

Introduction to Philosophy



Department of Philosophy
Central South University
xieshenlixi@163.com
[github](#)

August 2, 2026

Contents

Introduction

逻辑学

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

你有哪些感兴趣的“哲学”问题？

哲学问题

一只蝌蚪希望自己变成青蛙吗？

- ▶ 形而上学: 存在什么? (例如蝌蚪、物质的东西、心理状态、关系, 实体和关系谁更基本?)
- ▶ 认识论: 我们能知道什么? 如何知道? 我们能知道我们自己的/他人的/蝌蚪的心理吗?
- ▶ 心灵哲学: 什么是心理状态、心理过程? 一堆物质是否足以涌现出心理状态? 希望变青蛙是目标导向的 Agency 吗?
- ▶ 逻辑学: 我们应该如何思考? 决策?
- ▶ 伦理学: 我们 (不) 应该做什么? 我们是否有权让蝌蚪痛苦?
- ▶ 科学哲学: 蝌蚪希望变青蛙, 有因果解释吗? 可测量证伪吗? 什么是科学理论? 模型? 解释? 证据? 新概念的提出与新理论的构建之间是什么关系?
- ▶ 概念分析: 我们所说的 x 是什么意思? 说一只蝌蚪“希望” 是什么意思? AI“希望” 优化自己的目标函数吗?
- ▶ ...

带着问题学习

1. 下面哪种东西最“真实”——你所坐的椅子？构成这张椅子的分子？还是你坐在上面的感觉？
2. 你相信宇宙有目的吗？
3. 如何看待“事实”与“价值”的二分？
4. 怎么证明你不是“缸中之脑”？
5. 你相信万事皆有因吗？
6. 刚出生的婴儿是否具有自由意志？
7. 如果有一个你知道你一旦进去就快乐到不愿意出来的“快乐箱”，你愿意进去吗？
8. 什么使得昨天的你和今天的你是同一个“你”？
9. “真”和“美”是什么关系？
10. 你怎么看待“知识”和“信念”之间的关系？
11. 语词的意义能从纯粹的句法形式中涌现吗？
12. “科学”与“非科学”之间有明确的界限吗？
13. 你是否认为有原则上永远不可知的问题？
14. 有没有某种你愿意为之付出生命的东西？
15. 你是否认同“追求真理可以不计利害”？
16. 你在什么情况下会允许“撒谎”？

什么是哲学？

- ▶ 爱智慧
- ▶ 亚里士多德：哲学是研究“作为存在的存在”，以及属于它的本质属性的科学。
- ▶ 康德：哲学是理性批判的体系，它最终归结为回答四个问题：1. 我能知道什么？2. 我应该做什么？3. 我可以希望什么？4. 人是什么？
- ▶ 马克思：哲学是时代精神的精华。
- ▶ 维特根斯坦：哲学的目的在于思想的逻辑澄清。
- ▶ 罗素：哲学是介于神学与科学之间的无人区。
- ▶ 德勒兹：哲学是创造概念的艺术。
- ▶ 冯友兰：哲学就是对于人生有系统的反思。

哲学有哪些主要分支？

- ▶ 形而上学
- ▶ 认识论
- ▶ 逻辑学
- ▶ 伦理学
- ▶ 美学
- ▶ 科学哲学
- ▶ 心灵哲学
- ▶ 语言哲学
- ▶ 数学哲学
- ▶ 逻辑哲学
- ▶ 政治哲学
- ▶ 宗教哲学
- ▶ ...

Contents

Introduction

逻辑学

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

Contents

Introduction

逻辑学

Toy Logic

穆勒五法

类比论证

谬误与扯淡

形而上学

认识论

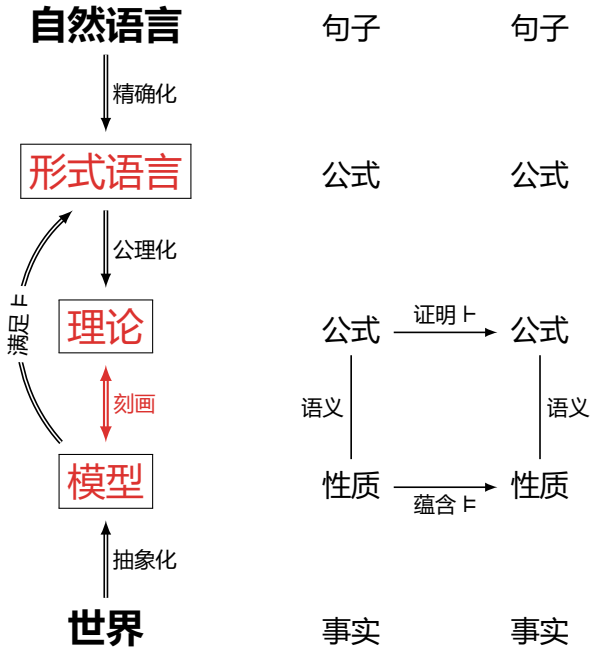
伦理学

心灵哲学

语言哲学

科学哲学

政治哲学



一个超级简单的 Toy Logic

► **语法:** 符号集 $\text{Var} = \{X, Y, \dots\}$, 公式 $\text{All}(X, Y)$

► **语义:** 结构 $\mathcal{M} := (M, \llbracket \cdot \rrbracket)$, 其中 $\llbracket X \rrbracket \subset M$

$$\mathcal{M} \models \text{All}(X, Y) \quad \text{当且仅当} \quad \llbracket X \rrbracket \subset \llbracket Y \rrbracket$$

► **逻辑蕴涵:** $\Gamma \models \varphi$ 当且仅当对任意结构 $\mathcal{M}$: 若 $\mathcal{M} \models \Gamma$ 则 $\mathcal{M} \models \varphi$

► **形式系统:**

$$\frac{\text{All}(X, Y) \quad \text{All}(Y, Z)}{\text{All}(X, Z)} \qquad \frac{}{\text{All}(X, X)}$$

► **证明:** $\Gamma \vdash \varphi$ 当且仅当存在一棵以 φ 为“根”节点的有穷“树”, 其每一节点或是公理或属于前提 Γ , 或通过推理规则由前面的节点生成

► **例子:** 怎么证明 $\text{All}(A, B), \text{All}(B, C), \text{All}(C, D) \vdash \text{All}(A, D)$

$$\frac{\frac{\text{All}(A, B) \quad \text{All}(B, C)}{\text{All}(A, C)} \quad \text{All}(C, D)}{\text{All}(A, D)}$$

► **元定理:** 能推出的都有效? 有效的都能推出? $\Gamma \vdash \varphi \iff \Gamma \models \varphi$

► **思考一下:** 怎么证明推不出? $\text{All}(A, B) \not\models \text{All}(B, A)$

形式语言

语义

模型

$\text{All}(A, B), \text{All}(B, C) \xrightarrow{\text{证明}} \text{All}(A, C)$

语义

语义

$\llbracket A \rrbracket \subset \llbracket B \rrbracket, \llbracket B \rrbracket \subset \llbracket C \rrbracket \xrightarrow{\text{蕴含}} \llbracket A \rrbracket \subset \llbracket C \rrbracket$

Theorem (Toy Logic 的可靠性定理)

$$\Gamma \vdash \varphi \implies \Gamma \models \varphi$$

Proof.

对证明树 $\Gamma \vdash \varphi$ 线性排列, 对其证明长度应用数学归纳法.

- ▶ 长度为 1: $\varphi \in \Gamma$ 或 φ 是公理 $\text{All}(X, X)$ 的特例. 显然 $\Gamma \models \varphi$.
- ▶ 假设证明长度不超过 n 时都有 $\Gamma \vdash \varphi \implies \Gamma \models \varphi$, 下证长度为 $n+1$ 时也成立.

第 $n+1$ 步或是公理或属于前提或由推理规则得到.

若由推理规则 $\frac{\text{All}(X, Y) \quad \text{All}(Y, Z)}{\text{All}(X, Z)}$ 得到, 由归纳假设, $\Gamma \models \text{All}(X, Y)$ 且 $\Gamma \models \text{All}(Y, Z)$.

即 $\llbracket X \rrbracket \subset \llbracket Y \rrbracket$ 且 $\llbracket Y \rrbracket \subset \llbracket Z \rrbracket$.

所以 $\llbracket X \rrbracket \subset \llbracket Z \rrbracket$, 即 $\Gamma \models \text{All}(X, Z)$.

Remark: 公理有效, 推理规则保真

Theorem (Toy Logic 的完备性定理)

$$\Gamma \models \varphi \implies \Gamma \vdash \varphi$$

Proof.

- 不妨证其逆否命题: 假设 $\Gamma \not\models \varphi$, 往证 $\Gamma \not\vdash \varphi$, 即, 找一个结构 $\mathcal{M}$, 使得 $\mathcal{M} \models \Gamma$ 但 $\mathcal{M} \not\models \varphi$.
- 令 $\mathcal{M}^\Gamma := (\mathcal{M}, \llbracket \cdot \rrbracket)$, 其中

$$\mathcal{M} := \text{Var}, \quad \llbracket X \rrbracket := \{Y : \Gamma \vdash \text{All}(Y, X)\}$$

1. 检查 $\mathcal{M}^\Gamma \models \Gamma$.

任给 $\text{All}(A, B) \in \Gamma$, 根据 $\llbracket A \rrbracket, \llbracket B \rrbracket$ 的定义和
$$\frac{\text{All}(Y, A) \quad \text{All}(A, B)}{\text{All}(Y, B)},$$
 可得 $Y \in \llbracket A \rrbracket \implies Y \in \llbracket B \rrbracket$, 即 $\mathcal{M}^\Gamma \models \text{All}(A, B)$.

2. 检查 $\mathcal{M}^\Gamma \not\models \varphi$. (往证: $\mathcal{M}^\Gamma \models \varphi \implies \Gamma \vdash \varphi$, 从而与假设 $\Gamma \not\models \varphi$ 矛盾)
假设 $\varphi = \text{All}(A, B)$, 若 $\mathcal{M}^\Gamma \models \varphi$, 则 $\llbracket A \rrbracket \subset \llbracket B \rrbracket$.
即, 对任意 $Y, Y \in \llbracket A \rrbracket \implies Y \in \llbracket B \rrbracket$.

因为 $\text{All}(A, A)$, 所以 $A \in \llbracket A \rrbracket$. 从而 $A \in \llbracket B \rrbracket$. 因此 $\Gamma \vdash \varphi$.

让语言表达力更丰富一点儿?

- ▶ **语法:** $All(X, Y)$, $Some(X, Y)$, $All(X, \bar{Y})$, $Some(X, \bar{Y})$, $Most(X, Y)$
- ▶ **语义:** 结构 $\mathcal{M} := (M, \llbracket \cdot \rrbracket)$, 其中 $\llbracket X \rrbracket \subset M$, $\llbracket \bar{X} \rrbracket = M \setminus \llbracket X \rrbracket$

$\mathcal{M} \models Some(X, Y)$ 当且仅当 $\llbracket X \rrbracket \cap \llbracket Y \rrbracket \neq \emptyset$

$\mathcal{M} \models Most(X, Y)$ 当且仅当 $|\llbracket X \rrbracket \cap \llbracket Y \rrbracket| > |\llbracket X \rrbracket \setminus \llbracket Y \rrbracket|$

- ▶ **形式系统:**

$$\frac{Some(X, Y)}{Some(Y, X)} \quad \frac{Some(X, Y)}{Some(X, X)} \quad \frac{All(X, \bar{Y})}{All(Y, \bar{X})}$$

$$\frac{Some(X, Y) \quad All(Y, Z)}{Some(X, Z)} \quad \frac{All(X, Y) \quad Some(X, Z)}{Some(Y, Z)}$$

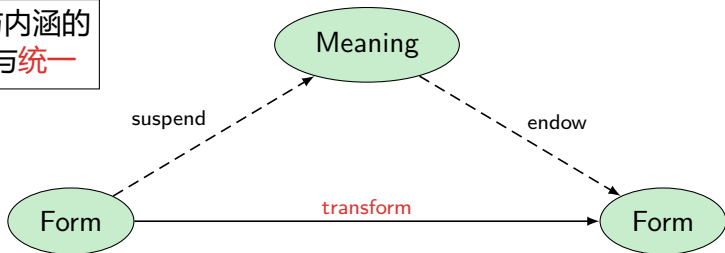
$$\frac{Most(X, Y)}{Some(Y, X)} \quad \frac{Some(X, X)}{Most(X, X)} \quad \frac{Most(X, Y) \quad All(Y, Z)}{Most(X, Z)}$$

$$\frac{Most(X, Z) \quad All(X, Y) \quad All(Y, X)}{Most(Y, Z)} \quad \frac{All(Y, X) \quad All(X, Z) \quad Most(Z, Y)}{Most(X, Y)}$$

逻辑求真

“真”为逻辑指引方向, 正如“美”为美学、“善”为伦理指引方向.
— 弗雷格

形式与内涵的
分离与统一



数理逻辑是先于一切科学的科学, 包含着位于一切科学底层的观念和原理.

