



图模型和因果推理基础

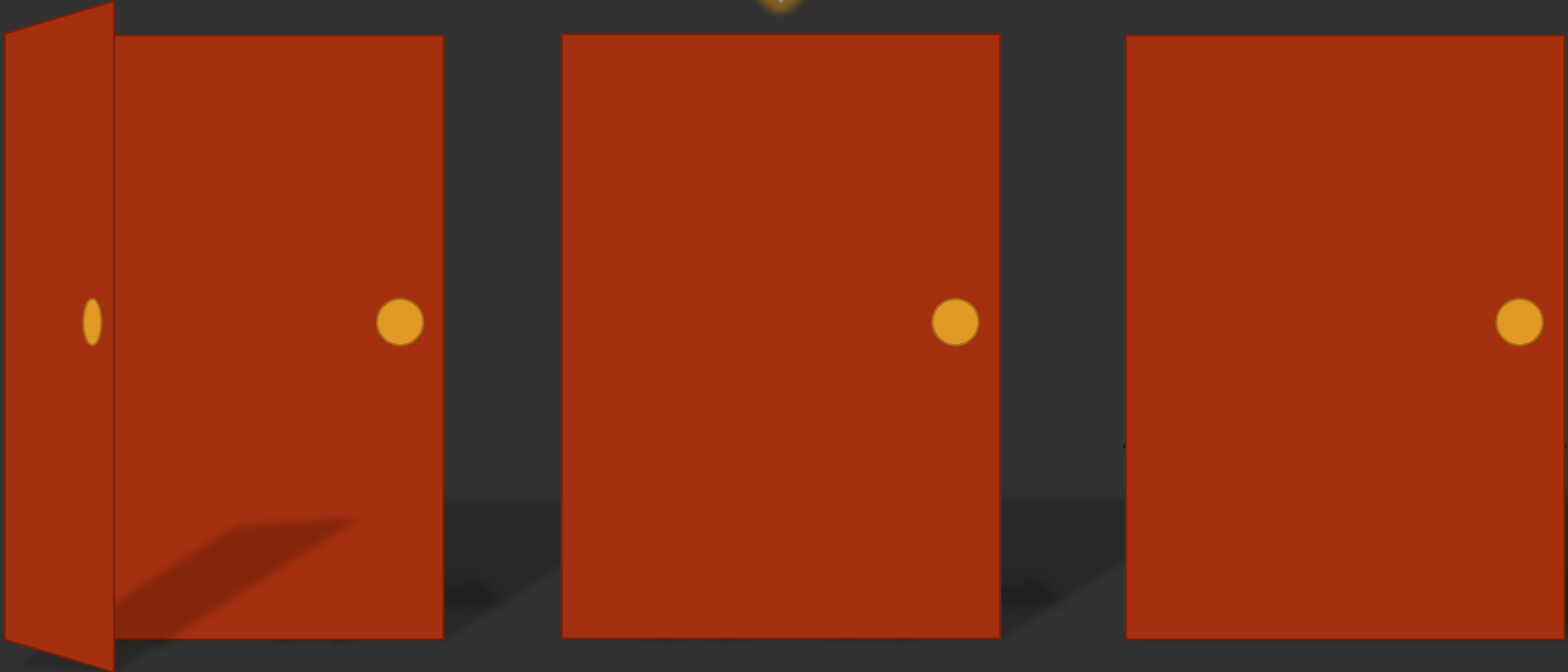
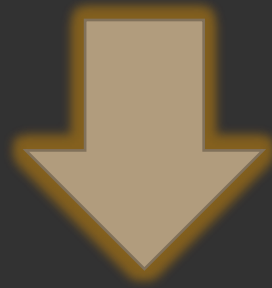
中国科学院计算技术研究所

