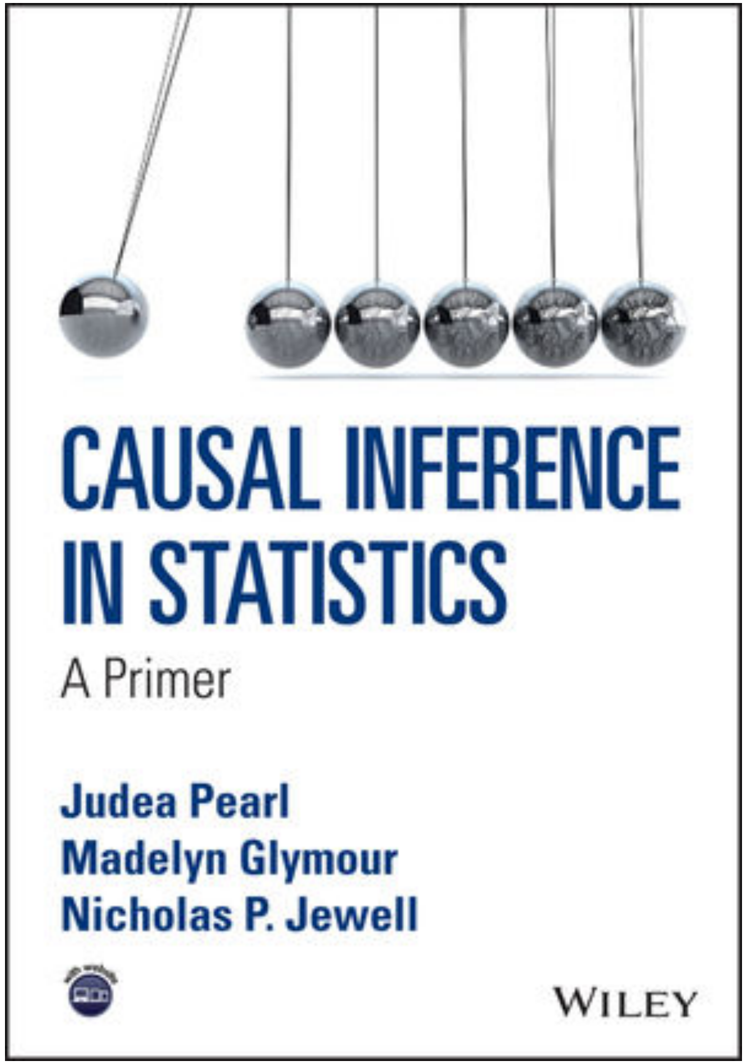
Learning And Mining from DatA
http://lamda.nju.edu.cn

因果强化学习概述

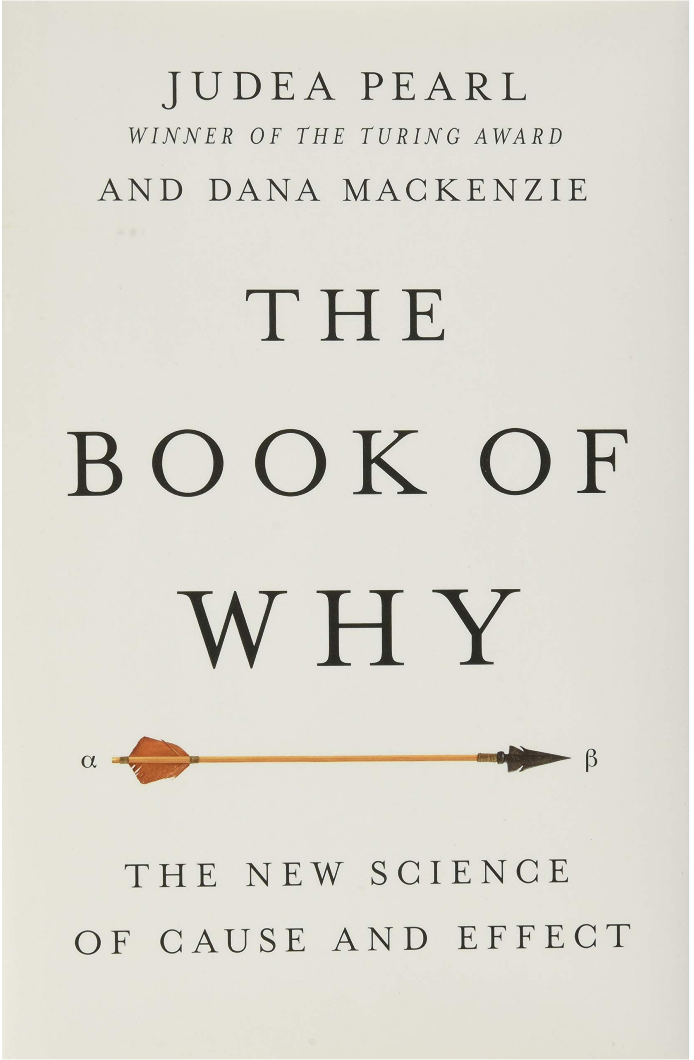
田鸿龙

Contents

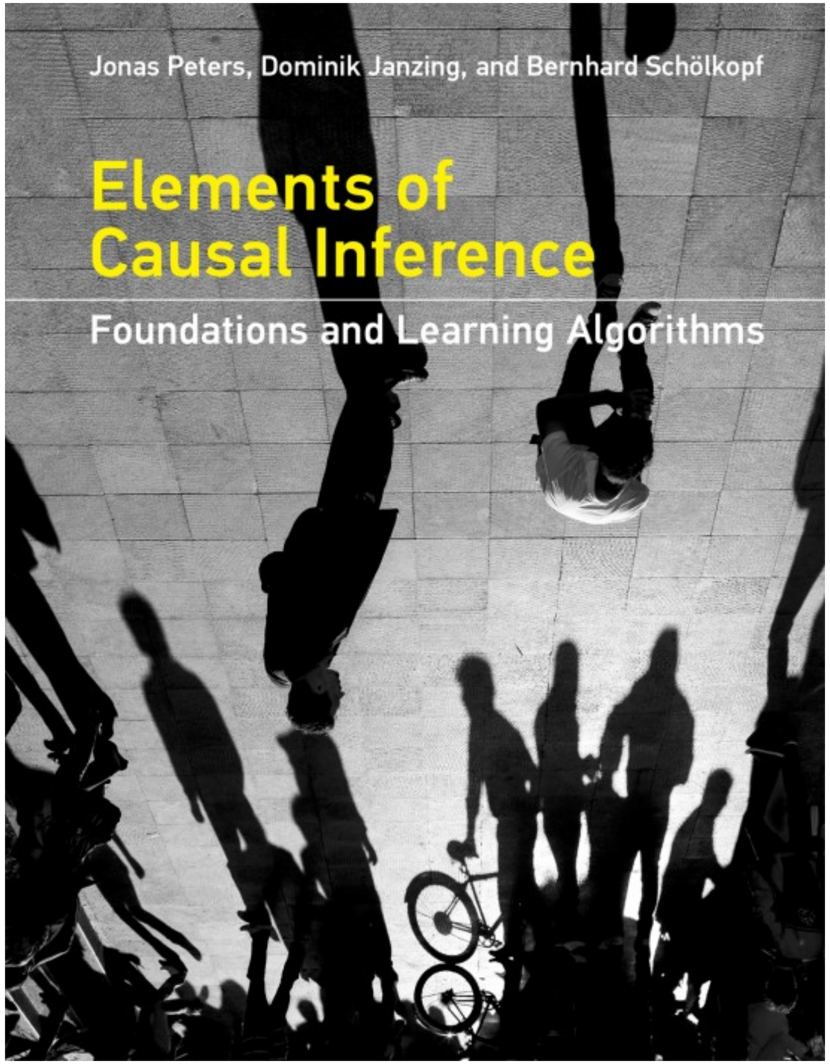
Books



CAUSAL INFERENCE IN STATISTICS
A Primer
Judea Pearl
Madelyn Glymour
Nicholas P. Jewell
WILEY



JUDEA PEARL
WINNER OF THE TURING AWARD
AND DANA MACKENZIE
THE BOOK OF WHY
α
β
THE NEW SCIENCE OF CAUSE AND EFFECT



Jonas Peters, Dominik Janzing, and Bernhard Schölkopf
Elements of Causal Inference
Foundations and Learning Algorithms