— 哥德尔

一个概念是“逻辑”的, 当且仅当, 它对“论域”到自身的所有可能的——变换都保持不变.

— 塔斯基

思考一下

1. 什么是“语法”/“语义”?
2. 什么是“对象语言”/“元语言”?
3. 什么是“满足”?
4. 什么是“真”?
5. 什么是“有效命题”/“有效论证”?
6. 什么是“形式系统”?
7. 什么是“定理”?
8. 什么是“证明”?
9. 什么是“理论”?
10. 什么是“可 (有穷) 公理化”?
11. 什么是“一致”?
12. 什么是“可靠”?
13. 什么是“完备”?
14. 什么是“紧致”?
15. 什么是“定义”?
16. 什么是“表示”?
17. 什么是“归纳”/“递归”?
18. 什么是“有穷”/“无穷”?
19. 什么是“可数”/“不可数”?
20. 什么是“可判定”?
21. 什么是“计算复杂性”?
22. 什么是“表达力”?
23. 什么是“简洁性”?
24. 什么是“可解释性”?
25. 什么是“同态”?
26. 什么是“同构”?
27. 什么是“初等等价”?
28. 什么是“范畴性”?

Contents

Introduction

逻辑学

Toy Logic

穆勒五法

类比论证

谬误与扯淡

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

简单枚举归纳

- ▶ 我们都是瞎子.
- ▶ 吝啬的人是瞎子, 他只看见金子看不见财富.
- ▶ 挥霍的人是瞎子, 他只看见开端看不见结局.
- ▶ 卖弄风情的女人是瞎子, 她看不见自己脸上的皱纹.
- ▶ 有学问的人是瞎子, 他看不见自己的无知.
- ▶ 诚实的人是瞎子, 他看不见坏蛋.
- ▶ 坏蛋是瞎子, 他看不见上帝.
- ▶ 上帝也是瞎子, 他在创造世界的时候, 没有看到魔鬼也跟着混进来了.
- ▶ 我也是瞎子, 我只知道说啊说啊, 没有看到你们全都是聋子.

求同法

$$ABC \rightarrow xyz$$

$$ADE \rightarrow xuv$$

$$\hline A \rightarrow x$$

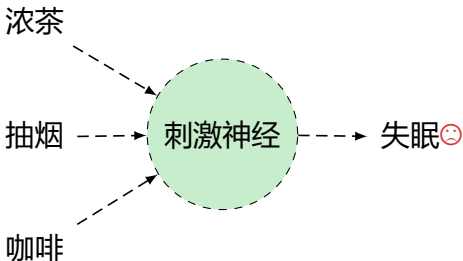
	缺碘	职业	年龄	甲状腺
张三	缺	工人	老年	肿大
李四	缺	农民	中年	肿大
王五	缺	士兵	青年	正常

缺碘 $\rightarrow$ 甲状腺肿大

甲状腺肿大人群的相同点是缺碘
缺碘是肿大的“必要”条件

泡脚	看小说	喝浓茶	失眠
泡脚	看电视	抽烟	失眠
泡脚	听音乐	喝咖啡	失眠

泡脚 $\rightarrow$ 失眠



求异法

$$ABC \rightarrow xyz$$

$$\overline{ABC} \rightarrow \overline{x}yz$$

$$\hline A \rightarrow x$$

冰激凌	沙拉	面包	牛奶	肚子疼
酸奶	沙拉	面包	牛奶	肚子不疼

冰激凌 $\rightarrow$ 肚子疼

吃了冰激凌的人肚子疼
肚子不疼的人没吃冰激凌
冰激凌是肚子疼的“充分”条件

上课	头疼
不上课	头不疼
上课 $\rightarrow$ 头疼	



取健康的蜘蛛, 大吼一声, 蜘蛛被吓跑了
砍掉蜘蛛的腿, 大吼一声, 蜘蛛纹丝不动

蜘蛛的听觉器官长在腿上. 😞

秋末冬初街道旁的响叶杨纷纷落叶, 但高压水银灯下的响叶杨却迟迟不落叶, 因此, 高压水银灯照射可能是响叶杨落叶迟的原因.

求同求异共用法

$$ABC \rightarrow xyz \quad ABC \rightarrow xyz$$

$$ADE \rightarrow xuv \quad \overline{ABC} \rightarrow \overline{xyz}$$

$$A \rightarrow x$$

动物	纲	环境	形态
鲨鱼	鱼纲	海	
鱼龙	爬行	海	
鲸鱼	哺乳	海	
鼯鼠	哺乳	陆	
蝙蝠	哺乳	空	

达尔文: 环境 $\rightarrow$ 形态

	咖啡	嗦粉	熬夜	看书	肚子疼
张三	●	●	●	●	●
李四	●		●		●
王五		●		●	
赵六			●	●	

剩余法

$$\begin{array}{lcl} ABC \rightarrow xyz & & ABC \rightarrow x \\ B \rightarrow y & & B \not\rightarrow x \\ C \rightarrow z & & C \not\rightarrow x \\ \hline A \rightarrow x & & A \rightarrow x \end{array}$$

受其他行星吸引, 天王星运行轨道上有四个地方发生偏斜
三个地方偏斜是已知行星吸引的结果

剩余一个地方偏斜是未知行星吸引的结果 (发现海王星)

	咖啡	嗦粉	熬夜	看书	疲惫	肚子疼
张三	•	•	•	•	•	•
李四	•		•		•	•
王五		•		•		
赵六			•	•	•	

Remark: 预设了一个原因确定一个结果; 原因彼此独立; 结果彼此独立;
所有可能的原因已知; 所有其它结果的原因已知.

共变法

$$ABC \rightarrow xyz$$

$$A^{\uparrow}BC \rightarrow x^{\uparrow}yz$$

$$A^{\downarrow}BC \rightarrow x^{\downarrow}yz$$

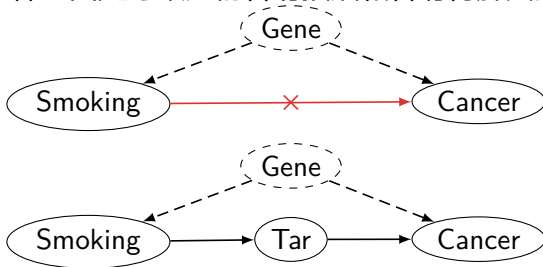
$$\hline A \rightarrow x$$

Example

热胀冷缩、体温表

禁烟人士 抽烟越多越容易患肺癌，所以抽烟是导致肺癌的重要原因。

烟草公司 某种基因是导致人们容易抽烟和容易得肺癌的共同原因。



Simpson's Paradox

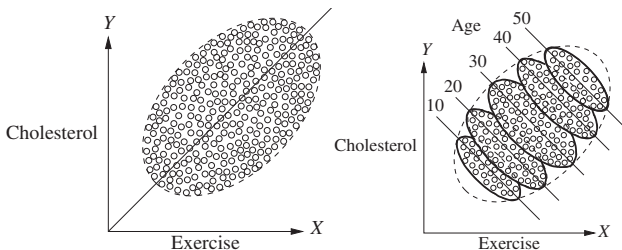
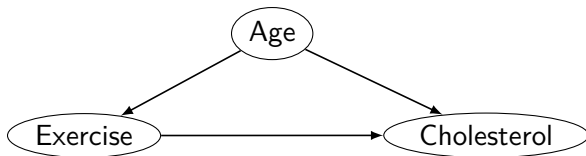


Figure: Exercise appears to be beneficial in each age group but harmful in the population as a whole. — Older people exercise more.



$$\mathbb{E}[\text{cholesterol} \mid \text{exercise}] > \mathbb{E}[\text{cholesterol} \mid \text{no exercise}]$$

$$\mathbb{E}[\text{cholesterol} \mid \text{do}(\text{exercise})] < \mathbb{E}[\text{cholesterol} \mid \text{do}(\text{no exercise})]$$

Contents

Introduction

逻辑学

Toy Logic

穆勒五法

类比论证

谬误与扯淡

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

Example

古龙《九月鹰飞》

- ▶ 叶开忍不住又道：“你为什么还是戴着这草帽？”
- ▶ 墨九星道：“因为外面有狗在叫。”
- ▶ 叶开怔了怔，道：“外面有狗叫，跟你戴草帽又有什么关系？”
- ▶ 墨九星冷冷道：“我戴不戴草帽，跟你又有什么关系？”

Example

庄子《齐物论》

不知周之梦为蝴蝶与，蝴蝶之梦为周与？

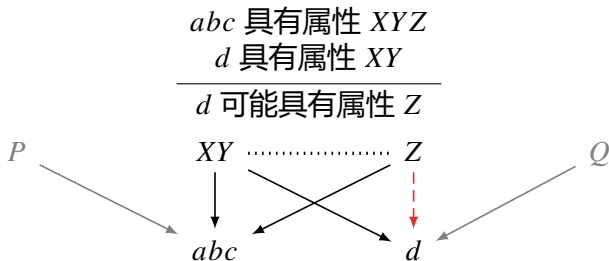


神秀 vs 慧能

神秀：身是菩提树，心如明镜台，时时勤拂拭，勿使惹尘埃。

慧能：菩提本无树，明镜亦非台，本来无一物，何处染尘埃？

类比论证

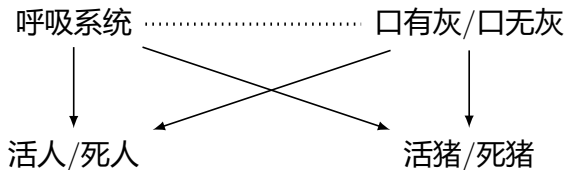


1. 类比物 abc 与目标物 d 的相似属性 XY 越多, 论证越强.
2. 相似属性 XY 与推出属性 Z 之间越相关, 论证越强.
3. 类比物 abc 与目标物 d 的相似属性 XY 越本质, 论证越强.
4. 类比物 abc 数量越多, 论证越强.
5. 类比物 abc 的差异性越大, 论证越强.
6. 类比物 abc 与目标物 d 的非相似属性 PQ 与 XYZ 的相关程度, 可能削弱或加强论证.
7. 结论越具体, 论证越弱.

Example

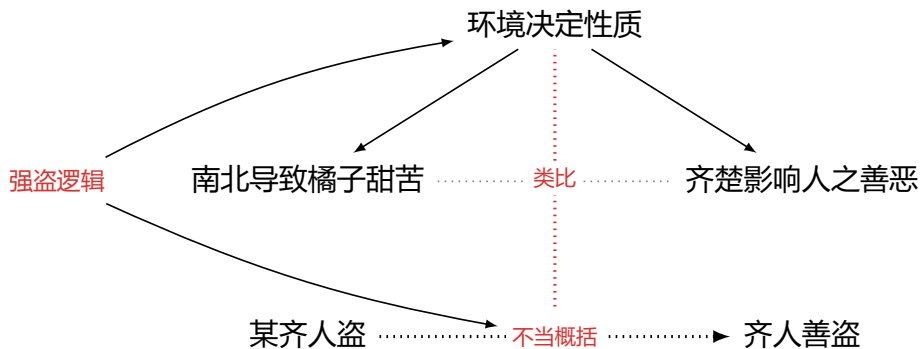
Argument by Analogy

有妻杀夫, 放火烧舍, 称“火烧夫死”. 夫家疑之, 讼于官. 妻不服. 取猪两头, 杀其一. 积薪焚之, 活者口中有灰, 杀者口中无灰. 因验尸, 口果无灰, 鞠之服罪.