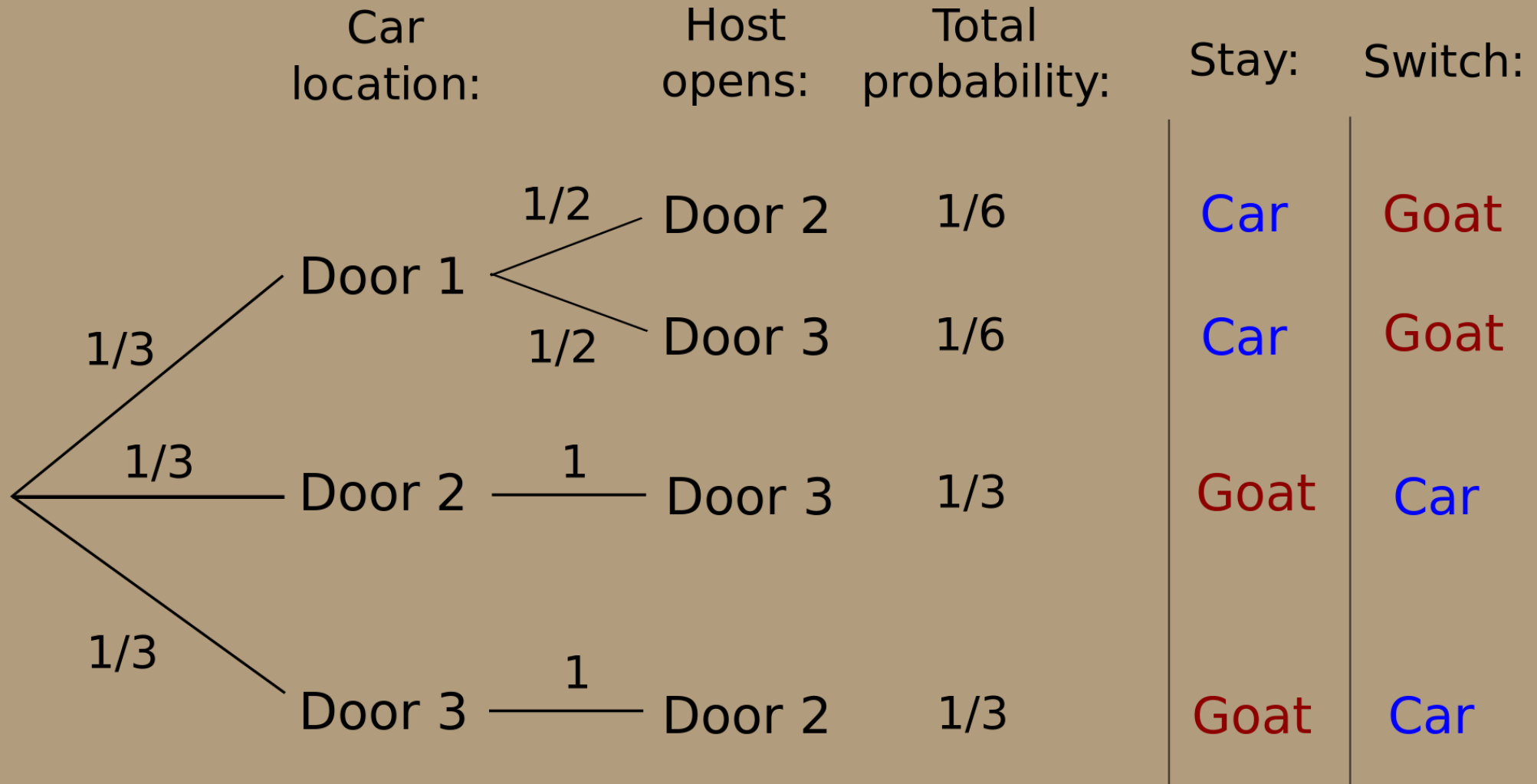
李奉治

因果科学与Causal AI读书会

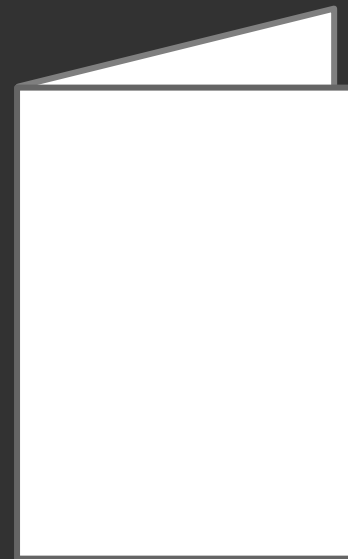
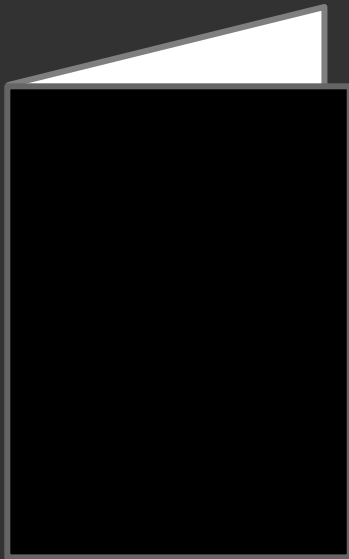
09/27/2020