Causal Inference

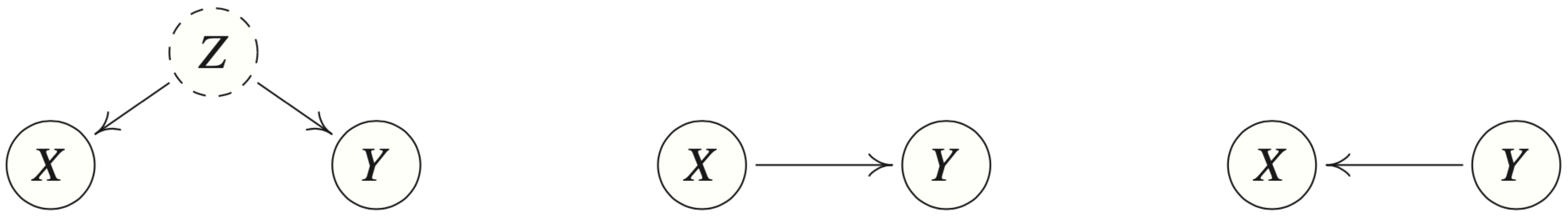
- The science of “why”.
- 相关性不等于因果性。
- 对因果的认知是人类的“先验”，~~甚至可能是最基本的先验~~。（因为人类永远无法完全得到任何完整的因果结构，但是人类相信这样的结构存在，下意识的思考事物之间的因果关系。）
- 因果的存在让人类形成了“后悔”的能力，进一步讲，形成了道德的概念，从而形成法律的概念。
- 因果的存在同时解释了一部分情感的存在。
- “道德机器”、“情感机器”的诞生，（起码）需要让人工智能理解因果。

Causal Inference(cont.)

- 相关性不等于因果性。
- 但是相关性和因果性“相关”。
- 相关->因果：
  - Reichenbach’s common cause principle
  - 两件事物相关，必有背后的因果。（未必有直接的因果）
- 因果->相关：
  - 大部分因果关系表现为相关性。

Causal Inference(cont.)

- **Reichenbach's common cause principle**: If two random variables X and Y are statistically dependent ($X \not\perp Y$), then there exists a third variable Z that causally influences both. (As a special case, Z may coincide with either X or Y .) Furthermore, this variable Z screens X and Y from each other in the sense that given Z, they become independent, $X \perp Y \mid Z$.



Causal Inference(cont.)

Judea Pearl

Donald Rubin

Resolving disputes between J. Pearl and D. Rubin on causal inference.

Why is there a dispute between Judea Pearl and Rubin with respect to the theoretical frameworks used in causal modelling?

Structural Causal Models

Definition(Structural Causal Models): A structural causal model M (or, data generating model) is a tuple (V, U, F, P(u)), where

- $V = \{V_1,\ldots,V_n\}$ are endogenous variables.
- $U = \{U_1,\ldots,U_m\}$ are exogenous variables.
- $F = \{f_1,\ldots, f_n\}$ are functions determining V, for each V_i, $V_i \leftarrow f_i(Pa_i, U_i)$, where $Pa_i \subset V$, $U_i \subset U$.
- P(u) is a distribution over U.

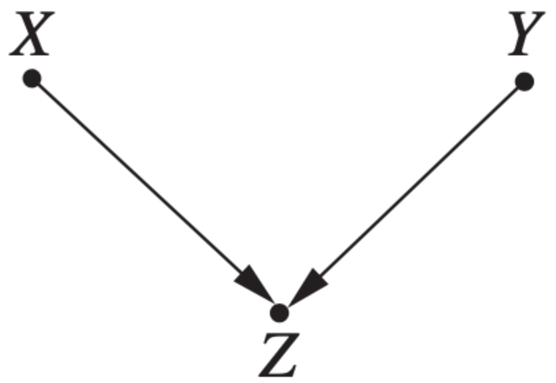
- 假设：世界是一个SCM，F是物理学定律。

Causal Bayesian Network

- Causal Bayesian Network和Bayesian Network的形式相同，但是表达的含义不同。
- Causal Bayesian Network也是Directed Acyclic Graph，X->Y表达的含义是：X是Y的直接因。
- Causal Bayesian Network的结构本身蕴涵了大量的先验知识，这些知识可能是错误的（意味着需要不断学习和调整）。
- 本质上，因果推断正是充分利用了这些先验的“因果知识”，提升了模型的表示能力。

Example1

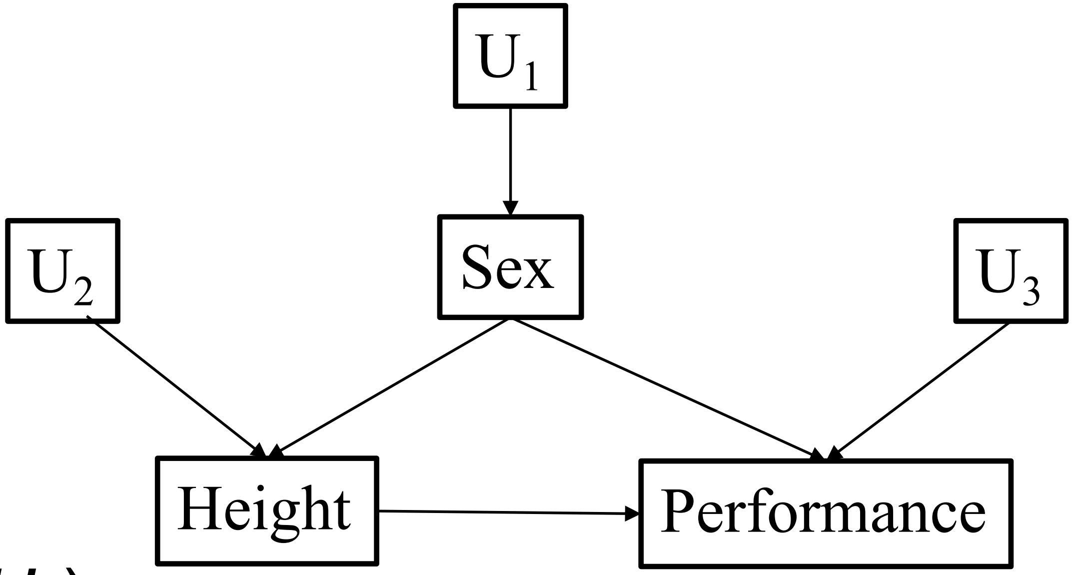
- SCM：
  - $U=\{X, Y\}$
  - $V=\{Z\}$
  - $F=\{f_Z\}$, f_Z :$Z=2X+3Y$



- This model represents the **salary (*Z*)** that an employer pays an individual with ***X*** **years of schooling** and ***Y*** **years in the profession**. X and Y both appear in f_Z , so X and Y are both direct causes of Z. If X and Y had any ancestors, those ancestors would be potential causes of Z.

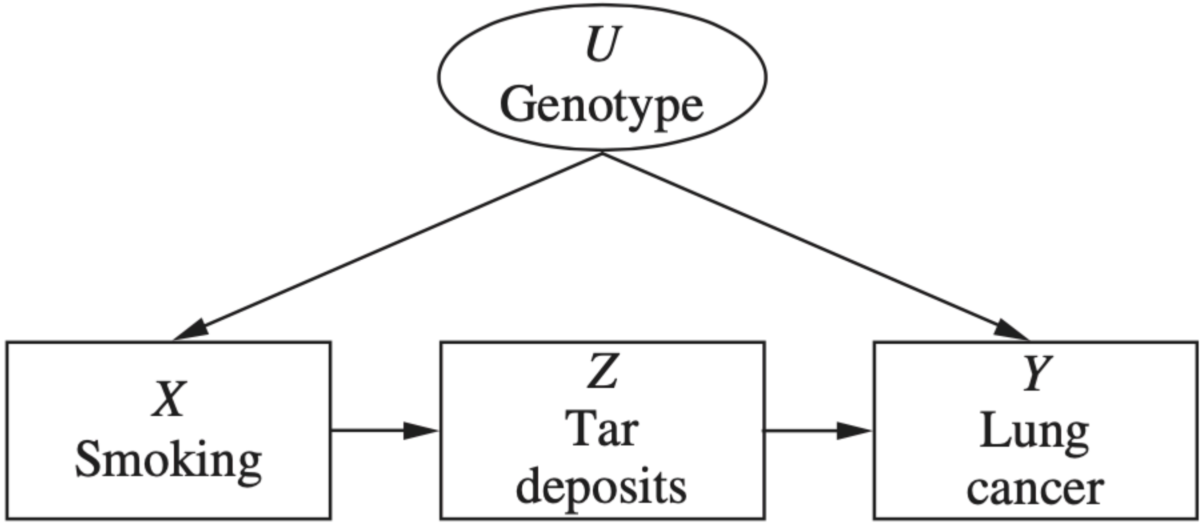
Example2

- SCM:
  - V = {Height, Sex, Performance}
  - $U = \{U_1, U_2, U_3\}$
  - $F = \{f_1, f_2\}$
    - Sex = U_1
    - Height = f_1(Sex, U_2)
    - Performance = f_2(Height, Sex, U_3)