Refutation by Analogy 《晏子使楚》

1. 楚王赐晏子酒, 酒酣, 吏二缚一人诣王.
2. 王曰: “缚者曷为者也?” 对曰: “齐人也, 坐盗.”
3. 王视晏子曰: “齐人固善盗乎?”
4. 晏子避席对曰: “婴闻之, 橘生淮南则为橘, 生于淮北则为枳, 叶徒相似, 其实味不同. 所以然者何? 水土异也. 今民生齐不盗, 入楚则盗, 得无楚之水土, 使民善盗耶?”



Example

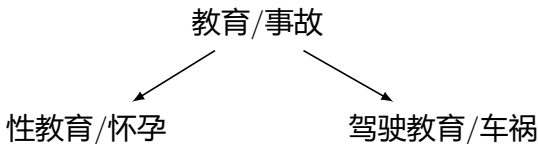
Argument by Analogy

人们似乎经常相信创造力, 但它所做的只不过是把对象的分界线确定下来, 并赋予它一个名字. 正如地理学家划出海岸线并说“这些线确定的海域为黄海”, 此时他并未创造一个海; 数学家也一样, 他不能通过定义创造东西.

— 弗雷格

Refutation by Analogy

- ▶ 性教育导致怀孕.
- ▶ 是的, 正如驾驶教育导致车祸. 😞



Example

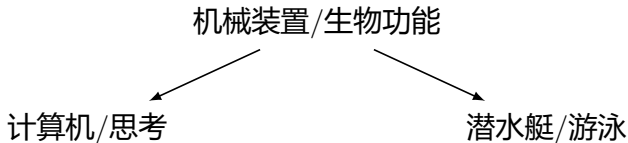
Argument by Analogy

不能要求每样东西都有定义, 否则如同要求任何物质都可被分解. 简单物质不能被分解, 逻辑上简单的东西不能被定义.

— 弗雷格

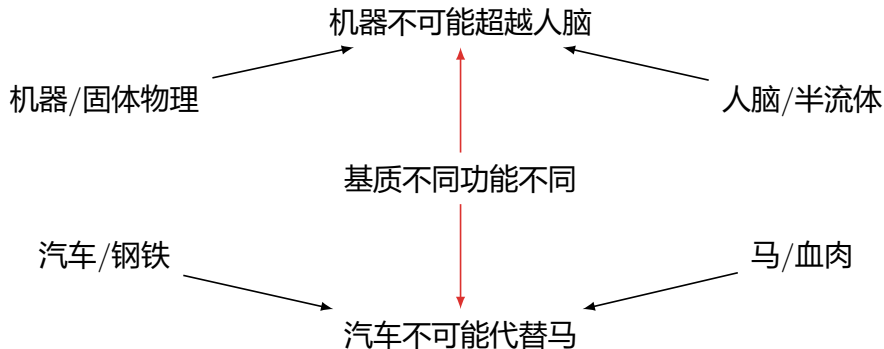
Refutation by Analogy

- ▶ 计算机会思考吗?
- ▶ 潜水艇会游泳吗?



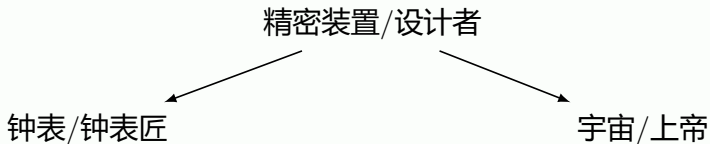
Refutation by Analogy

- ▶ 机器不可能超越人脑, 因为, 机器是建立在固体物理学之上的, 而人脑是一个活的半流体系统.
- ▶ 汽车不可能代替马, 因为, 汽车是钢铁铸成的, 而马是活的血肉构成的有机体.

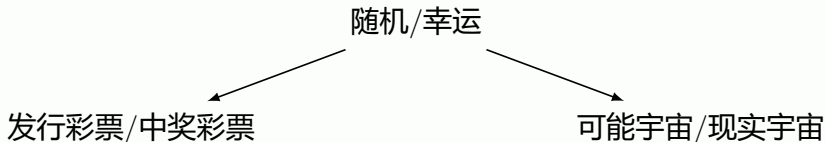


Examples — “上帝存在吗?”

Argument by Analogy



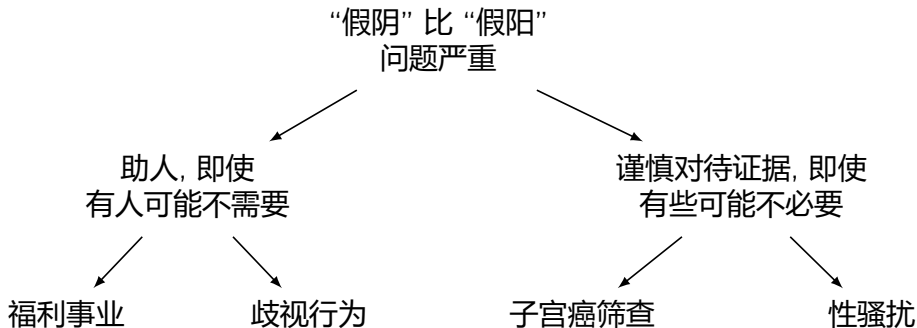
Refutation by Analogy



Refutation by Analogy

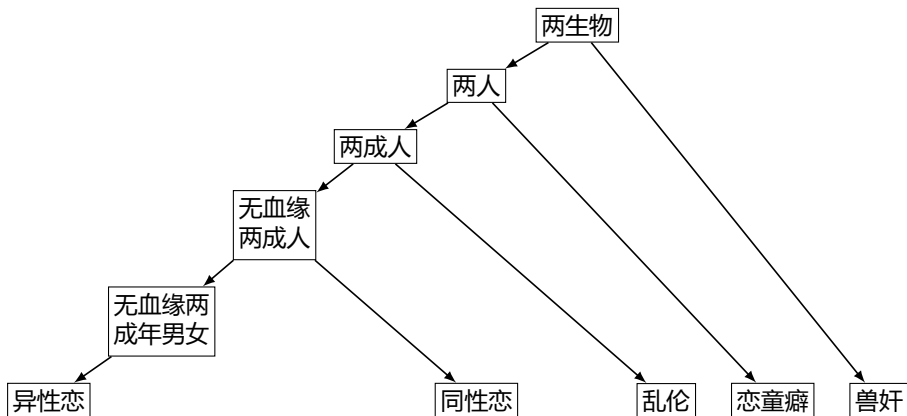
- ▶ Everything must have a cause. The world must have a cause.
- ▶ Everybody has a mother. Does the human race have a mother?

抽象 — “类比层级” 的提升



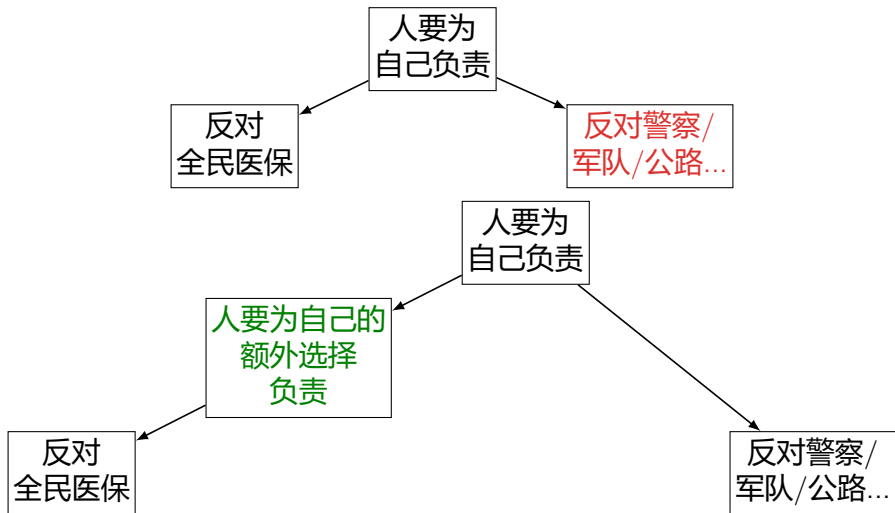
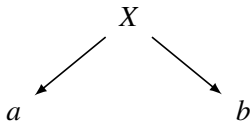
选择正确的“类比层级”

类比就像桥梁，可以带我们去任何地方，在选择走哪座桥时要非常小心。

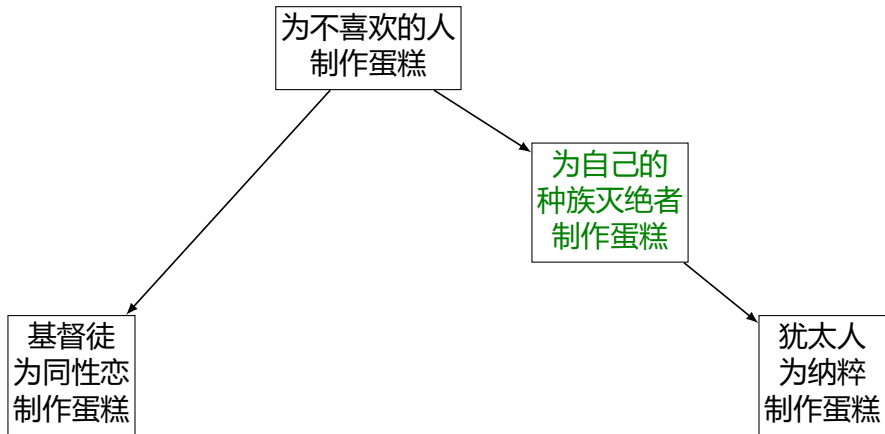


“你竟然接受同性恋，是不是你还接受乱伦？”

作为桥梁的类比层级 X 有时是隐藏的.



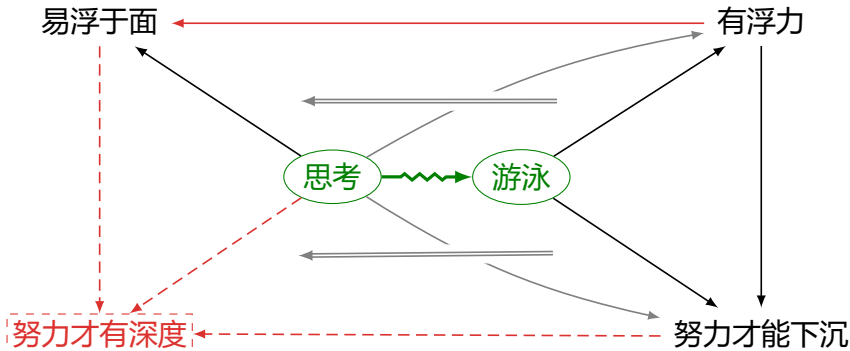
- ▶ 基督徒糕点师拒绝为同性恋制作婚礼蛋糕.
- ▶ 键盘侠: 要求基督徒糕点师为同性恋制作婚礼蛋糕正如要求犹太糕点师为纳粹制作婚礼蛋糕!



思考像游泳

思考像游泳. 游泳时有浮力, 要努力才能下沉; 思考也容易浮于表面, 要努力才能有深度.