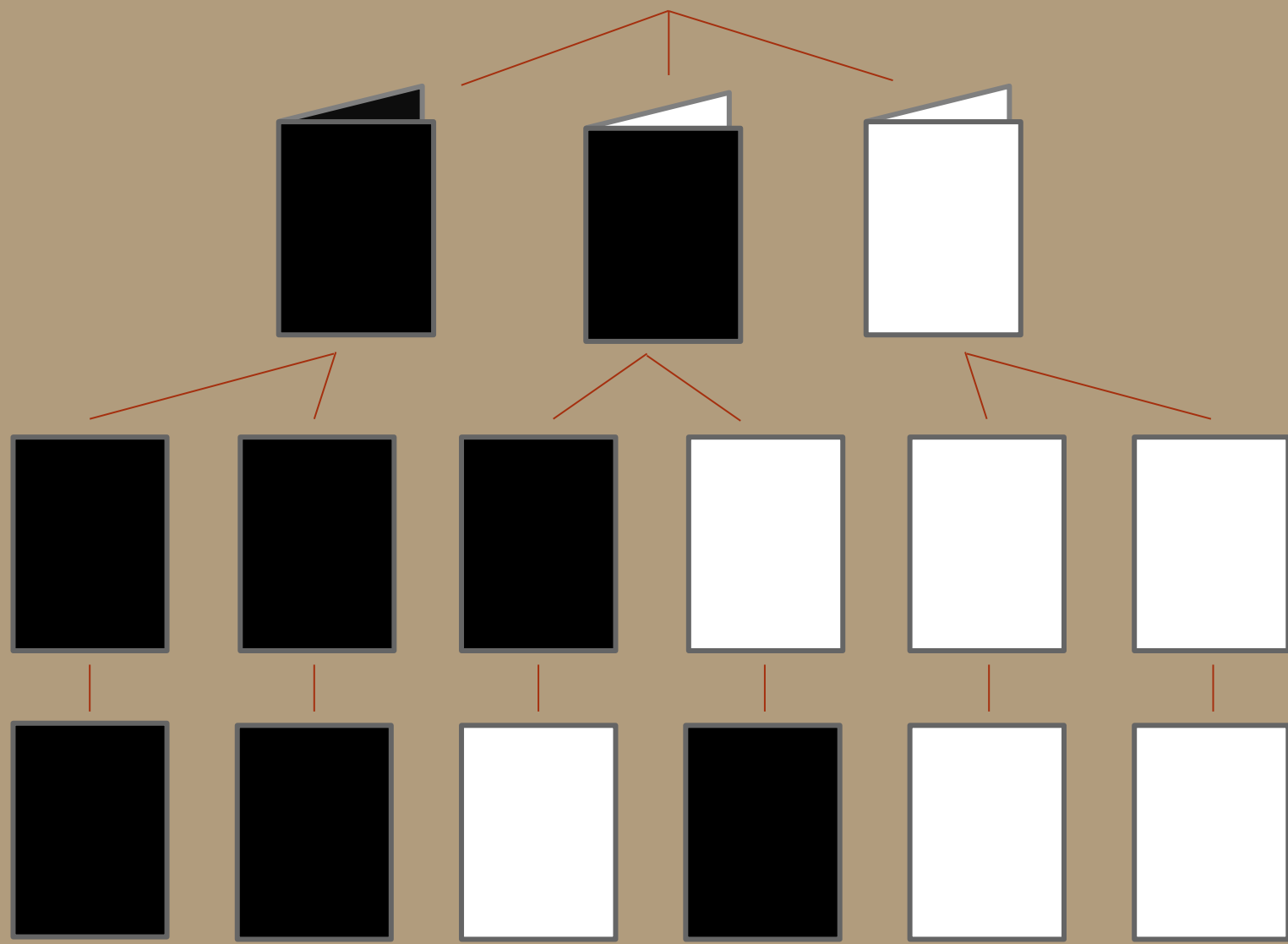
The Monty Hall Paradox



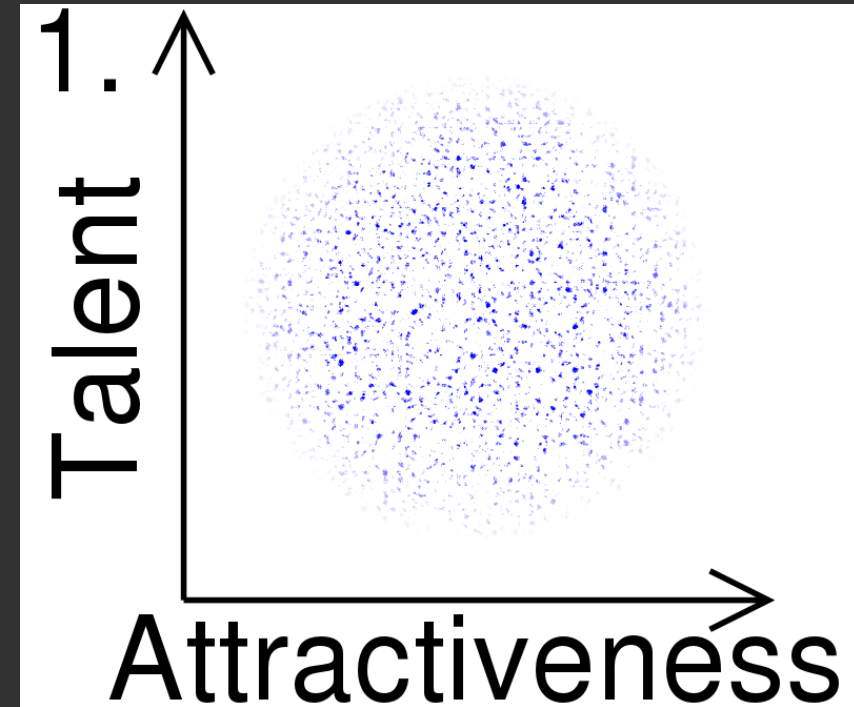
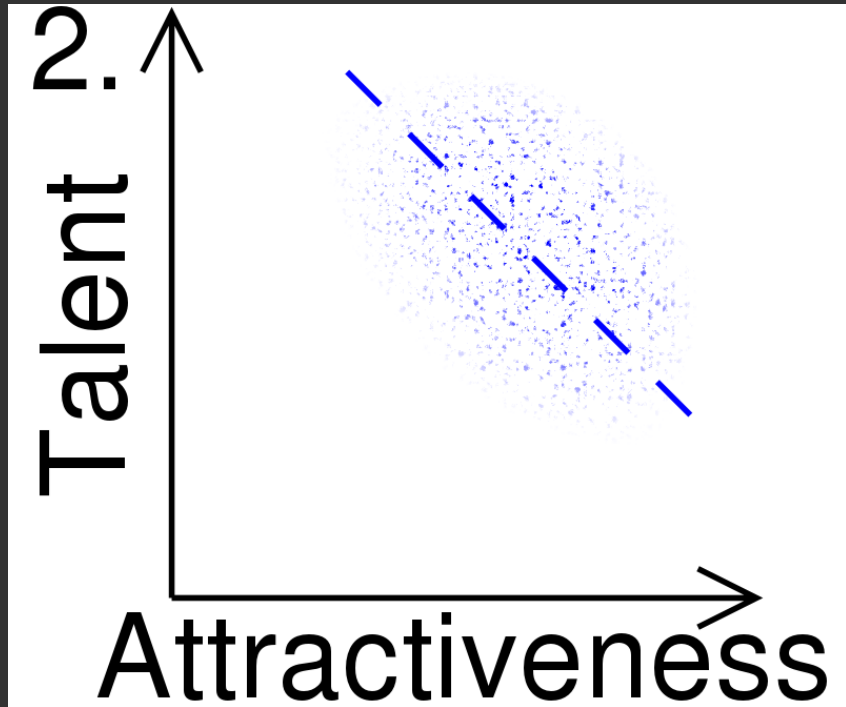