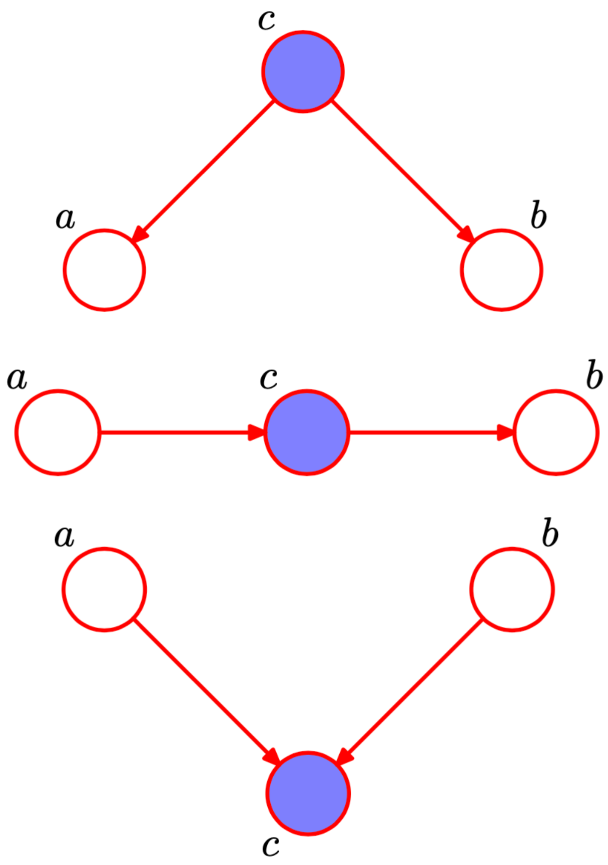
Example3

- 省略了其他外部变量
- 假设Genotype是不可观测的
- 试图计算Smoking对Lung cancer的因果效应



Causal Bayesian Network(cont.)

- Bayesian Network的三种基本结构：
- Fork(tail-to-tail)
  - $p(a,b \mid c) = \dfrac{p(a|c)p(b|c)p(c)}{p(c)} = p(a \mid c)p(b \mid c)$
- Chain(head-to-tail)
  - $p(a,b \mid c) = \dfrac{p(\mathrm{c|a})p(b|c)p(\mathrm{a})}{p(c)} = p(a \mid c)p(b \mid c)$
- Collider(head-to-head)
  - $p(a,b) = p(a)p(b)$
  - $p(a,b \mid c) = \dfrac{p(\mathrm{a})p(\mathrm{b})p(\mathrm{c|a,b})}{\mathrm{p(c)}} \neq \mathrm{p(a \mid c)p(b \mid c)}$



Causal Bayesian Network(cont.)

- **Definition(*d*-separation):** A path p is blocked by a set of nodes Z if and only if
- 1. p contains a chain of nodes A→B→C or a fork A←B→C such that the middle node B is in Z (i.e., B is conditioned on), or
- 2. p contains a collider A → B ← C such that the collision node B is not in Z, and no descendant of B is in Z.

Causal Bayesian Network(cont.)

- Causal Bayesian Network和Bayesian Network的关系：
  - Bayesian Network描述的是变量之间的相关性关系；
  - Causal Bayesian Network描述的是变量之间的因果性关系；
  - 两个变量相关往往不能确定它们之间的因果关系
    - 太阳升起和公鸡打鸣
    - 吸烟，肺癌和基因型
  - 两个变量之间有因果关系往往说明它们相关
    - 可以构造出反例

$$V = \{X, Y, Z\}, U = \{U_X, U_Y, U_Z\}, F = \{f_X, f_Y, f_Z\}$$

$$f_X : X = U_X$$

$$f_Y : Y = \begin{cases} a & \text{IF } X = 1 \text{ AND } U_Y = 1 \\ b & \text{IF } X = 2 \text{ AND } U_Y = 1 \\ c & \text{IF } U_Y = 2 \end{cases}$$

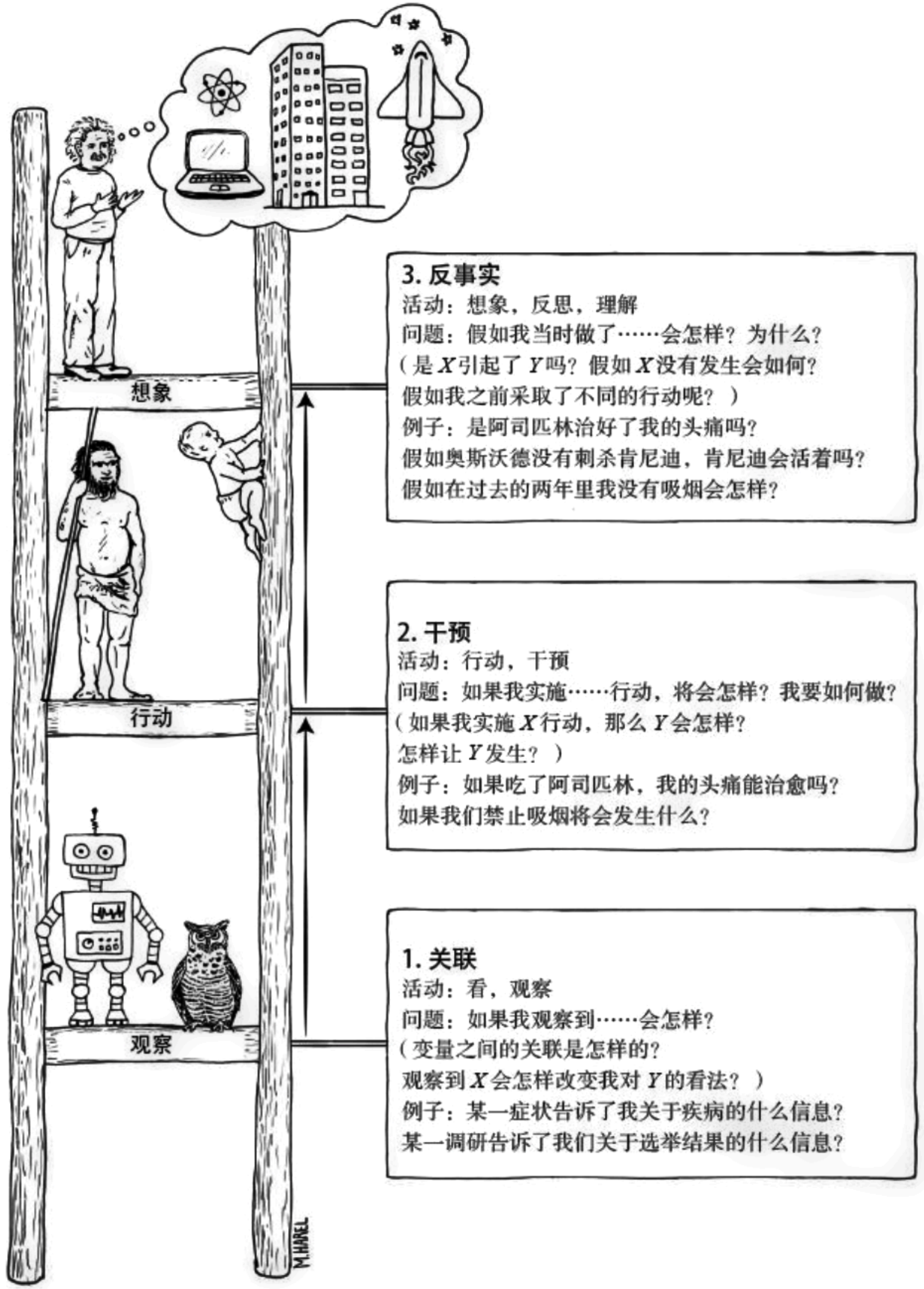
$$f_Y : Z = \begin{cases} i & \text{IF } Y = c \text{ OR } U_Z = 1 \\ j & \text{IF } U_Z = 2 \end{cases}$$