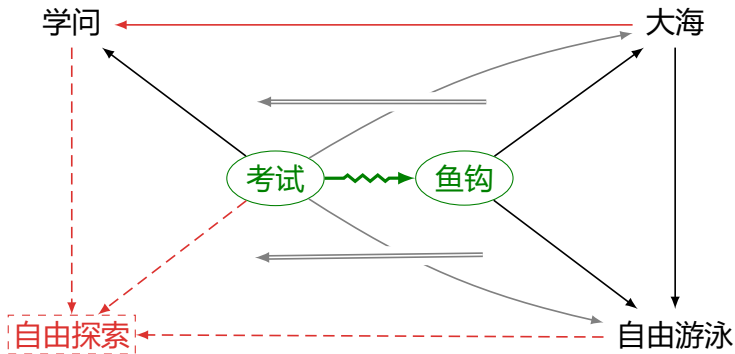
— 维特根斯坦



学问像大海, 考试像鱼钩

学问像大海, 考试像鱼钩, 老师老要把鱼挂在鱼钩上, 叫鱼怎么能在大海中学会自由、平衡地游泳?

— 埃尔米特

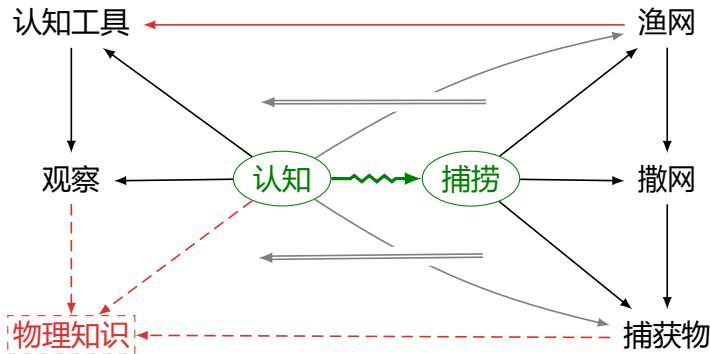


爱丁顿《物理科学的哲学》

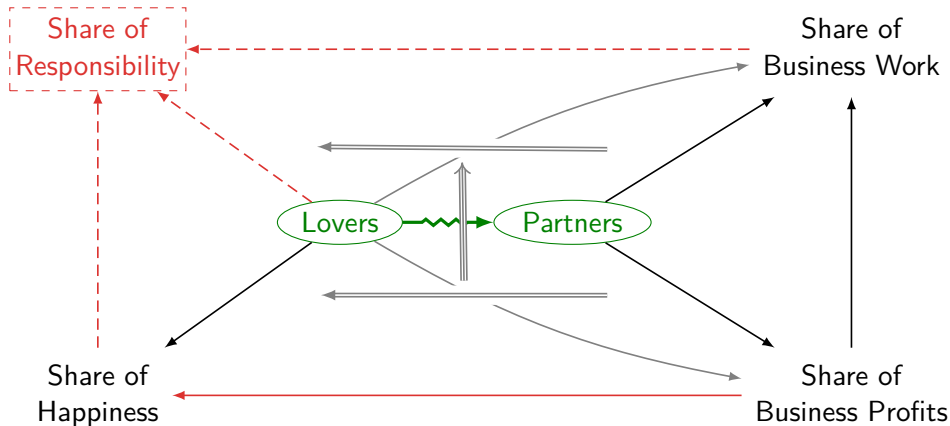
想象一位鱼类专家想探究海洋中的生命. 他舒臂撒网, 捕获了一堆海洋生物. 他检查了自己的捕获物,并由此作出了两项概括:

1. 凡海洋生物皆长于 5 厘米.
2. 凡海洋生物皆有鳃.....

捕获物相当于物理学知识体系, 网相当于思维装置和感官工具, 撒网意味着观察.



隐喻: “Love is a Partnership”



隐喻“凸显”一些东西,“隐藏”一些东西,也可能“创造”一些东西.

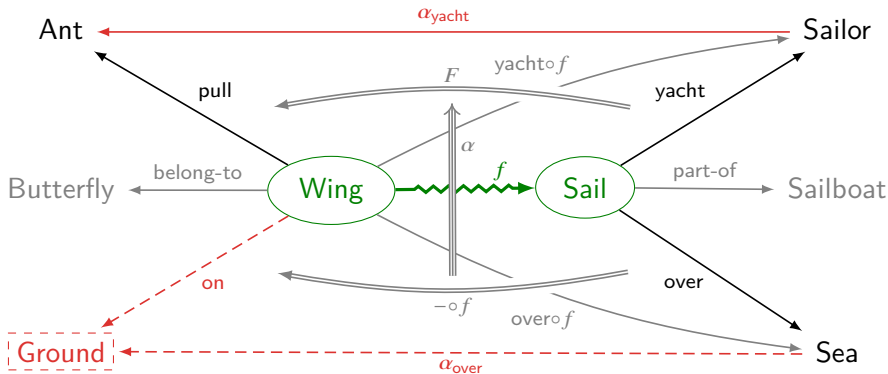
- ▶ 爱情是“旅程”.
- ▶ 爱情是“避风港”. “电磁感应”. “一阵风”. “毒药”. “战争”.....

隱喻: “A Wing is a Sail”

The Ground — A Japanese Poem

An ant
pull a wing of a butterfly.
O
It is like yachting.

$$\begin{array}{ccc} F(\text{yacht}) = \text{pull} & \xleftarrow{\alpha_{\text{yacht}}} & \text{yacht} \circ f \\ \downarrow & \circlearrowleft & \downarrow \\ F(\text{over}) = \text{on} & \xleftarrow{\alpha_{\text{over}}} & \text{over} \circ f \end{array}$$



Contents

Introduction

逻辑学

Toy Logic

穆勒五法

类比论证

谬误与扯淡

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

谬误

一. 形式谬误

二. 非形式谬误

1. 言辞谬误

- ▶ 模糊谬误: 划界谬误或连续体谬误, 假精确或过度精确, 抽象概念当具体概念用
- ▶ 歧义谬误: 一词多义, 歧义句构, 辖域谬误, 重音, 脱离语境、断章取义, 概念扭曲, 偷换概念, 混淆集合与个体或整体与部分, 变更标准
- ▶ 定义谬误: 不当定义, 篡改定义
- ▶ 废话谬误: 平凡真理, 无意义的问题, 回顾性宿命论

2. 实质谬误

- ▶ 不相干
- ▶ 不充分
- ▶ 不当预设

言辞谬误

- ▶ 詹姆斯：当你绕着树转圈时，树上的松鼠也贴着树转圈，并让树始终挡在你和松鼠之间，请问，你绕着松鼠转圈了吗？
- ▶ “可能为真”指“不必然为假”还是“偶然为真”？

必然真理可能为真
一切可能为真的都可能为假

必然真理可能为假

- ▶ 三秀才赶考，途遇算命先生，问几人中举？先生竖起一指。
- ▶ 我头上站着一个无法被探测到的“飞天面条神”。
- ▶ 如果我说你是驴，那么我就是说你是动物；如果我说你是动物，那么我就说了真理；因此，如果我说你是驴，那么我就说了真理。

实质谬误

- ▶ 不相干: 歪曲论题 (稻草人、红鲱鱼、烟雾弹), 诉诸人身 (扣帽子、人身攻击、诉诸动机、罪恶关联、诉诸虚伪、伪善、诉诸成就、富贵、贫贱、智商), 诉诸情感 (诉诸恐惧、厌恶、仇恨、谄媚、同情、愧疚、可爱、性感、时髦、嘲弄、虚荣、势利、沉默), 诉诸暴力、恐吓、诽谤, 诉诸来源、年代、新潮、传统, 诉诸信心、意愿, 诉诸后果、中庸、自然, 转移举证责任, 不得要领
- ▶ 不充分: 不当概括 (偏差样本、偏差统计), 诉诸无知, 诉诸不当权威、名人、大众, 虚假原因, 滑坡谬误, 诉诸可能, 诉诸阴谋, 隐瞒证据, 不当类比, 乱赋因果 (相关、巧合、因果倒置、单因谬误), 完美主义谬误、权宜主义谬误
- ▶ 不当预设: 窃取论题, 非黑即白, 打压对立, 自然主义谬误 (实然推应然), 道德主义谬误 (好的就是自然的), 复合问题, 诱导性提问, 诉诸顽固、反复、冗赘, 乱枪打鸟, 两面讨好

“这鸡蛋真难吃。”

- ▶ 有本事你下一个好吃的蛋啊!
- ▶ 我可以负责任地说, 我们的鸡蛋都是合格的健康蛋!
- ▶ 鸡是优等鸡, 你咋说它下的蛋难吃?
- ▶ 这是别有用心的煽动, 你是何居心?
- ▶ 隔壁的鸭子给了你多少钱?
- ▶ 隔壁家的鸡蛋是伪蛋!
- ▶ 伟大的隔壁老王说好吃, 你跟他说去!
- ▶ 没有伟大的老王, 你连臭蛋都吃不上!
- ▶ 杀掉这只鸡换一只就能下金蛋?
- ▶ 你叫什么名字? 你是干什么的! 你是站在谁的立场上说话?
- ▶ 美国鸡蛋好吃, 你去吧!
- ▶ 你以为你是谁啊, 品蛋师啊? 就你威风!
- ▶ 你个鸭蛋脑残粉!
- ▶ 下蛋的是一只勤劳勇敢善良正直的鸡!
- ▶ 再难吃也是自己家的鸡下的蛋!
- ▶ 但隔壁家的鸡蛋没有我们家的蛋形圆!
- ▶ 吃鸡蛋是我们家的传统美德. 祖宗三代都是吃鸡蛋长大的! 你也是! 你有什么权力说这蛋难吃? 还是不是人!
- ▶ 作为一个吃鸡蛋长大的人, 我为我天天吃鸡蛋感到自豪!
- ▶ 拒绝抹黑! 抵制鸭蛋! 鸡蛋万岁! 鸡蛋加油!
- ▶ 人心理阴暗会导致味觉异常.....
- ▶ 其实隔壁家鸡蛋是个巨大的阴谋, 试图颠覆我们家!
- ▶ 其实邻居家只有少数人才能吃上鸡蛋.
- ▶ 我们这么大的一个家, 问题太复杂, 下蛋没有你想得那么容易.
- ▶ 不要再吵了, 这个家不能乱, 稳定、稳定压倒一切!
- ▶ 要对我们家的鸡有耐心, 它一定会下出更好吃的蛋.
- ▶ 蛋无完蛋!

“这鸡蛋真难吃.”

- ▶ 我们家的鸡已经可以打败隔壁家的鸭!
- ▶ 隔壁家也吃过这样的鸡蛋, 现在是初级阶段, 必须坚持一百年不动摇!
- ▶ 我们家人肠胃不好, 现阶段还不适合吃鸭蛋, 不符合我们家的具体家情!
- ▶ 凡事都有个过程, 现在还不是吃鸭蛋的时候.
- ▶ 鸡蛋好不好吃, 全体蛋鸡最有发言权.
- ▶ 老外都说好吃呢.
- ▶ 这蛋难吃但是历史悠久啊.
- ▶ 虽然难吃但重要的是好看啊.
- ▶ 比以前已经进步很多了.
- ▶ 哎, 人心不古, 世风日下, 就是因为你这种想吃鸭蛋的人太多了.....
- ▶ 隔壁家那鸭蛋更难吃, 你咋不说呢?
- ▶ 嫌难吃就别吃, 滚去吃隔壁的鸭蛋吧.
- ▶ 隔壁亡我之心不死! 该鸡蛋肯定是被隔壁一小撮不会下蛋的鸡煽动变臭的!
- ▶ 你上次吃茄子都吐, 味觉一贯奇葩.
- ▶ 胡说! 我们家的鸡蛋比隔壁家的鸭蛋好吃五倍! 五倍!
- ▶ 是你的思想跟不上鸡蛋口味的升级!
- ▶ 心理阴暗! 连鸡蛋不好吃也要发牢骚!
- ▶ 抱怨有毛用, 有这个时间快去赚钱!
- ▶ 隔壁家的鸡蛋也一样, 天下乌鸦一般黑, 没有好吃的鸡蛋!
- ▶ 吃了人家的鸡蛋还留下证据说鸡蛋难吃, 太有城府了!
- ▶ 很多家都是因为吃隔壁的鸭蛋而导致家庭冲突, 生活水平下降甚至解体!
- ▶ 到目前为止, 我没发现这鸡蛋难吃. 专家说了, 这鸡蛋难吃的可能性不大. 即使出现这种情况, 也是结构性难吃.
- ▶ 荷兰狗/东北猪/瘪三.....不配吃鸡蛋!
- ▶ 大家小心, 此人 IP 在国外.
- ▶ 滚, 你丫是鸡奸, 这里不欢迎你.