The Bertrand Paradox

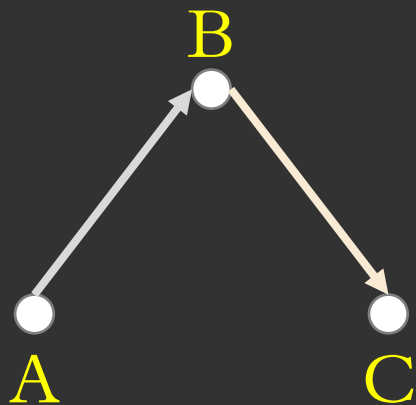




The Berkson's paradox

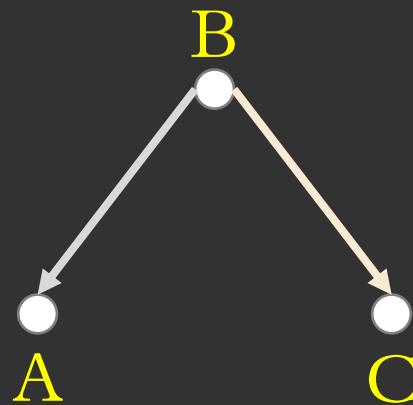


中介变量
Mediator



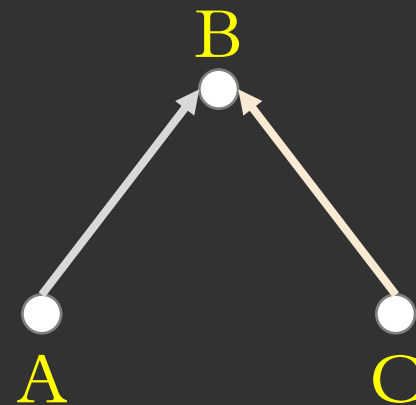
链接合
Chain

混杂因子
Confounder



叉接合
Fork

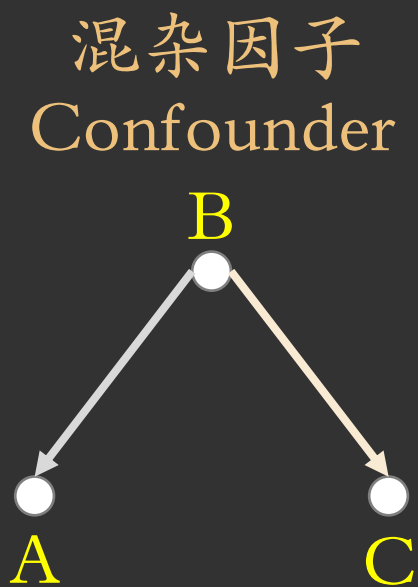
对撞因子
Collider



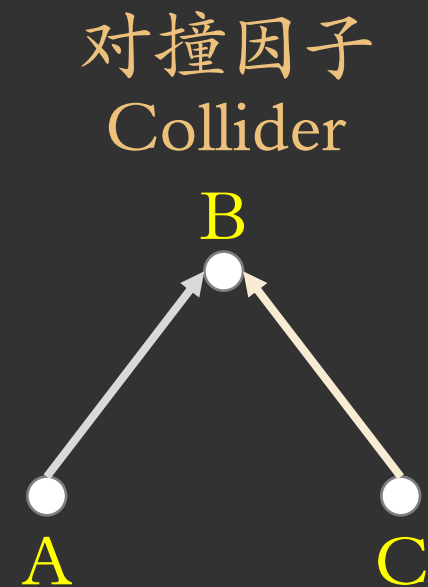
对撞接合
Collider



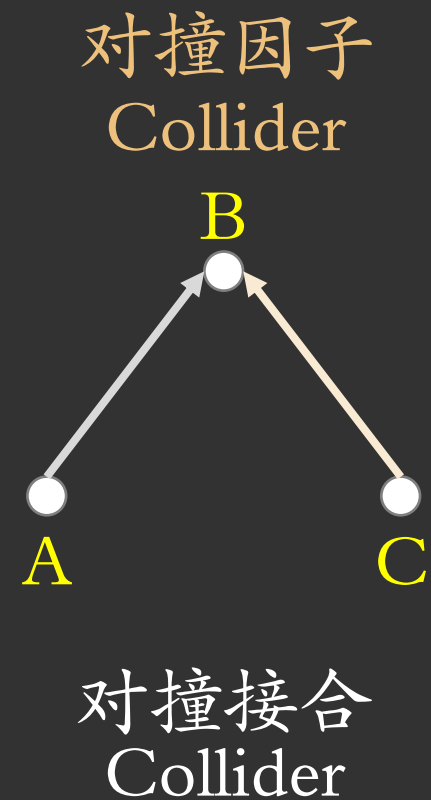
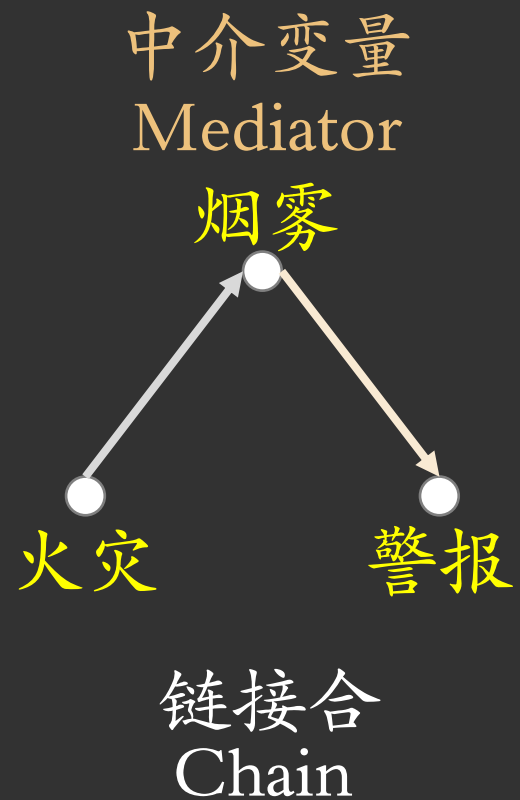
链接合
Chain