Pearl’s Causal Hierarchy

PCH又被称为因果之梯，描述了三种不同层次的因果关系，分别是关联，干预和反事实。

三种不同的关系分别对应了人类的三种活动，即观察，行动和想象。

其中关联即传统统计学中的概率论，**干预**和**反事实**是因果推断关注的主要内容。



Interventions 干预

- Observational studies: 用数学的方式模拟随机对照试验
- Interventions主要使用以下技术，适用的范围由浅到深，其中Adjustment Formula适用于完全观测的因果图，Do-calculus在任何因果图上都具有完备性的证明
- The Adjustment Formula
- The Back-Door Criterion
- The Front-Door Criterion
- Do-Calculus
- Conditional Interventions

Randomized Controlled Trial

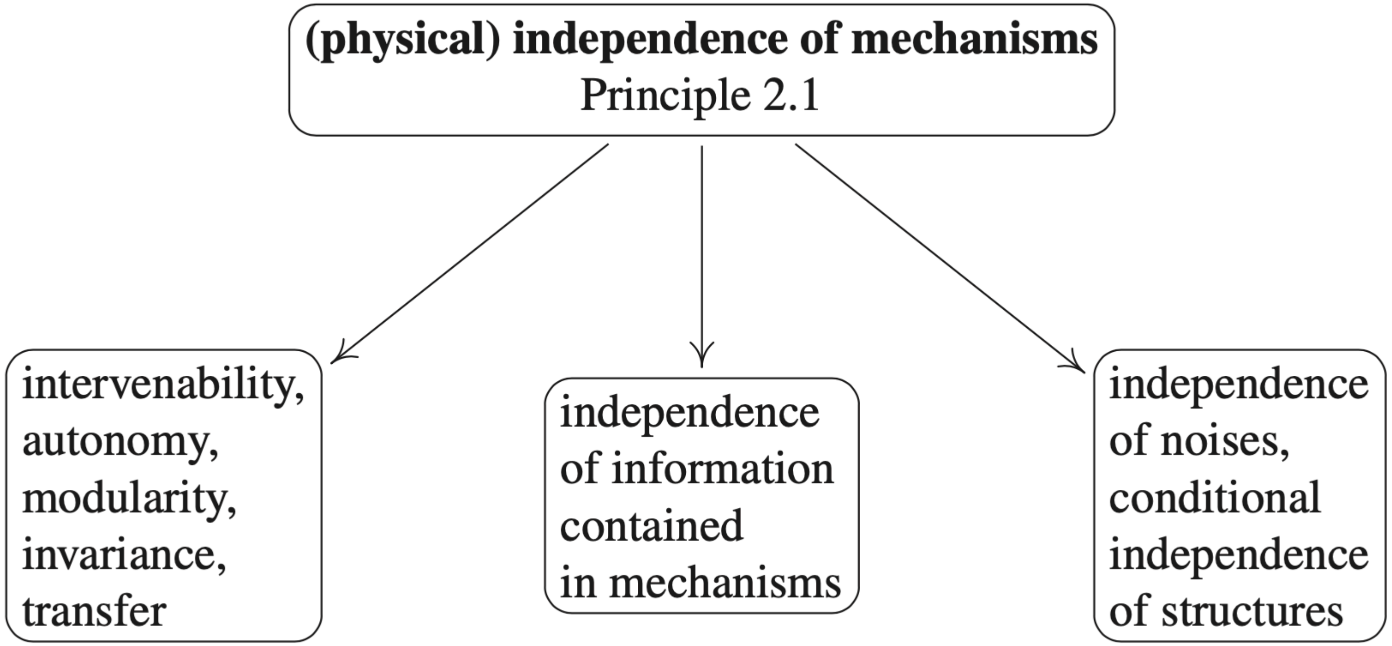
- 随机对照试验的基本方法是，将研究对象**随机分组**，对不同组实施不同的**干预**，在这种严格的条件下对照效果的不同。在研究对象数量足够的情况下，这种方法可以抵消已知和未知的**混杂因素**对各组的影响。
- 例子：测试化肥的效果。
  - 简单的对照试验：在不同的时间或地点控制化肥用量。
  - 随机对照试验：在对照试验的基础上随机分组。

Observational studies

- The art and science of inferring causation without experiments
- This can only be accomplished if extras assumptions are added
- 进行Observational studies的目的是避免实验，例如：
  - When experiments are impossible for unethical/practical reasons(The case for smoking/lung cancer link)
  - When there are many experiments to perform(biological systems)
- extras assumptions往往指的是因果图（最好是SCM，但几乎不可能），如何得到这些extras assumptions是Casual Learning要考虑的内容

Independent mechanisms

- **Principle**(Independent mechanisms): The causal generative process of a system’s variables is composed of autonomous modules that do not inform or influence each other.



Independent mechanisms (cont.)

- Example：
  - altitude A
  - average annual temperature T
- 从数学上：$p(a,t) = p(a|t)\ p(t) = p(t|a)\ p(a)$
- 实际上： $p(a,t) = p(t|a)\ p(a)$
- $p(t|a)$和$p(a)$ 是两个彼此独立的mechanisms，可以把两者看成具有输入输出功能的黑箱。
- Independent mechanisms保证了Interventions的有效性。

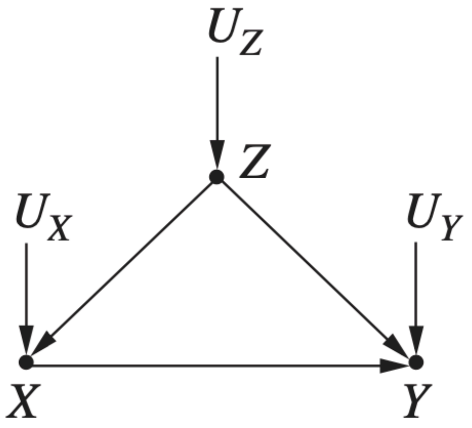
The do operator (Pearl's notation)

- 观察：P(A | B = b): the probability of A being true given an observation of B = b
  - That is, no external intervention
  - P(A | B = b)被称为conditional distribution
- 干预：P(A | **do**(B = b)): the probability of A given an intervention that sets B to b
  - P(A | do(B)): some shorter notation for do(B) = true
  - P(A | do(B))被称为interventional distribution

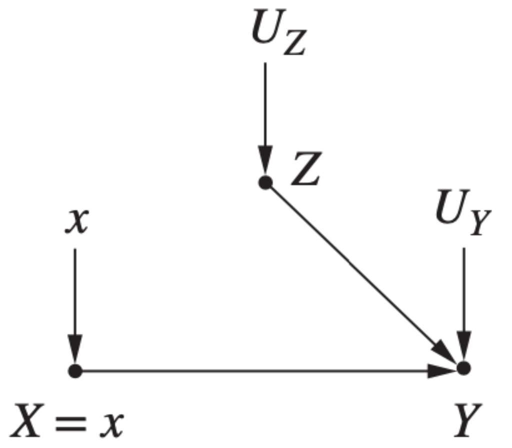
The Adjustment Formula

- Simpson's Paradox

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

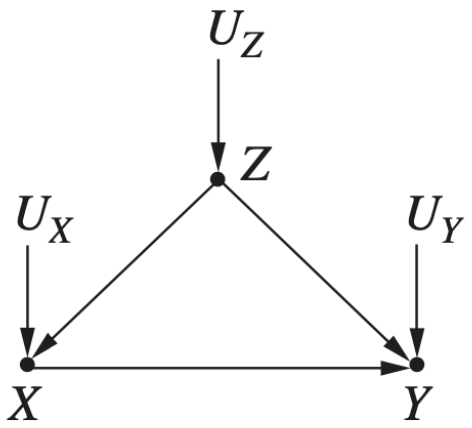


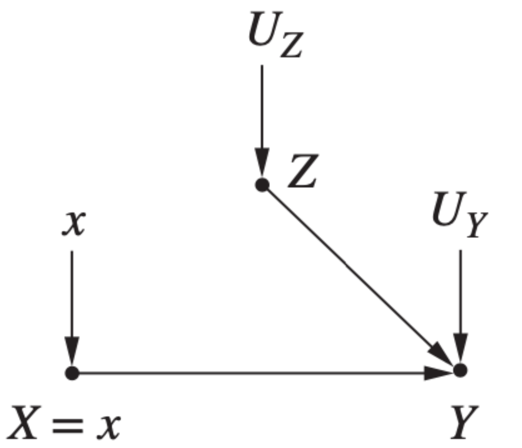
- *X* standing for drug usage, and *Y* standing for recovery, *Z* representing gender.
- 关联：$P(Y=1|X=1)-P(Y=1|X=0)$
- 干预：$P(Y=1|do(X=1))-P(Y=1|do(X=0))$



The Adjustment Formula(cont.)