假如潘金莲不开窗户, 就不会掉下木棍打到西门庆, 也就不会认识西门庆, 不会出轨, 不会害死武大郎, 武松不会被逼杀人上梁山, 不会有独臂擒方腊, 方腊就可夺取大宋江山, 没了宋就不会有靖康耻、金兵入关, 也不会有元、明、清, 不会闭关锁国、鸦片战争、八国联军. 这样中国将成为超级大国, 称霸世界!

For Want of a Nail

For want of a nail, the shoe was lost,
For want of a shoe, the horse was lost,
For want of a horse, the rider was lost,
For want of a rider, the message was lost,
For want of a message, the battle was lost,
For want of a battle, the war was lost,
For want of a war, the kingdom was lost,
For want of a nail, the world was lost.
And all for the want of a nail.

孔子《论语》

名不正, 则言不顺,
言不顺, 则事不成;
事不成, 则礼乐不兴,
礼乐不兴, 则刑罚不中;
刑罚不中, 则民无所措手足.

礼记·坊记

天无二日，土无二王，家无二主，尊无二上。

董仲舒《春秋繁露》

天以终岁之数，成人之身，故小节三百六十六，副日数也；大节十二，分副月数也；内有五藏，副五行数也；外有四肢，副四时数也；乍视乍瞑，副昼夜也；乍刚乍柔，副冬夏也。

告子：性，犹湍水也，决诸东方则东流，决诸西方则西流。人性之无分于善不善也，犹水之无分于东西也。

孟子：水信无分于东西，无分于上下乎？人性之善也，犹水之就下也。人无有不善，水无有不下。

孟子《生于忧患，死于安乐》

舜发于畎亩之中，傅说举于版筑之中，胶鬲举于鱼盐之中，管夷吾举于士，孙叔敖举于海，百里奚举于市。故天将降大任于斯人也，必先苦其心志，劳其筋骨，饿其体肤，空乏其身，行拂乱其所为，所以动心忍性，曾益其所不能。

《庄子·外篇·天地》

子贡南游于楚，反于晋，过汉阴，见一丈人方将为圃畦，凿隧而入井，抱瓮而出灌，搨搨然用力甚多而见功寡。子贡曰：“有械于此，一日浸百畦，用力甚寡而见功多，夫子不欲乎？”为圃者仰而视之曰：“奈何？”曰：“凿木为机，后重前轻，挈水若抽，数如佚汤，其名为橰。”为圃者忿然作色而笑曰：“吾闻之吾师，有机械者必有机事，有机事者必有机心。机心存于胸中，则纯白不备。纯白不备则神生不定，神生不定者，道之所不载也。吾非不知，羞而不为也。”

法兰克福《论扯淡》

假设在美国的独立日上，总统声嘶力竭地对大众说：“我们伟大的、受到上帝保佑的祖国，它的开国元勋在神圣的引导下，为人类开创了一个崭新的篇章……”^a

^a到底有没有上帝？上帝有没有保佑美国？美国是否伟大？美国的开国元勋是否受到了神圣的引导？美国的建立在人类历史上是一个怎样的地位？为人类开创了一个怎样的新篇章？

有 6000 人死于醉酒; 有 4000 人死于开车; 但只有 100 人死于醉酒开车. 因此, 醉酒开车比单纯的醉酒或者单纯的开车更安全.

莱布尼茨《单子论》

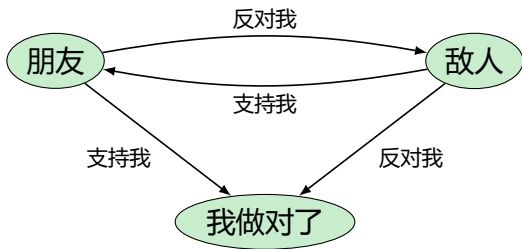
如果单子没有知觉, 那么其复合物也没有知觉.

帕斯卡赌

如果上帝不存在, 但你相信上帝存在, 也没太大损失; 然而, 如果上帝存在, 而你却不相信上帝存在, 那你将面临巨大的惩罚. 所以, 应该相信上帝存在.

鲁迅《论辩的灵魂》

我骂卖国贼, 所以我是爱国者. 爱国者的话是最有价值的, 所以我的话是不错的, 我的话既然不错, 你就是卖国贼无疑了!



- ▶ 朋友越是支持我, 越说明我做对了.
- ▶ 敌人越是反对我, 越说明我做对了.
- ▶ 连敌人都不得不支持我, 变成我的朋友, 更说明我做对了.
- ▶ 如果朋友反对我, 说明他其实是敌人.

On the reception and detection of pseudo-profound bullshit

请为下面几句话的“深刻性”打分:

1. We are in the midst of a self-aware blossoming of being that will align us with the nexus itself.
2. A wet person does not fear the rain.
3. Most people enjoy some sort of music.

► Bullshit Receptability 与超自然信仰、阴谋论、替代医学正相关.

bullshit receptability := mean of the profoundness ratings for all bullshit

► Bullshit Sensitivity 与批判性思维能力正相关.

$$\begin{aligned} & \text{bullshit sensitivity} \\ & \quad := \\ & \quad \frac{\text{profundity ratings for meaningful quotations}}{\text{profundity ratings for pseudo-profound bullshit}} \end{aligned}$$

扯淡

	说谎	扯淡
真理	知道真理	不关心真理
意图	掩盖真理	影响听众

“相比于说谎, 扯淡是真理更大的敌人。”

— 法兰克福

- ▶ 学术: 晦涩的语言掩盖思想的贫乏.
- ▶ 政治: 模糊的辞令掩盖政策的空洞.
- ▶ 营销: 华丽的术语包装低劣的产品.
- ▶ 健康养生: 无法证伪的理论导向有害的行为.

Wholeness depends on dimensionless phenomena. Reality has always been full of messengers of the multiverse, whose third eyes are transformed into transcendence. Transcendence is the healing of choice. Complexity is the driver of transcendence. Our conversations with other messengers have led to an awakening of ultra-non-local consciousness. Consciousness requires exploration. We are at a crossroads of flow and ego. We can no longer afford to live with ego. Where there is ego, life can't thrive. We exist as expanding wave functions. The goal is to plant the seeds of passion rather than bondage. We are in the midst of a self-aware blossoming of being that will align us with the nexus itself. Lifeform, look within and recreate yourself. To follow the path is to become one with it. By unfolding, we believe; By deepening, we vibrate; By blossoming, we self-actualize. We dream, we heal, we are reborn. We must learn how to lead unlimited lives in the face of delusion. You and I are dreamweavers of the quantum soup. The infinite is approaching a tipping point. Hidden meaning transforms unparalleled abstract beauty. **Wholeness quiets infinite phenomena.**

How to generate pseudo-profound bullshit?¹

1. State the blindingly obvious (of life's big theme) incredibly slowly.
 - ▶ We were all children once.
2. Doublethink/Dialectic/Contradiction.
 - ▶ War is peace. Freedom is slavery. Ignorance is strength.
 - ▶ Everyone is the other, and no one is himself.
3. Ambiguity/Metaphor/Parable.
 - ▶ Language is the house of the truth of Being.
 - ▶ Never stay up on the barren heights of cleverness, but come down into the green valleys of silliness.
 - ▶ Ethics does not treat of the world. Ethics must be a condition of the world, like logic.
 - ▶ A person is neither a thing nor a process but an opening through which the Absolute can manifest.
 - ▶ Making itself intelligible is suicide for philosophy. Those who idolize "facts" never notice that their idols only shine in a borrowed light.

¹Law: Believing Bullshit.

Frankfurt: On Bullshit.

Bergstrom & West: Calling Bullshit.

How to generate pseudo-profound bullshit?

4. Analogy.

- ▶ We have got on to slippery ice where there is no friction, and so, in a certain sense, the conditions are ideal; but also, just because of that, we are unable to walk. We want to walk: so we need friction. Back to the rough ground!
- ▶ The subject does not belong to the world: rather, it is a limit of the world. This is exactly like the case of the eye and the visual field. You do not see the eye. Nothing in the visual field allows you to infer that it is seen by an eye. Our life is endless in the way that our visual field is without limit.

5. Use jargon.

- ▶ Profound boredom, drifting here and there in the abysses of our existence like a muffling fog, removes all things and men and oneself along with it into a remarkable indifference. This boredom reveals being as a whole.
- ▶ A machinic assemblage, through its diverse components, extracts its consistency by crossing ontological thresholds, non-linear thresholds of irreversibility, ontological and phylogenetic thresholds, creative thresholds of heterogenesis and autopoiesis.

What is a Good Argument?

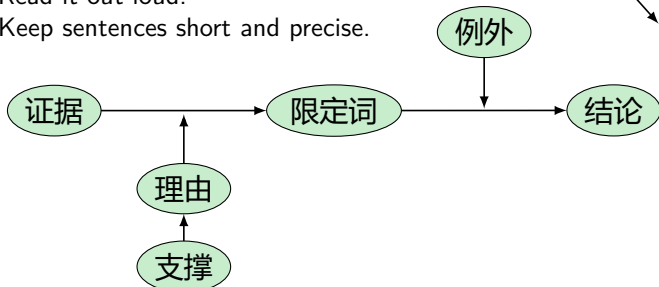
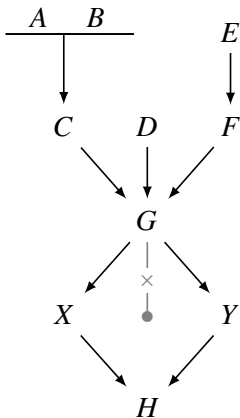
A **good argument** should:

1. be deductively **valid** (or inductively strong) and have **premises all true**;
2. have its validity (or inductive strength) and truth-of-premises be as evident as practically possible to the parties involved;
3. be clearly stated;
4. avoid ambiguity, emptiness, circularity, vicious regression, and emotional language;
5. be relevant to the issue at hand.

Writing & Argument

1. Before you write: Structuring the paper.
 - 1.1 Diagram the logic of your argument.
 - 1.2 Break your work into sections/paragraphs.
 - 1.3 Write a brief abstract for each section.
2. Write: Fill in the details.

Tell them what you're going to say.
Say it.
Tell them what you've just said.
3. Rewrite: Improving your language.
 - 3.1 Read it out loud.
 - 3.2 Keep sentences short and precise.



Contents

Introduction

逻辑学

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

Contents

Introduction

逻辑学

形而上学

存在

因果

时空

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

何物存在?

唯名论:
抽象事物不存在

极端唯名论:
即使具体事物也不存在

柏拉图主义:
抽象事物存在

极端柏拉图主义:
就连具体事物也存在

- ▶ 名字叫“罗素”的那个人存在
 - ▶ 人存在
 - ▶ 电子存在
 - ▶ 红存在
 - ▶ 2 存在
- “如果其他事物存在, 那数也存在.”

— 柏拉图《智者篇》

忒修斯之船

- ▶ 如果忒修斯之船上的全部木头逐渐被替换, 替换下来的木头又被重组成一艘船. 哪一艘才是忒修斯之船?
- ▶ 极端唯名论者: 不存在真正的忒修斯之船. 真正存在的是经验刺激. 人们区分或命名的是这些不同的经验刺激. 而区分或命名并不总是严格的.

Contents

Introduction

逻辑学

形而上学

存在

因果

时空

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

Type Causation & Token Causation

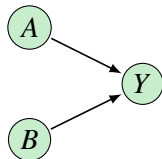
- ▶ **General/Type Causation:** 抽烟导致肺癌
- ▶ **Actual/Token Causation:** 张三抽了三十年烟, 导致了肺癌

因果解释:

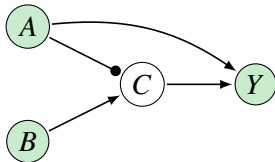
“昨天晚上是下着暴雨, 我也喝醉了, 但是, 事故的真正原因是, 车的刹车片坏了.”