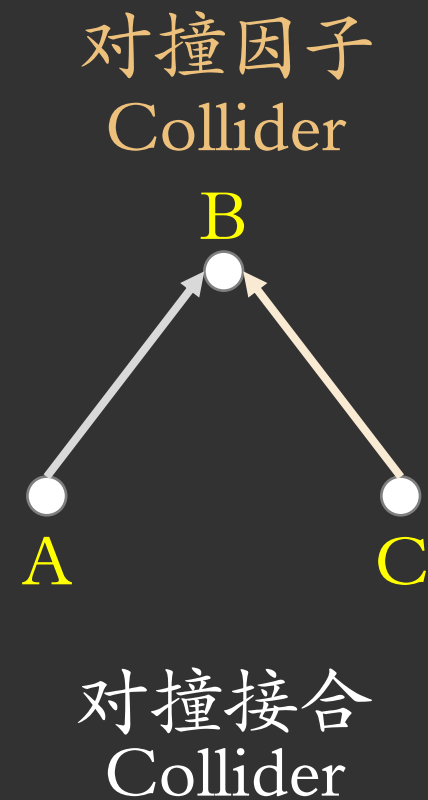
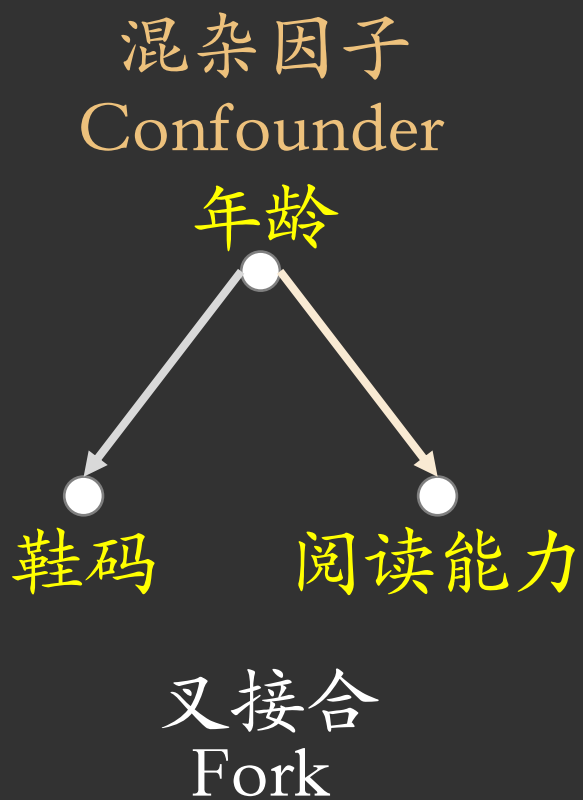
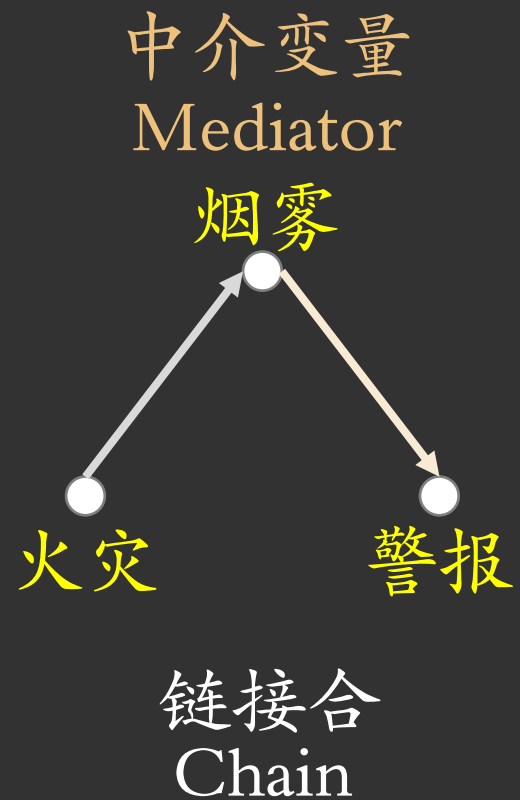