- $p(y|x)$表示的是x和y的相关性， $p(y|\mathrm{do}(x))$表示的是x和y的因果性。
- 随机对照试验得到的图是右边的图，所谓“adjustment”就是通过左边的图的数据，计算右边的图对应的结果。
- 将左边的图转化为右边的图的过程称为“manipulate”，用下标m表示，于是： $p(y|\mathrm{do}(x)) = p_m(y|x)$





The Adjustment Formula(cont.)

- $p_m(y|x) = \sum_z p_m(y,z|x) = \sum_z p_m(y|x,z)p_m(z|x)$
- $p_m(y|x,z) = p(y|x,z)$
- $p_m(z|x) = p_m(z) = p(z)$
- The Adjustment Formula:

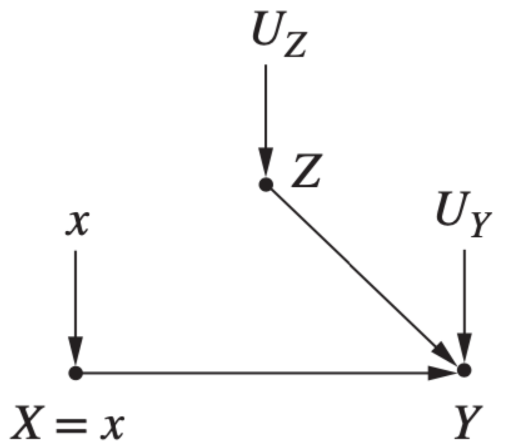
$$p(y|\mathrm{do}(x)) = \sum_z p(y|x,z)p(z)$$

- $p_m(y|x) = \sum_z p_m(y,z|x) = \sum_z p_m(y|x,z)p_m(z|x)$
- $p_m(y|x,z) = p(y|x,z)$
- $p_m(z|x) = p_m(z) = p(z)$
- The Adjustment Formula:

$$p(y|\mathrm{do}(x)) = \sum_z p(y|x,z)p(z)$$

$P(Y=1|do(X=1)) = 0.832$

$P(Y=1|do(X=0)) = 0.7818$



The Adjustment Formula(cont.)

- **Principle**(The Causal Effect Rule):
- *Given a graph G in which a set of variables PA are designated as the parents of X, the causal effect of X on Y is given by*

$$P(Y = y|do(X = x)) = \sum_{z} P(Y = y|X = x, PA = z)P(PA = z)$$

- *where z ranges over all the combinations of values that the variables in PA can take.*
- 假设所有的变量都可以被观察到。

Do-Calculus

- 使用三条规则和概率论公理，根据因果图进行干预。
- Rule1: 如果我们观察到变量W与Y无关（其前提可能是以其他变量Z为条件），那么Y的概率分布就不会随W而改变。
  - $p(y|do(x),z,w) = p(y|do(x),z)$
- Rule2: 如果变量集Z阻断了从X到Y的所有后门路径，那么以Z为条件（对Z进行变量控制），则do（X）等同于see（X）。
  - $p(y|do(x),z) = p(y|x,z)$
- Rule3: 如果从X到Y没有因果路径，我们就可以将do（X）从$p(y|do(x))$中移除。
  - $p(y|do(x)) = p(y)$

Counterfactuals

- 关联，干预和反事实是层层递进的，关联是一种特殊的干预，干预是一种特殊的反事实。
- 反事实是对**既定事实**的否定。
- 例如：从A地到B地，可以走城市公路或者高速公路，假设某人某次从城市公路走用时一小时，请问走高速公路预计的时间？
- 本质上是在计算driving time的期望，然而这个期望不能用do算子表示，这反应了反事实高于干预。
- 同时，反事实是无法通过现实世界的实验计算的，不可能让同一个人在同一个时间做出两个不同的选择：反事实建立在**多个平行的世界**中，反事实寻找除了给定变量（如高速公路还是城市公路）外，**最形似的两个世界**。

Counterfactuals(cont.)

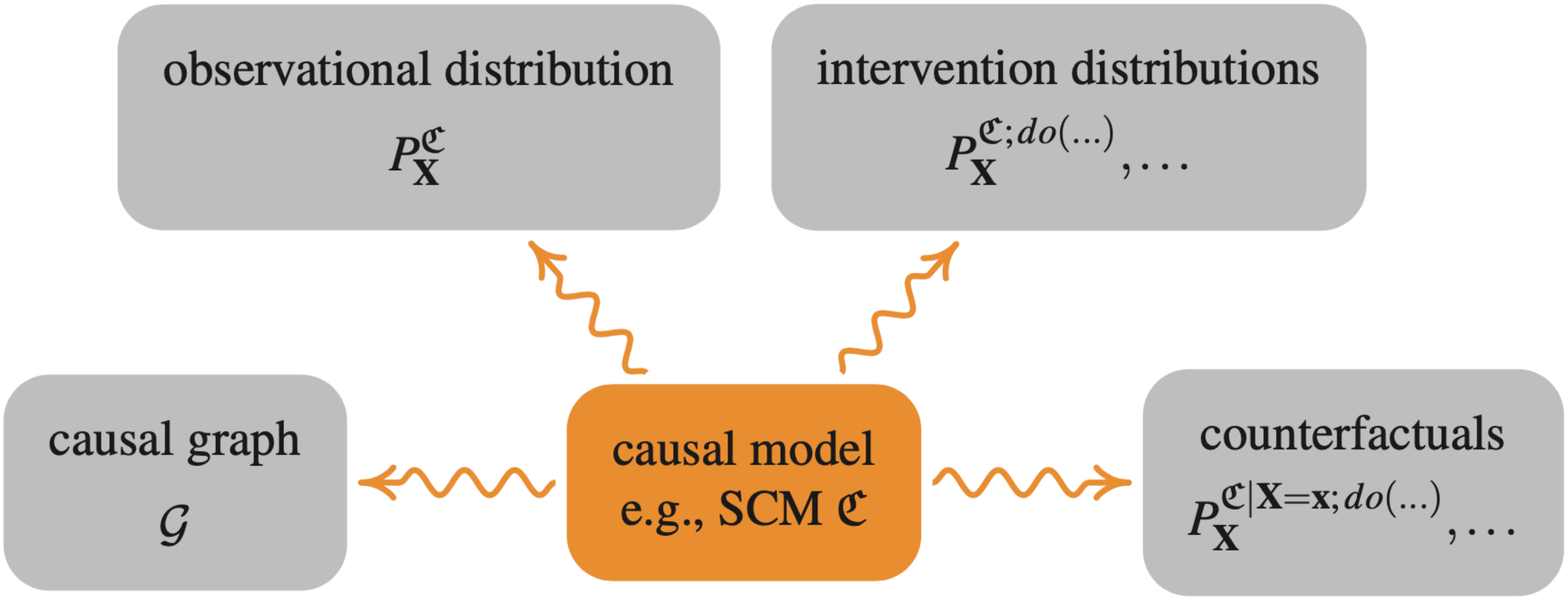
- 例子：有针对一种eye disease的一种treatment
  - 99%的病人用了会痊愈，不用会失明
  - 1%的病人不用会痊愈，用了会失明
- 现在已知一个病人来了，医生用了药，导致这个病人失明了，请问（这是一个反事实问题），如果医生没有用药病人的病情会怎样？
- SCM：
  - $V = \{T, B\}$ $U = \{U_T, U_B\}$
  - $T = U_T$
  - $B = T \cdot U_B + (1-T) \cdot (1-U_B)$
- 可以根据上面的例子得出$U_T = U_B = 1$
- 接下来只需要do(T=0)即可

Counterfactuals(cont.)

