- No causation without manipulation (D. Rubin)
- Allegory on Three world:
 - Laplacian world;
 - intervenable world;
 - random world
- “Causal” Stoicism



Unresolved disputes ...





Thank you!

"***Felix, qui potuit rerum cognoscere causas***"
—Vigil, verse 490 of Book 2 of the
"*Georgics*"