叉接合
Fork



对撞接合
Collider







链接合
Chain



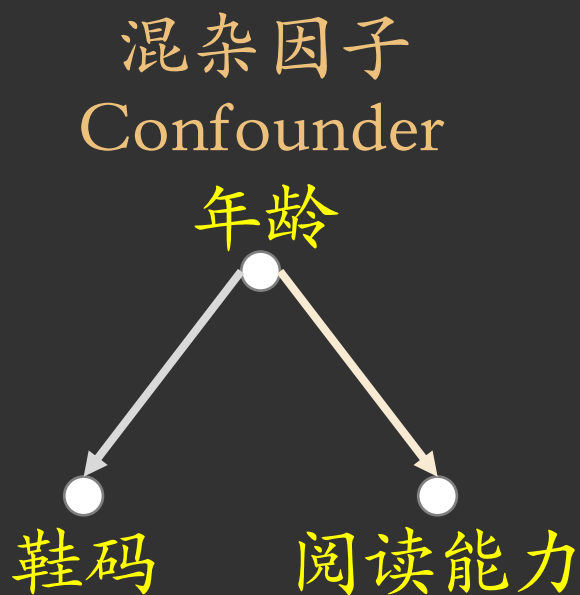
叉接合
Fork



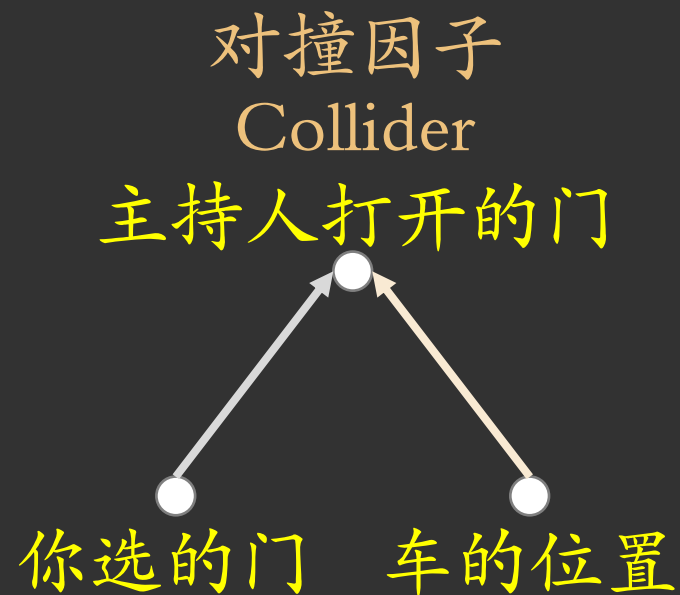
对撞接合
Collider



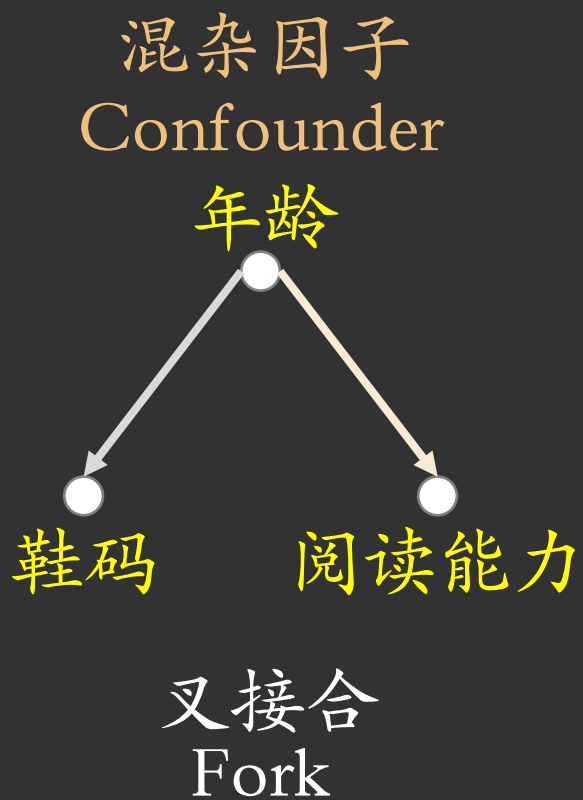
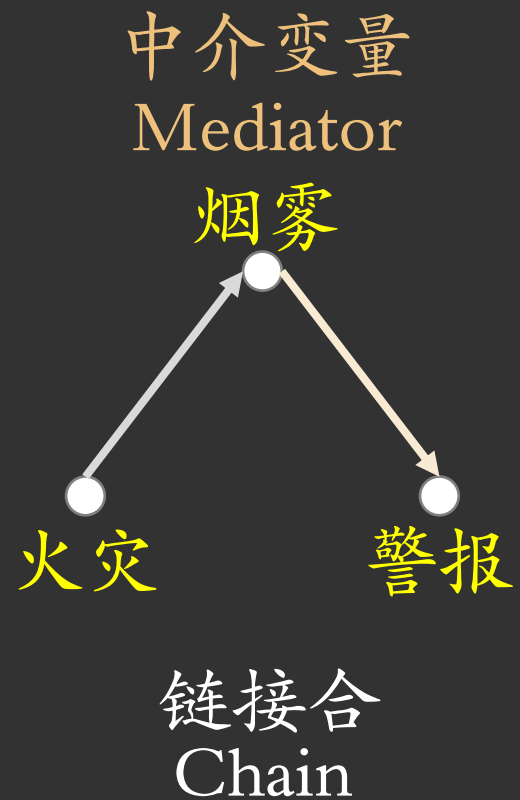
链接合
Chain