- In such a deterministic model, every assignment $U = u$ to the exogenous variables corresponds to a single member of, or “unit” in a population, or to a “situation” in nature.
- 反事实让因果推断从**群体层面**进化到了**个体层面**。
- 对于不确定的情况，只需要定义p(u)即可。
- 两种常见的反事实计算：
  - 假设和事实矛盾（高速公路还是城市公路）
  - 需要针对个体计算（让所有数学成绩不及格的学生假期作业翻倍会怎样？）

概念之间的联系

- SCM蕴涵了关于因果的全部信息。
- Causal Graph以及因果之梯的三个层次的分布都可以从SCM中推断出来，这个过程被称为因果推断。



Causal Hierarchy Theorem

- **Theorem** (Causal Hierarchy Theorem): With respect to Lebesgue measure over (a suitable encoding of L3-equivalence classes of) SCMs, the subset in which any PCH 'collapse' is measure zero.

- **Corollary**: To answer question at Layer i (about a certain interaction), one needs knowledge at layer i or higher.
- 换句话说，底层的信息几乎不能决定高层。

Causal Learning

- 在只给定消极的观察数据(passive observations)的情况下，不可能知道因果图的任何情况。

Proposition 4.1 (Non-uniqueness of graph structures) *For every joint distribution $P_{X,Y}$ of two real-valued variables, there is an SCM*

$$Y = f_Y(X, N_Y), \quad X \perp\!\!\!\perp N_Y,$$

where f_Y is a measurable function and N_Y is a real-valued noise variable.

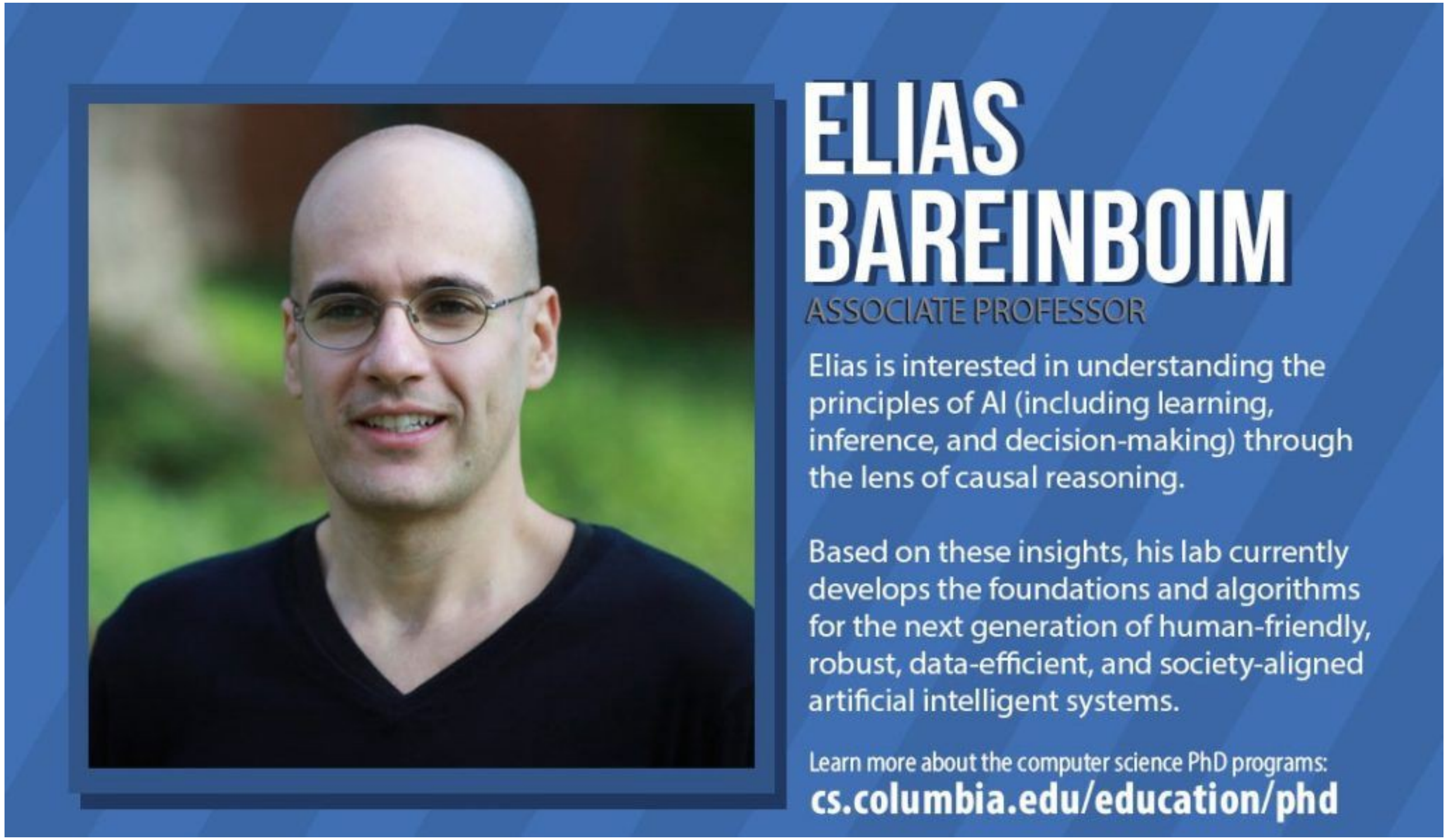
Discovering Causal Structure

- There are fundamentally three ways to get the DAG:
  - Prior knowledge
  - Guessing-and-testing
  - Discovery algorithms
- Guessing-and-testing主要根据DAG的两个性质：
  - 如果存在x指向y的路径，改变x应该会让y发生变化
  - 使用d-separates得到的，变量之间的条件独立的关系
- 根据DAG的性质可以开发出具体的算法，例如PC算法，fast PC算法等

[redacted]
complexity and credit assignment.

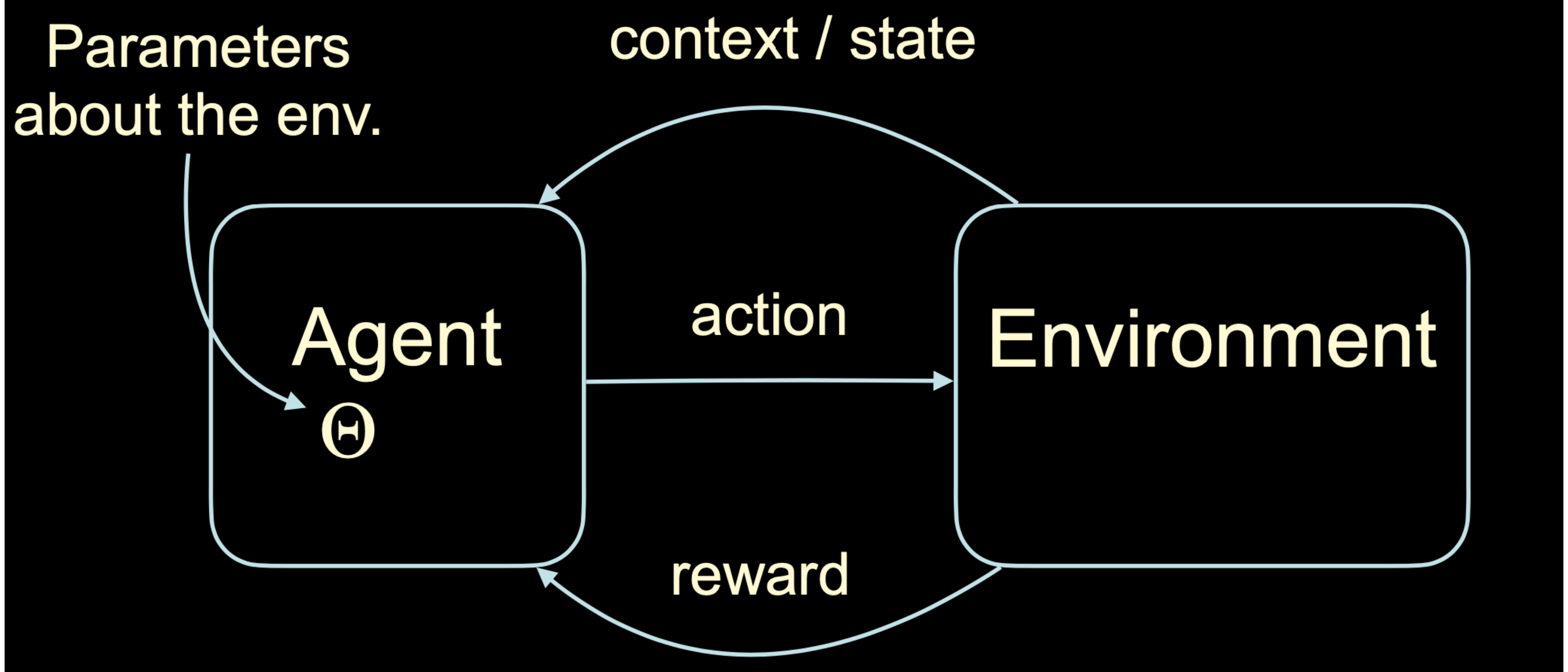
- Causal Inference(CI) is great at leveraging **structural invariances** across settings and conditions.
- **Causal RL = CI + RL**
- Our goal: Provide a cohesive framework that takes advantage of the capabilities of both formalisms, and that allows us to develop the next generation of AI systems.