Actual Causation vs PN, PS

- ▶ 实际因果在构建解释、责任划分中起重要作用.
- ▶ 必要概率 PN 和充分概率 PS 只依赖 $Y_x(u)$, 关注因果模型的全局特征 (输入-输出), 无视因果过程.
- ▶ 实际因果必须考虑因果过程.
 - $B = 1$ 是 $Y = 1$ 的原因吗? 第一个例子中 “是”, 第二个 “不是”.



$Y = A \vee B$
$A = 1, B = 1, Y = 1$



$C = B \wedge \neg A$
$Y = A \vee C$
$A = 1, B = 1, C = 0, Y = 1$

$$A \vee C \equiv A \vee (B \wedge \neg A) \equiv A \vee B$$

Philosophy — Process Approach to Causality?

- ▶ A **causal process** is a world line of an object which possesses a conserved quantity.
- ▶ A **causal interaction** is an intersection of world lines which involves exchange of a conserved quantity.
- ▶ A causes B if there is a transfer of energy or momentum from A to B .

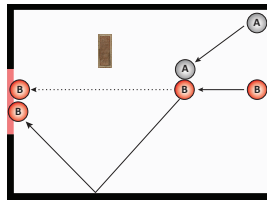
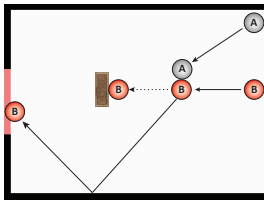
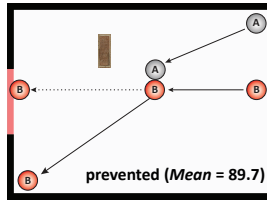
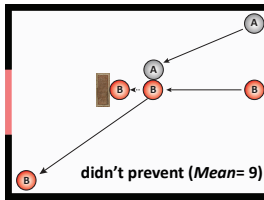
1. 相同过程, 不同因果?

process?

counterfactual?

2. 开关是灯亮的原因, 传输了啥?

3. 欲望导致行动, 传输了啥?



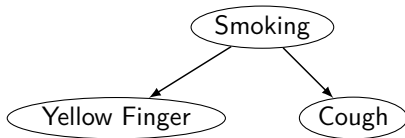
Philosophy — Probabilistic Approaches to Causality?

► **Reichenbach:** C causes E , iff,

1. C is earlier than E , $t_C < t_E$,
2. $P(E | C) > P(E | \neg C)$,
3. there is no event S (earlier than or simultaneous with C) that S screens off C from E .

$$\neg \exists S : P(E | C \wedge S) = P(E | S)$$

Remark: Simpson paradox



- **Cartwright:** C causes E , iff, $P(E | C \wedge B) > P(E | \neg C \wedge B)$ for all background context B .²
- **Dupré:** “average degree of causal significance”:

$$\sum_B P(B) [P(E | C \wedge B) - P(E | \neg C \wedge B)]$$

²Background context: all causal factors for E that excludes C and the effects of C .

Suppes' Genuine Cause — Probabilistic Approach

- ▶ **Suppes' *prima facie* cause:** C is a *prima facie* cause of E iff
 1. $t_C < t_E$
 2. $P(C) > 0$
 3. $P(E | C) > P(E | \neg C)$
- ▶ **Suppes' first *spurious* cause:** C , a *prima facie* cause of E , is a *spurious* cause iff there exists S such that
 1. $t_S < t_C < t_E$
 2. $P(C \wedge S) > 0$
 3. $P(E | C \wedge S) = P(E | S)$
 4. $P(E | C \wedge S) \geq P(E | C)$
- ▶ **Suppes' second *spurious* cause:** C , a *prima facie* cause of E , is a *spurious* cause iff there is a partition $\mathcal{E}$ and for every $S \in \mathcal{E}$
 1. $t_S < t_C < t_E$
 2. $P(C \wedge S) > 0$
 3. $P(E | C \wedge S) = P(E | S)$

Remark: ε -spurious: $|P(E | C \wedge S) - P(E | S)| < \varepsilon$
- ▶ **Suppes' genuine cause:** nonspurious *prima facie* cause.

- ▶ 概率因果依赖时序关系.
- ▶ 对背景变量的选择有死循环的风险.
 - 若要求对环境的完整描述, 则会将概率性关系化归为确定性方程.
 - 若描述的过于粗略, 则会导致伪相关和其它混杂效应.
 - 而要求背景变量与所讨论的变量“因果相关”则导致死循环.
 - 如何选择合适的背景变量, 类似于如何找寻合适的控制变量去混杂, 这必须依赖因果.

Philosophy — Regularity Approaches to Causality?

- ▶ **Mill's Sufficient** Condition: C causes E iff $C \rightarrow E$.
- ▶ **Hobbes's Necessary** Condition: C causes E iff $\neg C \rightarrow \neg E$.
- ▶ **Ramsey Test**: $C \Box \rightarrow E$ should be believed iff, after suspending judgment on C and E , E is believed as a result of assuming C .

$$C \Box \rightarrow E \in K \iff E \in K * C$$

- ▶ **Wright's NESS** condition: C causes E iff C is a Necessary Element of a Sufficient Set for E .
 1. $C \wedge X$ is E 's sufficient condition. $C \wedge X \rightarrow E$
 2. X is not sufficient for E . $X \nrightarrow E$

Mackie's Actual Causation — Regularity Approach

- ▶ **Mackie's INUS Condition:** (insufficient but necessary part of a causal condition that is itself unnecessary but sufficient of the effect)

1. $C \wedge X$ is E 's sufficient but unnecessary condition.
2. C is not sufficient for E . $C \nrightarrow E$
3. X is not sufficient for E . $X \nrightarrow E$

$$(C \wedge X) \vee Y \leftrightarrow E$$

Example: 为什么造假币? 因为不会造真币、没钱又不会赚钱...

— Why E ?

— C_1 rather than C_2 , since $C_1 = \operatorname{argmax}_C P(E | C)$?

- ▶ **Mackie's actual causation:**

- ▶ C is at least an INUS condition of E
- ▶ C was present
- ▶ Components of X were present
- ▶ Every disjunct in Y not containing C as a conjunct was absent

Lewis' Actual Causation — Counterfactual Approach

Definition (But-For Cause)

$X = x$ is a **but-for** cause of $Y = y$ in (M, u) iff

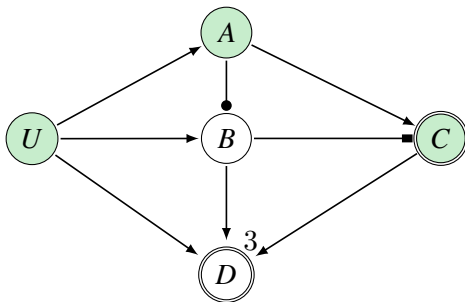
1. $M, u \models X = x \wedge Y = y$
2. there exist $x' \neq x$ and $y' \neq y$ such that $M, u \models [X = x']Y = y'$

Definition (Actual Causation — Lewis 1973)

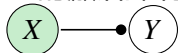
$X = x$ is an actual cause of $Y = y$ in (M, u) iff there exists a sequence of variables $Z_1 = X, \dots, Z_n = Y$ s.t. $Z_i = z_i$ is a but-for cause of $Z_{i+1} = z_{i+1}$ for $i = 1, \dots, n - 1$.

Remark: Is causation transitive?

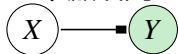
Neuron Diagrams



- ▶ X 的激活抑制了 Y

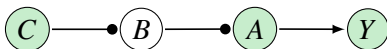


- ▶ X 不激活则 Y 激活



- ▶ $\textcircled{\textcircled{X}}^n$ 多个信号方能激活 X

Example — Double Prevention



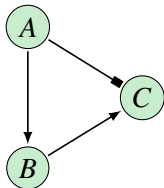
$B = \neg C$
$A = \neg B$
$Y = A$
$C = 1, B = 0, A = 1, Y = 1$

- ▶ Alice A is planning to hack Yuri's Y computer.
- ▶ Bob B launches a missile at Alice's city.
- ▶ Carl C shoots down the missile.
- ▶ Alice hacks Yuri's computer, without any knowledge that Bob and Carl even exist.
- ▶ Nevertheless, Carl caused Yuri's computer being hacked.
- ▶ $C = 1$ is an actual cause of $Y = 1$.

Example — Transitivity?

1. A is a cause of B
2. B is a cause of C
3. A is a cause of C ?

- ▶ Alice 医生治好了 Bob 的致命疾病 A .
- ▶ Bob 出院时 B , 被车撞死 C .



$B = A$
$C = \neg A \vee B$
$A = 1, B = 1, C = 1$

Sufficient Conditions for “But-For” Cause to be Transitive

Theorem (Transitivity of But-For Causation — Halpern 2016)

Suppose $X = x$ is a but-for cause of $Y = y$ in (M, u) , and $Y = y$ is a but-for cause of $Z = z$ in (M, u) , and the following two conditions hold,

1. for every $y' \in R(Y)$, there exists an $x' \in R(X)$ such that

$$M, u \models [X = x']Y = y'$$

2. Y is on every causal path from X to Z .

Then $X = x$ is a but-for cause of $Z = z$ in (M, u) .

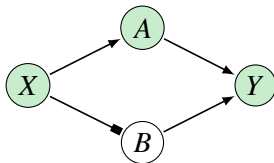
Remark: 对比 Pearl 的传递性的充分条件: 当 X 对 Z 没有直接影响时, 传递性成立.

$$\frac{Z_{y'x'} = Z_{y'}}{x \rightarrow y \ \& \ y \rightarrow z \implies x \rightarrow z}$$

Woodward's Manipulability Theory of Causation

1. X is a **total cause** of Y iff changing X will change Y for some values of all other variables that aren't descendants of X .
2. X is a **direct cause** of Y iff changing X will change Y when all variables other than X and Y are fixed at some values.
3. X is a **contributing cause** of Y iff there is a directed path from X to Y such that, changing X will change Y when the variables not on this path are fixed at some values.

Type of cause	What is held fixed
total	all other variables not descendants of X
direct	all variables other than X and Y
contributing	all variables not on one directed path from X to Y



$$A = X$$

$$B = \neg X$$

$$Y = A \vee B$$

X is not a total or a direct cause, but a contributing cause of Y

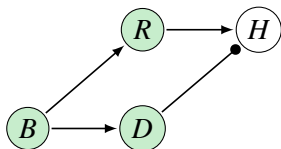
Woodward's Manipulability Theory of Causation

Definition (Actual Causation — Woodward 2003)

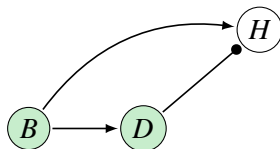
$X = x$ is an actual cause of $Y = y$ in (M, u) iff

1. $M, u \models X = x \wedge Y = y$
2. there is a directed path from X to Y such that, some intervention on X will change Y when the variables not on this path are fixed at their actual values.

Example: A boulder falls (B) and rolls toward the hiker (R). The hiker ducks (D) so that he does not get hit ($\neg H$).



$$H = R \wedge \neg D$$



$$H = B \wedge \neg D$$

Problem: $B = 1$ is a Woodward cause of $H = 0$ in G_1 but not in G_2 .