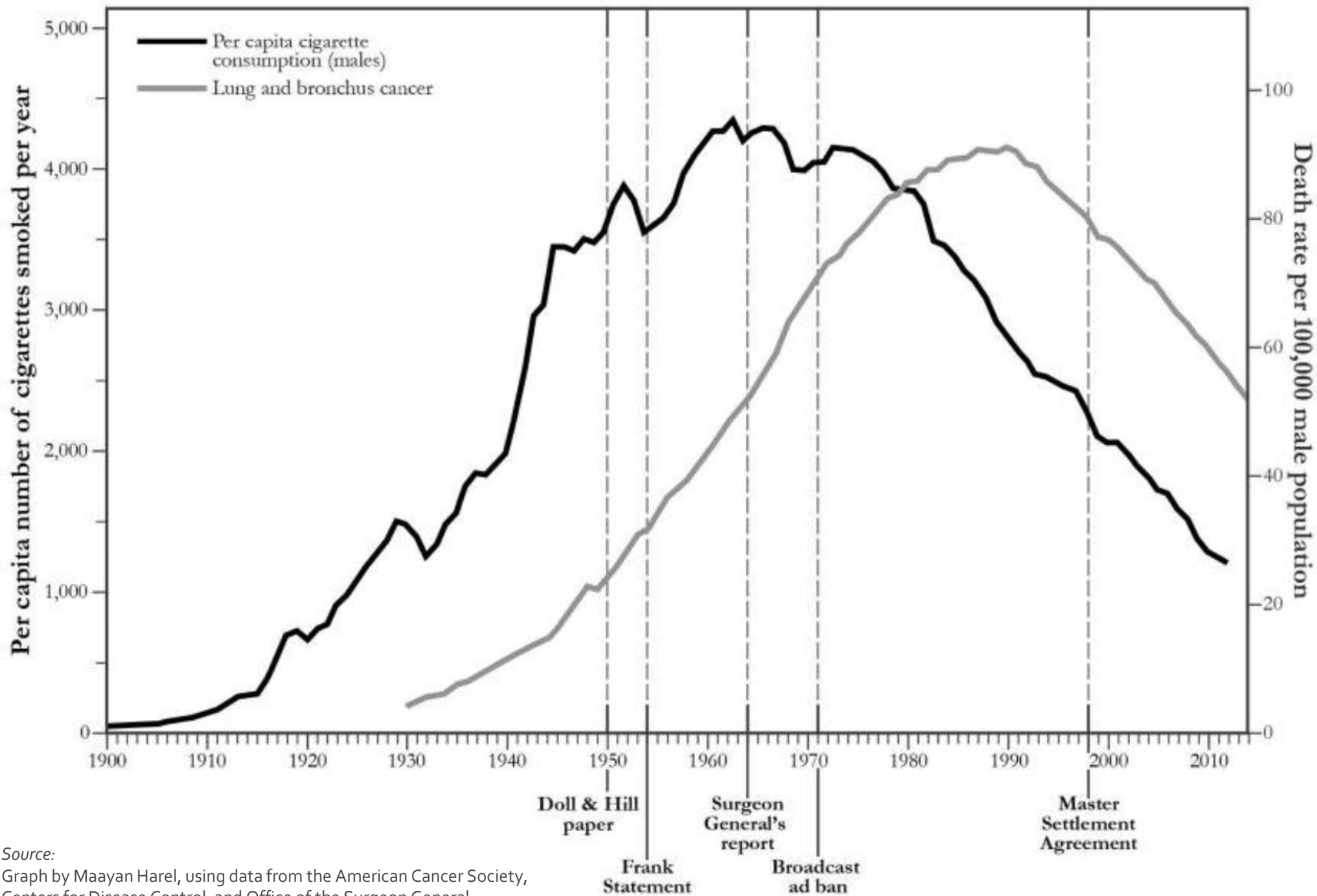


叉接合
Fork



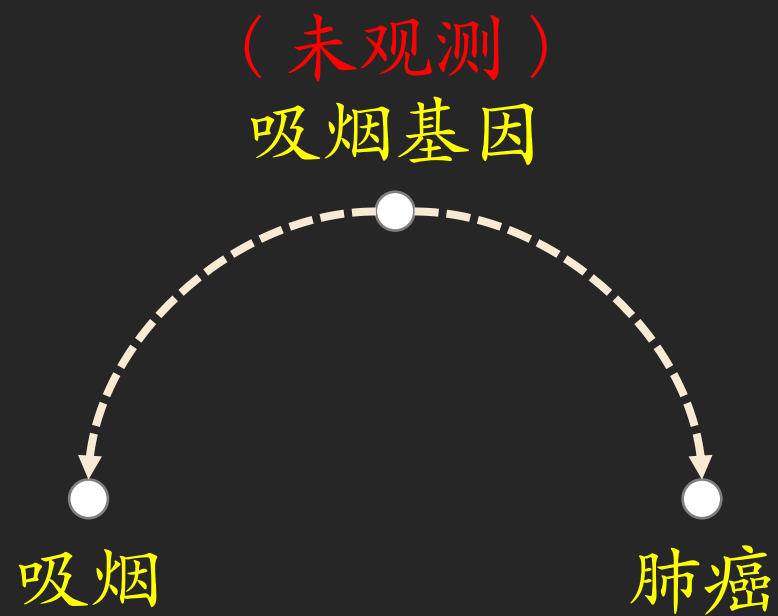
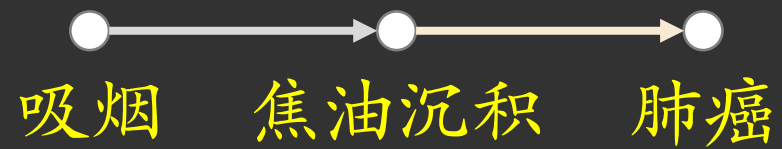
对撞接合
Collider





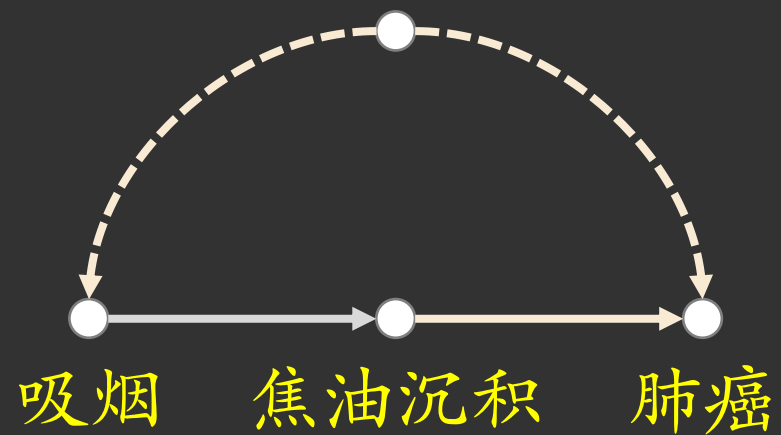
Source:

Graph by Maayan Harel, using data from the American Cancer Society, Centers for Disease Control, and Office of the Surgeon General.



(未观测)

吸烟基因, 心情, 工业化程度……



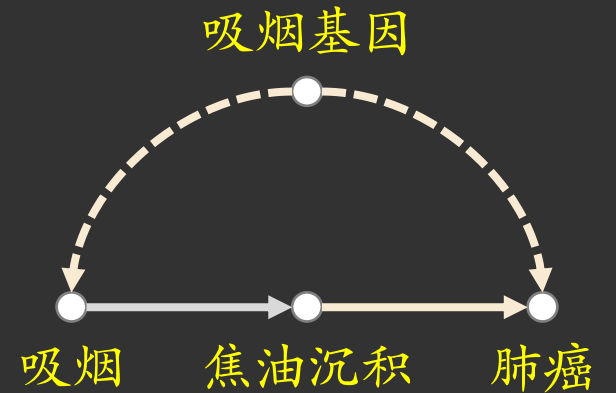
d-分离 (d-Separation)

令 X , Y 和 Z 是有向无环图 G 的结点集的三个不相交子集。若对于 X 和 Y 两个点集之间任意一条无向路径 p , 都有一个结点 w , 满足下方任意一条性质:

1. w 是 p 上的对撞因子, 且自身与其后代结点都不属于 Z

2. w 不是 p 上的对撞因子, 且属于 Z

则集合 Z 使 X 与 Y 之间在 G 上 d-分离, 记作 $(X \perp\!\!\!\perp Y \mid Z)_G$



概率因果模型 (Probabilistic Causal Model)

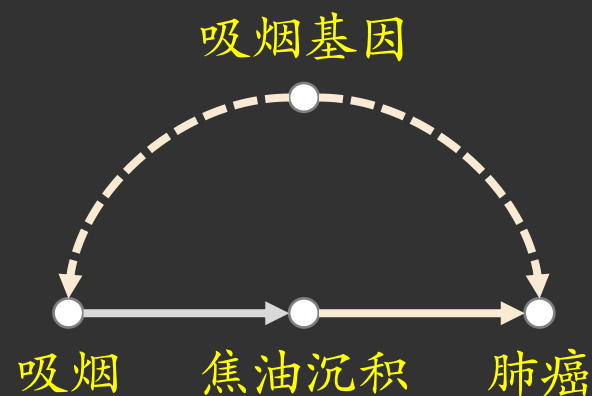
是一个四元组 $M = \langle U, V, F, P(U) \rangle$, 其中

1. U 是一组外生变量, 这些变量无法被观测或干预, 但是会影响到模型中的其他变量

2. V 是一组内生变量 $\{V_1, \dots, V_n\}$, 这些变量可以被观测, 其值依赖于 $U \cup V$ 的一个子集

3. F 是一组函数 $\{f_1, \dots, f_n\}$, 其中每个函数 f_i 是一个从 $U \cup V \setminus \{V_i\}$ 到变量 V_i 的映射, 即 $V_i = f_i(V_1, \dots, V_{i-1}, V_{i+1}, \dots, V_n, U_1, \dots, U_m)$