Causal Reinforcement Learning(cont.)



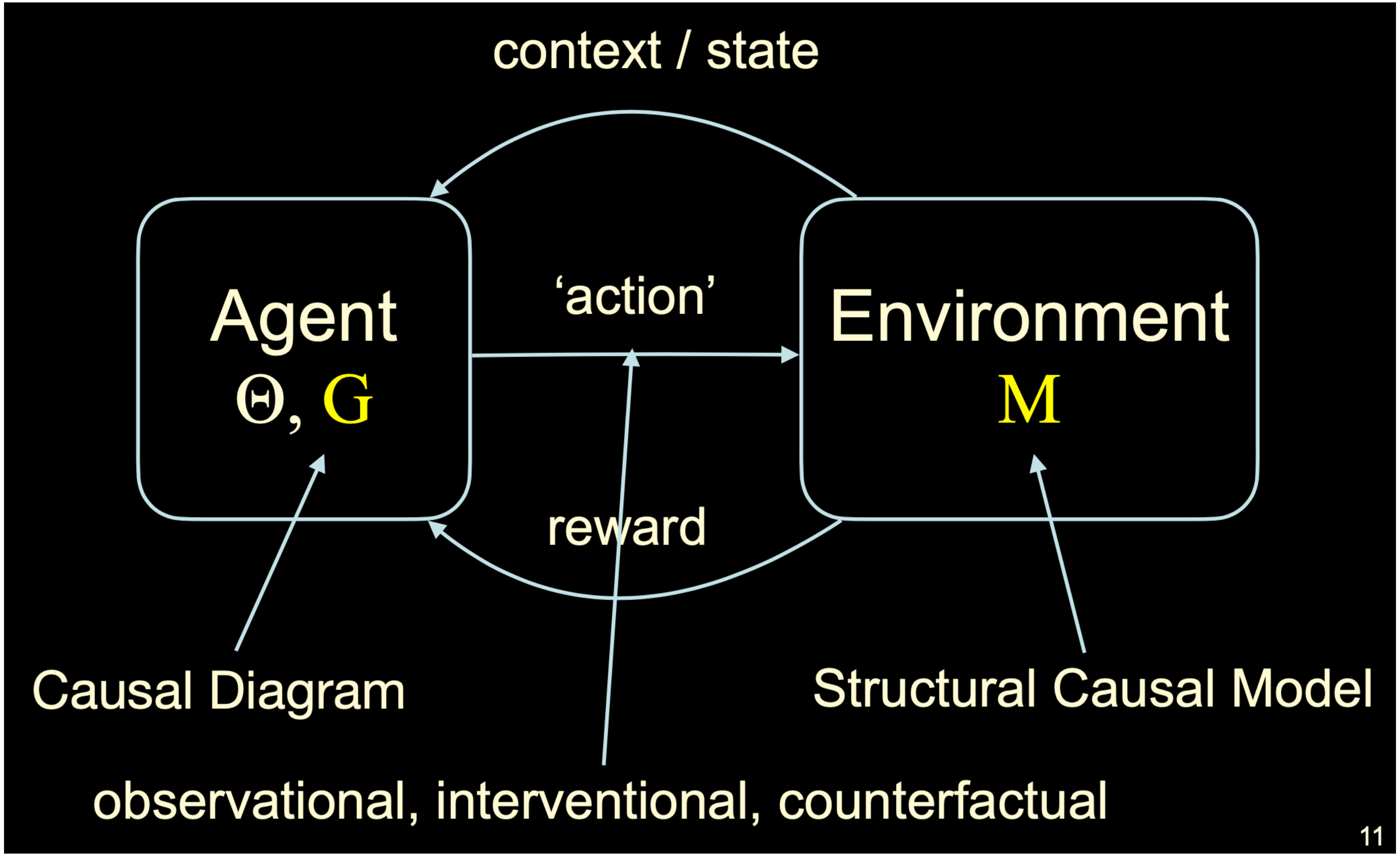
Towards Causal Reinforcement Learning(ICML, 2020)

RL - Big Picture



- Receive feedback in the form of rewards.
- Agent's utility is defined by the reward function.
- Must (learn to) act so as to maximize expected rewards.

Causal RL - Big Picture



RL and CI

- Goal of RL: Learn a policy π s.t. sequence of actions $\pi(.)$ = (X1, X2..., Xn) maximizes reward $E_{\pi}\,[Y|do(X)]$

Offline

1. Online learning ($\rightarrow$ do(x))
   - Agent performs experiments herself
   - Input: experiments $\{(do(X_i), Y_i)\}$; Learned: $P(Y \mid do(X))$
2. Off-policy learning (do(x) $\rightarrow$ do(x))
   - Agent learns from other agents' actions
   - Input: samples $\{(do(X_i), Y_i)\}$; Learned: $P(Y \mid do(X))$
3. Do-calculus learning (see(v) $\rightarrow$ do(x))
   - Agent observes other agents acting
   - Input: samples $\{(X_i, Y_i)\}$, G; Learned: $P(Y \mid do(X))$

Online Learning

- 强化学习本身就具备一定的因果推断能力，因为agent可以对环境进行干预

Explainable Reinforcement Learning Through a Causal Lens

Prashan Madumal, Tim Miller, Liz Sonenberg, Frank Vetere
The University of Melbourne
Victoria, Australia
pmathugama@student.unimelb.edu.au, {tmiller, l.sonenberg, f.vetere}@unimelb.edu.au