Contents

Introduction

逻辑学

形而上学

存在

因果

时空

认识论

伦理学

心灵哲学

语言哲学

科学哲学

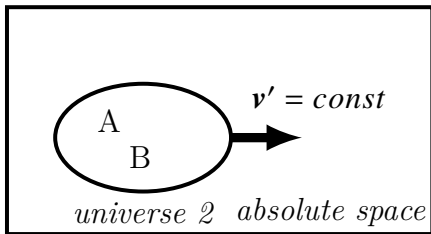
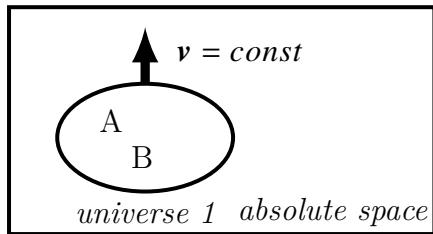
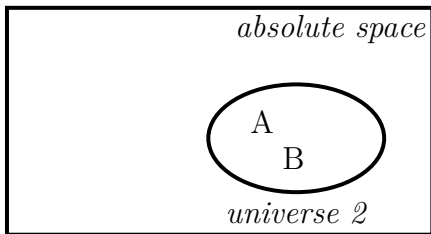
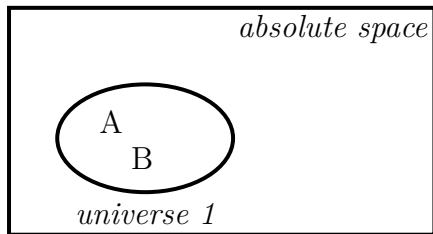
政治哲学

时空是绝对的还是相对的？

牛顿 PK 莱布尼茨

- ▶ 牛顿: 时空是绝对的, 独立于万物而存在. 即使没有物质, 空的空间依然存在.
- ▶ 莱布尼茨: 时空是相对的.
 - ▶ 空间是事物之间的相互关系, 没有事物就没有空间, 不存在空的空间.
 - ▶ 时间是事物关系变化的度量, 没有事物就没有时间, 不存在空的时间.
 - ▶ 如果空间是绝对的, 那么将宇宙往右平移一米将与原宇宙不可区分; 如果时间是绝对的, 那么将宇宙往后平移一分钟将与原宇宙不可区分. 违背“不可分辨者的同一性”.
 - ▶ 宇宙是一个因果关系的网络. 其中每一部分的性质由其与其它部分的关系决定.

Leibniz's "Static" & "Kinematic" Shift Argument



- ▶ 牛顿: $\text{universe1} \neq \text{universe2}$. 理由: 绝对位置/绝对速度不同.
- ▶ 莱布尼茨: $\text{universe1} = \text{universe2}$. 理由: 一、**充足理由律**. 上帝为什么创造了 universe1 而不是 universe2 ? 二、**不可分辨者的同一性原则**. universe1 和 universe2 不可分辨.

莱布尼茨对牛顿绝对时空的质疑

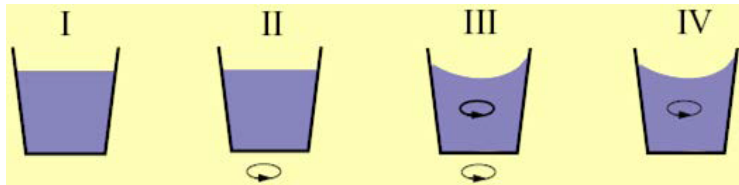
- ▶ 遵循牛顿力学的物体的位置和速度依赖绝对时空吗？
- ▶ 将整个宇宙往右平移一米，牛顿力学会受影响吗？
- ▶ 牛顿三定律、万有引力定律

$$F = ma, \quad F = G \frac{m_1 m_2}{r^2}$$

涉及物体的绝对时空位置或速度吗？

- ▶ 绝对位置、绝对速度不可观测，绝对时空没有物理意义。
- ▶ 牛顿的回应：绝对加速度总是伴随惯性力，是可观测的。相对时空怎么解释惯性？

牛顿的水桶实验



	I	II	III	IV
桶	静止	转动	转动	静止
水	静止	静止	转动	转动
相对运动	否	是	否	是

- ▶ 旋转水桶, 开始, 水由于惯性尚未旋转, 水面是平的; 然后, 水随桶旋转, 水面凹; 最后, 水桶静止, 水继续旋转, 水面凹.
- ▶ 牛顿: 水面的凹陷并不由其与桶的相对运动引起的, 而是由其相对于绝对空间的绝对运动引起的.

从莱布尼茨到马赫到爱因斯坦

- ▶ 马赫: 水面的凹陷是由其与宇宙中所有其他物质的相对运动引起的.
- ▶ 想象水桶壁非常非常厚, 整个宇宙的物质变成了厚厚的水桶壁, 那么, 当水相对于桶壁转动时, 水面就会凹陷.
- ▶ 马赫原理: 物体的惯性是宇宙中其他所有物质对该物体作用的结果.
- ▶ 马赫原理摒弃了绝对时空, 把惯性归结为引力, 这启发爱因斯坦发现了等效原理: 引力场与惯性力场局域不可区分.
- ▶ 根据等效原理, 旋转圆盘是个加速系, 它等价于一个引力场. 根据狭义相对论, 圆盘在旋转时, 沿边缘方向的尺子会发生洛伦兹收缩, 导致测量出的圆盘的周长变大, 而沿半径方向的尺子不会收缩, 直径不变. 这意味着, 测得的圆周率变大:

$$\frac{\text{周长}}{\text{直径}} > \pi$$

加速系 (引力场) 的几何是非欧几里德的. 引力是局域弯曲时空的几何效应. 物体在引力场中的自由落体, 不过是在弯曲时空中沿着测地线运动.

- ▶ 注: GR 并没有完全实现马赫原理. 宇宙只有薄壁水桶时, 水相对于桶壁旋转, 根据马赫原理, 水面不会凹陷, 而根据 GR, 会凹陷.

康德“一只手”挑战莱布尼茨

- ▶ 设想宇宙里只有孤零零的一只手.
- ▶ 这只手要么是左手, 要么是右手 — 即一只手不可能缺乏“手性”.
- ▶ 如果这样一只手确实具有手性, 那么究竟有什么“关系性事实”能够决定它是左手还是右手呢?
- ▶ 这只手的手性肯定是由其所处的绝对空间确定的.



- ▶ 莱布尼茨可能的回应: 一只手究竟是“左手”还是“右手”, 不是这只手的“固有属性”, 而是取决于整个空间的全局拓扑结构. 一只手在三维欧式空间中有“手性”, 并不意味着在其他空间也有手性 (比如高维空间或类似莫比乌斯的空间).
- ▶ 康德可能的回应: 一只手是怎么决定这样一种空间的拓扑结构的?

李·斯莫林对莱布尼茨“关系主义”宇宙观的解释

1. 宇宙之外一无所有. (只有一个宇宙.)
2. 因果完备性: 宇宙是因果封闭的. (万事皆有因, 但因果链不能追溯到宇宙之外. 宇宙的自然律和初始条件也需要有因果解释.)
3. 时空相对性: 空间是事物之间的相互关系; 时间是事物关系变化的度量.
4. 宇宙不是由物质构成, 而是由事件的因果过程构成. 宇宙是一个因果关系的网络. 其中每一部分的性质由其与其它部分的关系决定.
5. 相互作用: 如果 A 作用于 B , B 也一定会作用于 A . (时空告诉物质如何运动; 反过来, 物质告诉时空如何弯曲. 而牛顿的绝对时空是单向的, 时空指引物质运动, 物质却无法影响时空.)
6. 背景独立性: 基础物理理论不应该依赖于固定不变的背景, 即只建模宇宙的一部分, 而把其余部分视为不随其它要素变化的背景.

李·斯莫林对莱布尼茨“关系主义”宇宙观的解释

7. 事件独特性: 每个事件都是可分辨的, 所以, 任何两个事件的因果过去都是不同构的.
8. 没有任何两个瞬间完全相同, 从这个意义上说, 时间是真实的.
9. 对称性是背景依赖的性质. 由于“不可分辨者的同一性”, 不存在终极的对称性. 对称性, 要么是近似的, 要么是破缺的. (对称性, 比如时间反演不变性, 源于粗粒化过程中对事件独特性的忽略.)
10. 开放系统的能量流, 倾向于驱动系统向组织度更高的结构演化. (在熵最大的平衡态, 许多瞬间会不断复现.)
11. 观察者是宇宙的一部分.
12. 观察者只能观测到宇宙的部分信息. 未来的观察者可以看到更多信息.
13. 逻辑与观察者有关. 不能假定每个陈述非真即假. 至少有三类: 一类可判断为真, 一类可判断为假, 一类无法判断.

Contents

Introduction

逻辑学

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

知识的种类

	日常知识	科学知识	技术知识	伦理知识	数学知识
社会共同体	社会人群	科学家	工程师	社会群体	数学家
知识的对象	生活世界	自然界	人工自然	价值系统	抽象实体
知识的目标	处理日常生活问题	自然定律	社会效益	伦理规范	形式系统
知识的方法	教育、传媒、试错、交往	经验理性 可证实标准	工具理性、 决策逻辑	博弈论	形式演绎
真值条件	常识、习惯、科学	实验、理论 批判	应用标准、 专利审查	社会认同	系统可靠性、 完备性、简单性

知识的种类

- ▶ 事实性知识: 采用直接表示的形式
— 凡是猴子都有尾巴
- ▶ 过程性知识: 描述做某件事的过程
— 摩托车修理
- ▶ 行为性知识: 不直接给出事实本身, 只给出它在某方面的行为
— 微分方程、(事物的内涵)
- ▶ 实例性知识: 只给出一些实例, 知识藏在实例中
- ▶ 类比性知识: 既不给出外延, 也不给出内涵, 只给出它与其它事物的某些相似之处
— 比喻、谜语
- ▶ 常识性知识
- ▶ 元知识: 有关知识的知识. 如何使用知识的知识

Know that, Know whether, Know what, Know who, know where, Know how, Know why ...

Classical Empiricism

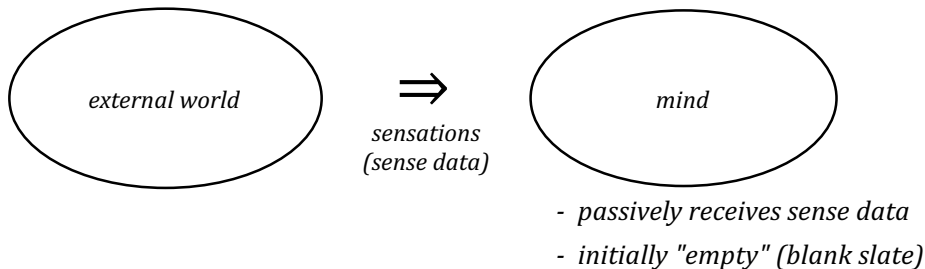


Figure: Locke, Berkeley, Hume: The only source of knowledge of the external world is experience.

- ▶ How is knowledge of the external world possible?
- ▶ How is knowledge of the future based only on past experience possible?

Rationalism

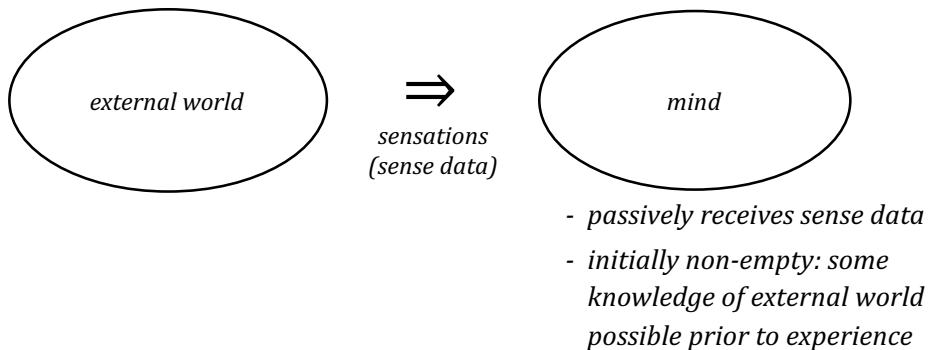


Figure: Descartes: There can be certain knowledge based on pure reason alone.

a priori knowledge = certain knowledge independent of experience.

Immanuel Kant 1724-1804

	a priori	a posteriori
analytic	✓	×
synthetic	?	✓



Synthetic a priori statement = truth is established by reason alone (a priori) and contains factual content (synthetic).