4. $P(U)$ 是在外生变量 U 上的一个联合概率分布

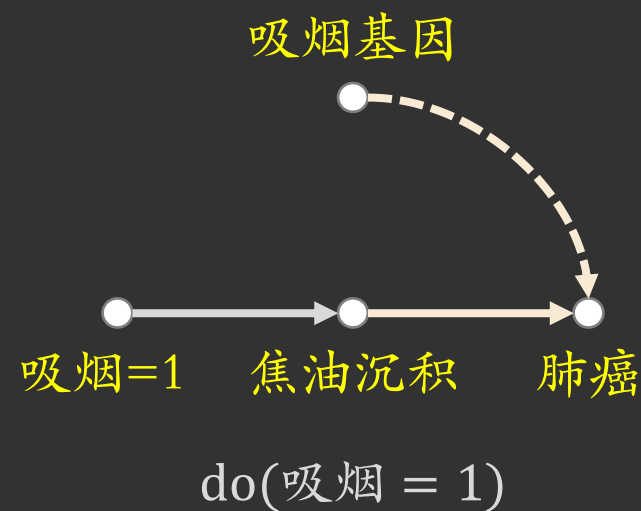


干预 (Intervention)

是对因果模型的一系列操作，使其中的一个或多个变量的观测值固定为特定值. 若对于模型

$M = \langle U, V, F, P(U) \rangle$ 中的变量集 X ($X \subseteq V$) 进行干预，使其值固定为 x ，可记为 $do(X = x)$

若要对 X 进行干预并考虑对其固定到各种不同值后对其他变量的影响时，会将此干预记为 $do(X)$



PART 1

Association & Causation

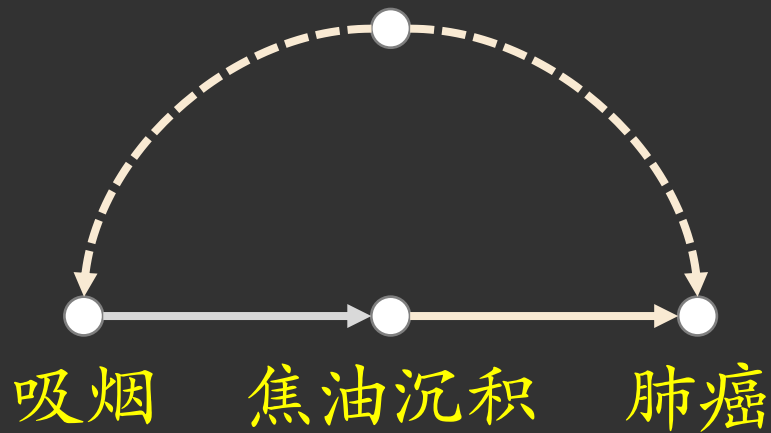
- Interesting Paradoxes Related to Causality
- Potential Outcomes

Graphical Models

- Related Notations & Terminology
- Do-Calculus & Intervention Graphs
- Structure Equation Models
- Comparison with Bayes Network

朱淑媛 清华大学工业工程系博士生，主要研究方向为行为运筹管理

(未观测)
吸烟基因



$$P(\text{癌症} \mid \text{吸烟}) == P(\text{癌症} \mid \text{do}(\text{吸烟}))$$

?

公理一 增添或删除观察

$$P(y \mid do(x), z, w) = P(y \mid do(x), w) \quad \text{if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\bar{X}}}$$

公理二 干预与观察交换

$$P(y \mid do(x), do(z), w) = P(y \mid do(x), z, w) \quad \text{if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\bar{X}\underline{Z}}}$$

公理三 增添或删除干预

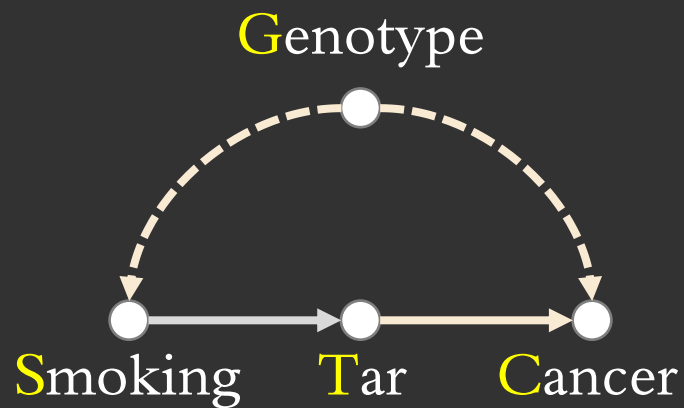
$$P(y \mid do(x), do(z), w) = P(y \mid do(x), w) \quad \text{if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\bar{X}, \overline{Z(W)}}}$$

PART 2

Identification

- Backdoor Adjustment with Examples
- Frontdoor Adjustment with Examples
- Instrumental Variable Analysis
- Do-calculus
 - Basic Notations
 - 3 Rules of Do-calculus
 - Examples to Show How It Works
 - Relation with Backdoor & Frontdoor & Instrumental Variable

付鑫玉 匹兹堡大学商学院管理信息系统博士生，研究方向是人机协作



$$P(c | do(s)) = \sum_t P(c | do(s), t) P(t | do(s))$$

概率公理

$$= \sum_t P(c | do(s), do(t)) P(t | do(s))$$

公理二

$$= \sum_t P(c | do(s), do(t)) P(t | s)$$

公理二

$$= \sum_t P(c | do(t)) P(t | s)$$

公理三

$$= \sum_{s'} \sum_t P(c | do(t), s') P(s' | do(t)) P(t | s)$$

概率公理

$$= \sum_{s'} \sum_t P(c | t, s') P(s' | do(t)) P(t | s)$$

公理二

$$= \sum_{s'} \sum_t P(c | t, s') P(s' | t) P(t | s)$$

公理三

PART 3

Identifiability

- Identification in causal inference engine
- Overview of identification methods
 - Definition
 - Approaches
- Scenarios of identification
 - Counterfactual responses to interventions that set variables to constants or via a known function of other variables
 - Counterfactual responses corresponding to path-specific effects
- Hidden variables & Latent edge
- Linear SEMs

徐培 就职于ThoughtWorks，为IT咨询和技术管理相关工作