Online learning improved by CI

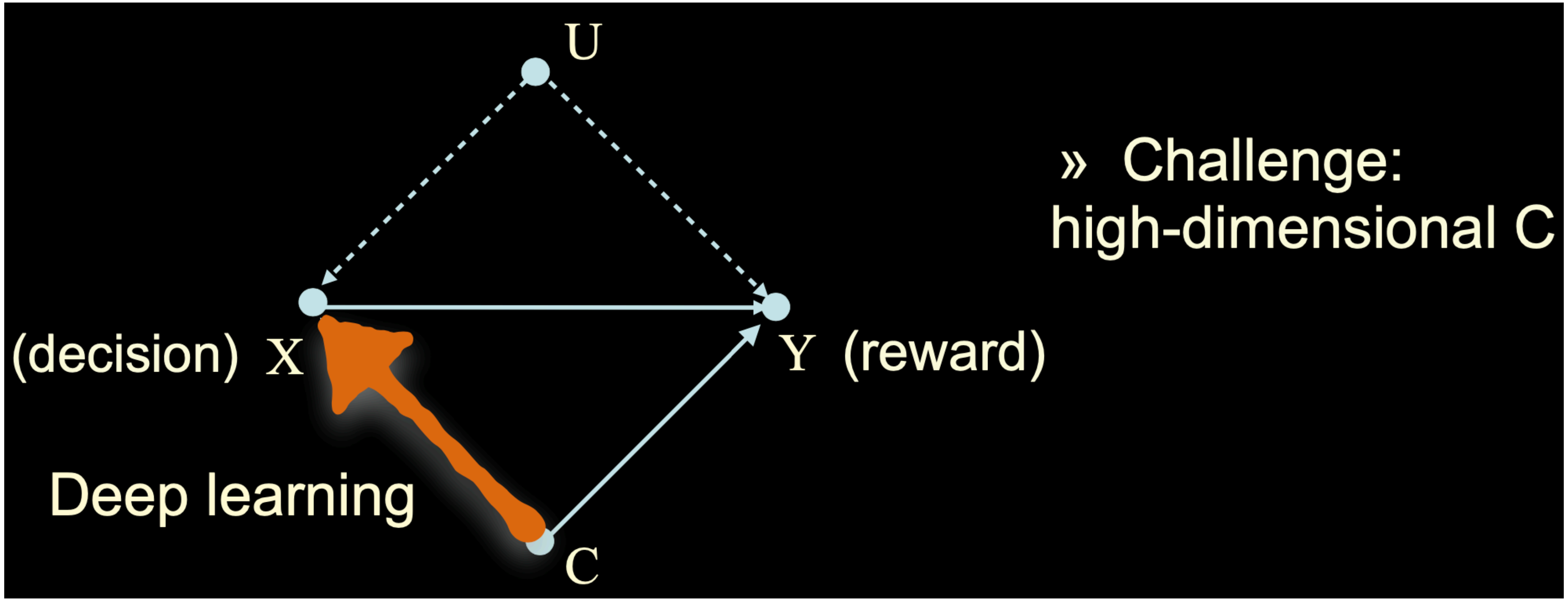
Near-Optimal Reinforcement Learning in Dynamic Treatment Regimes

Junzhe Zhang
Department of Computer Science
Columbia University
junzhez@cs.columbia.edu

Elias Bareinboim
Department of Computer Science
Columbia University
eb@cs.columbia.edu

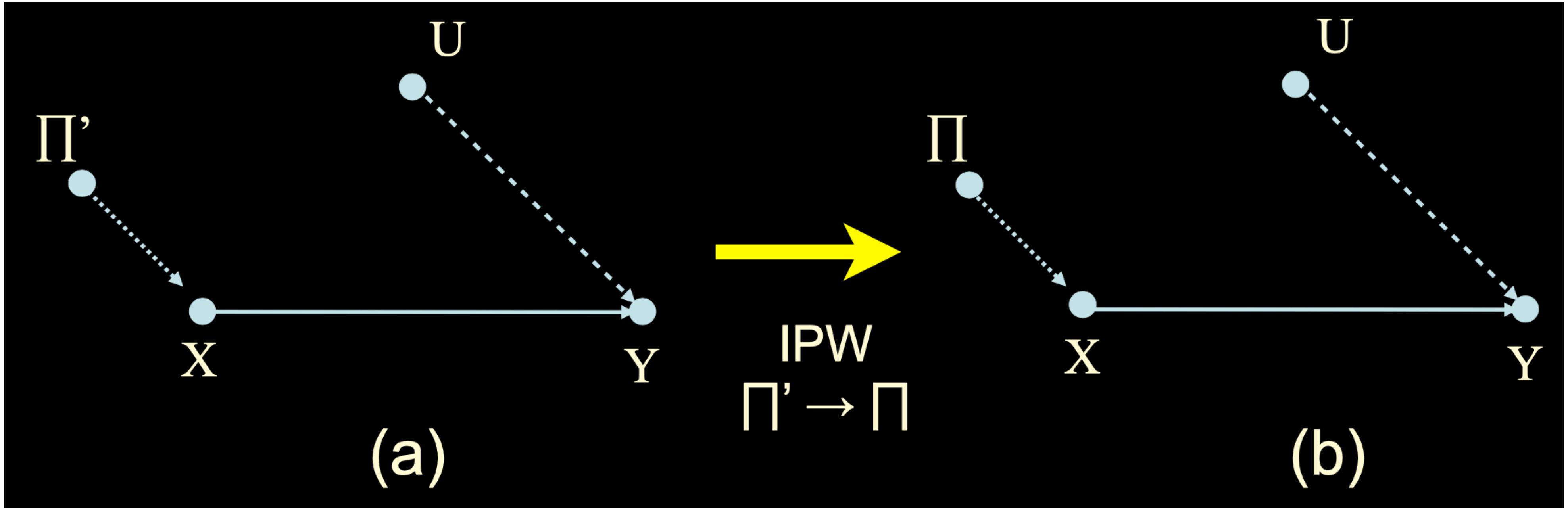
context-specific RL

- Context是从环境中提取的上下文，可以建立从context到decision的联系。



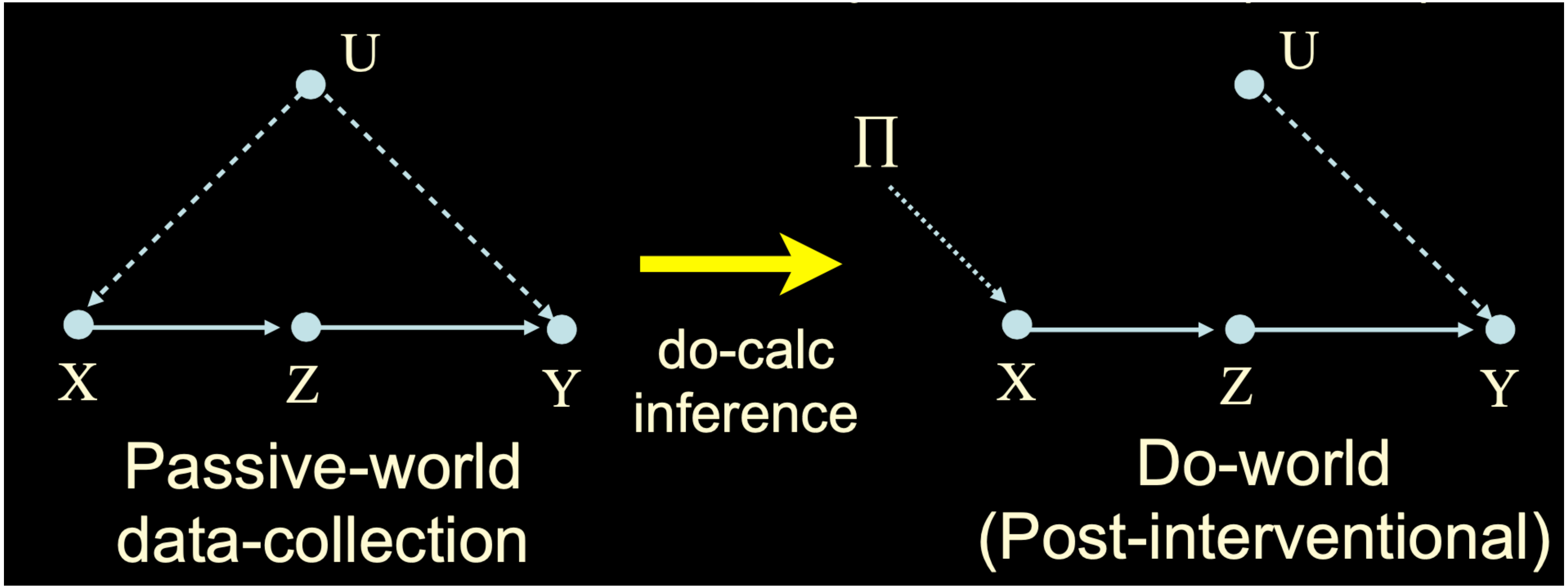
Off-Policy Learning

- 通过重要性采样IPW，用behavior策略采样的数据计算当前策略



Do-calculus learning

- E[Y | do(X)] can be estimated from non-experimental data (also called *natural / behavioral regime*)
- Do-calculus learning适用于off-line learning的setting



Challenge of CRL

1. Generalized Policy Learning (combining online + offline learning)
2. When and where to intervene? (refining the policy space)
3. Counterfactual Decision-Making (changing optimization function based on intentionality, free will, and autonomy)
4. Generalizability & robustness of causal claims (transportability & structural invariances)
5. Learning causal model by combining observations (L1) and experiments (L2)
6. Causal Imitation Learning