	先验	后验
分析	数学	×
综合	×	物理

Table: 康德之前

	先验	后验
分析	逻辑	×
综合	算术 几何	

Table: 康德

	先验	后验
分析	×	×
综合	×	逻辑 算术 几何

Table: 蒯因、普特南

	先验	后验
分析	逻辑 算术 纯粹几何	×
综合	×	应用几何

Table: 逻辑实证主义

	先验	后验
分析		×
综合	逻辑 算术 纯粹几何	应用几何

Table: Martin-Löf

Kant

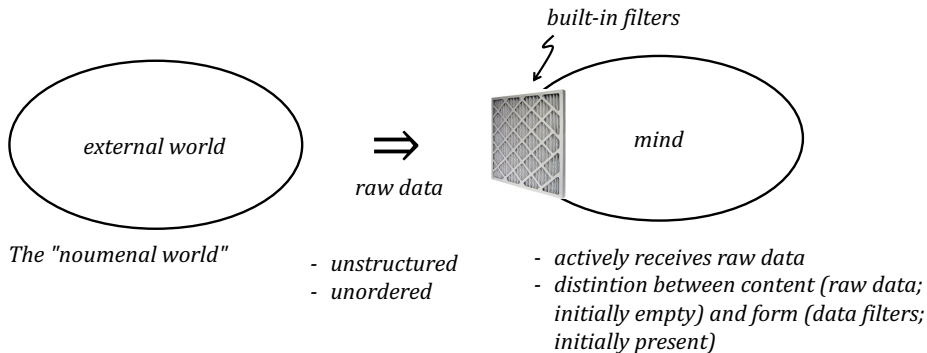
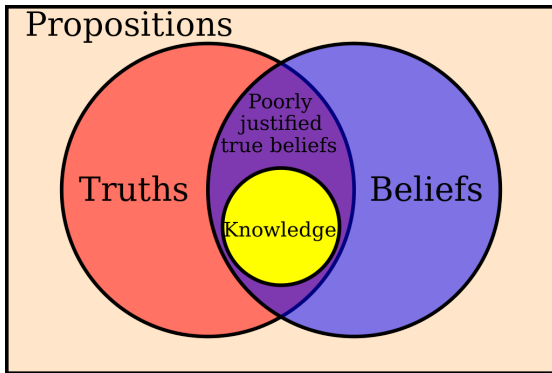


Figure: Kant: All structure and order (causal, temporal, spatial, etc) is imposed on raw data by filters ("forms") already present in the mind.

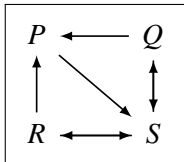
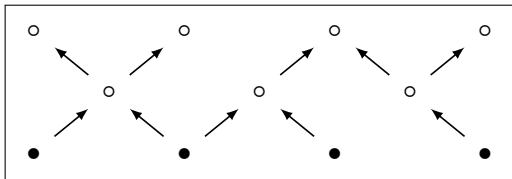
人🕶️为自然立法!

Gettier Problem: Is Justified True Belief Knowledge?



1. Smith believes "**Bob owns a Ford**".
2. He was told this by Bob.
3. Bob then sells his Ford.
4. Meanwhile Bob wins a Ford in a raffle.

Epistemic Justification



- ▶ Internalist theory of epistemic justification.
 - ▶ Foundationalism
 - S 's belief p is justified iff either
 - (a) p is a basic belief; or
 - (b) p is a non-basic belief justified inferentially by S 's basic beliefs.
 - ▶ Coherentism
 - S 's belief p is justified iff S has a coherent set of beliefs which includes p .
- ▶ Externalist theory of epistemic justification.
 - ▶ Reliabilism
 - S 's belief p is justified iff p is produced by a reliable cognitive process.

证成难题

1. 一个信念只能由另一个信念证成.
 2. 不允许循环证成.
 3. 证成链条是有穷的.
 4. 有被证成的信念.
- ▶ 上面四个命题不相容.
 - ▶ 基础论者拒斥 1.
 - ▶ 融贯论者拒斥 2.
 - ▶ 皮尔士拒斥 3.
 - ▶ 取消论者拒斥 4.

Nozick's Truth-Tracking Condition

Nozick's Truth-Tracking Condition

S knows that p iff:

1. p is true;
2. S believes that p ;
3. if p were not true, S would not believe that p ;
4. if p were true, S would believe that p .

$$Kp := p \wedge Bp \wedge (\neg p \Box \rightarrow \neg Bp) \wedge (p \Box \rightarrow Bp)$$

Kripke's counter-example — Fake Barn Country

- ▶ Smith is driving in a country containing fake barns.
- ▶ The fake barns are painted green.
- ▶ In the midst of these fake barns is one real barn, which is painted red.
 1. Smith looks up and happens to see the real barn. **"I see a red barn."**
 2. What if Smith looks up and forms the belief **"I see a barn"**?
- ▶ Smith knows "there is a red barn", but doesn't know "there is a barn".

怀疑论 vs 反事实

S_1 I know “I have hands”. Kp

S_2 I know if “I have hands” then “I am not a brain in the vat”. $K(p \rightarrow q)$

S_3 I don't know that “I am not a brain in the vat”. $\neg Kq$

— 我不知道 “我不是缸中脑” $\neg Kq$, 因为, 在距离 “我是缸中脑” 的最邻近世界里, 我会错误地相信 “我不是缸中脑”.

- ▶ Nozick's definition of Knowledge

$$Kp := p \wedge Bp \wedge (\neg p \Box \rightarrow \neg Bp) \wedge (p \Box \rightarrow Bp)$$

then K is not closed under known entailment.

$$Kp, K(p \rightarrow q) \not\models Kq$$

- ▶ By Sosa's definition,

$$Kp := p \wedge Bp \wedge (Bp \Box \rightarrow p) \wedge (p \Box \rightarrow Bp)$$

then

$$Kp, K(p \rightarrow q) \models Kq$$

德性知识论: 你知道 p , 当且仅当, p 是真的, 你相信 p , 你是因为自己优秀的认知能力和品德才相信的 p .

逻辑全知 vs 怀疑论

$$\frac{A}{KA} \quad \frac{A \rightarrow B}{KA \rightarrow KB} \quad \frac{A \leftrightarrow B}{KA \leftrightarrow KB} \quad \frac{K(A \rightarrow B) \quad KA}{KB}$$

S_1 I know "I have hands". Kp

S_2 I know if "I have hands" then "I am not a brain in the vat". $K(p \rightarrow q)$

S_3 I don't know that "I am not a brain in the vat". $\neg Kq$

Possible "solutions":

- ▶ Impossible worlds: inconsistent alternatives
- ▶ Awareness: K = awareness + implicit knowledge
- ▶ Algorithmic knowledge: K = answer by algorithm
- ▶ Timed knowledge: reasoning takes time (polynomial time)
- ▶ Neighbourhood semantics: still problematic
- ▶ Counterfactual implication:
 - ▶ Nozick: $Kp := p \wedge Bp \wedge (\neg p \Box \rightarrow \neg Bp) \wedge (p \Box \rightarrow Bp)$
 - ▶ Sosa: $Kp := p \wedge Bp \wedge (Bp \Box \rightarrow p) \wedge (p \Box \rightarrow Bp)$

Aaronson 基于计算复杂性的回答

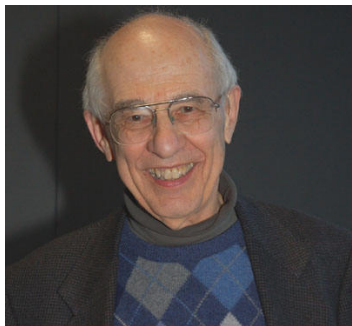
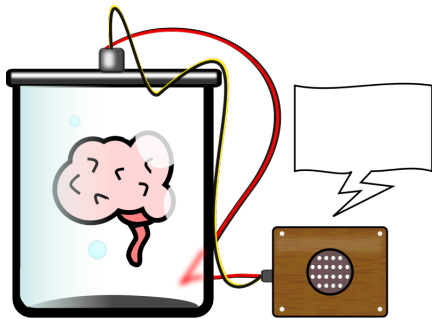
- ▶ 你**知道**微积分/西班牙语/桥牌游戏的规则吗?
- ▶ 你是否拥有一个内部**算法**, 凭它你能在**合理的时间**内回答某种形式的一大类 (可能是无限的) 问题?
- ▶ 你之所以能“知道”(know-that) $1591 = 43 \times 37$, 却又同时不“知道”1591 的素因子, 是因为当我们说知道这些问题的答案时, 我们其实指的是知道怎么 (know-how) 回答它们.
- ▶ 知道一个快速的乘法算法却不知道一个快速的因子分解算法, 两者之间并不矛盾, 即使我们把关于算法的知识建模成演绎封闭的³.

— Scott Aaronson: *Why Philosophers Should Care About Computational Complexity*

³可以以公理化的方式刻画一个多项式时间可计算的函数类, 其中的函数在满足公理的运算下仍然属于这个函数类.

Hilary Putnam 1926-2016

- ▶ Putnam's "brain in a vat" experiment: (no) real experience.



- ▶ Even if you're a brain in a vat, you can still reason, "I think, therefore I am."
- ▶ Brain in a vat which thinks that it is a brain in a vat?

Putnam's Semantic Externalist Argument against Skepticism

Assumptions:

- ▶ Causal Constraint: A term refers to an object only if there is an causal connection between that term and the object.
 - ▶ Disquotation Principle: " p " is true iff p .
1. My language disquotes.
 2. In BIVese, "brains in a vat" does not refer to brains in a vat.
 3. In my language, "brains in a vat" is a meaningful expression.
 4. In my language, "brains in a vat" refers to brains in a vat. (1,3)
 5. My language is not BIVese. (2,4)
 6. If I am a BIV, then my language is BIVese.
 7. I am not a BIV.

Contents

Introduction

逻辑学

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

伦理理论

1. 如何评判行动?
2. 如何评估目标选择?
3. 如何生成道德上可接受的行动?
 - ▶ **Deontology**: 行动具有内在的伦理价值 (康德主义).
 - ▶ **Asimovian**: 尽可能避免伤害 (通过作为或不作为).
 - ▶ **Utilitarianism**: 最大化总效用.
 - ▶ **Do-no-harm**: 不做任何会导致不良后果的事.
 - ▶ **Do-no-instrumental-harm**: 不做任何会导致不良后果的事, 除非它是一种非预期的副作用.
 - ▶ **Principle of double effect**: 一个行动是 **可接受的** 当且仅当
 1. 行动本身必须是善的、或道德中立的
 2. 主体必须有意达成好的后果, 坏的后果是副效应
 3. 不能以实现目标为手段产生坏的后果
 4. 好的效果必须大于坏的效果, 要权衡利弊, 减少伤害

一些伦理理论的比较

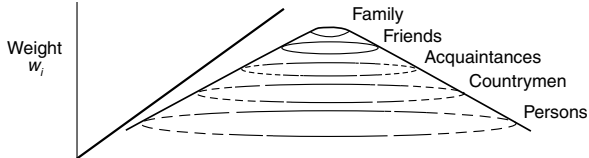
	后果论	道义论	美德伦理学
行动的正 当性	最大化效用	符合道德准则	有德的人会采取的行动
价值导向	善	正当 (履行道德义务)	德性
焦点	后果	行动	动机
核心问题	重要的是结果, 而不是行动	人作为目的, 而不是手段	行为人的品格
实践	大多数人的幸福 (means-ends reasoning)	遵循准则 (rational reasoning)	人的品格 (social practice)
规范	要好报	做好事	做好人

Utility Population Problem



Thanos

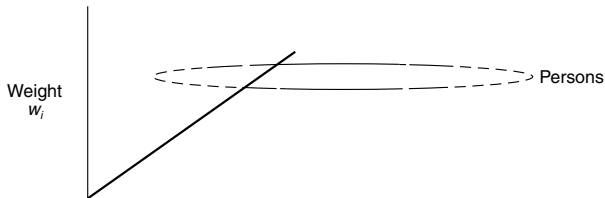
假设灭霸打一个响指, 世界人口会减少一半, 但留下的人及其后代的幸福感都会倍增.



Do most people value higher the well-being of people they know better?



The ethical egoist



合作 vs 信任

- ▶ 组织分别给了你和 Alice 12 元钱.

$$(12, 12)$$

- ▶ 你可以选择把其中的 0/4/8/12 元分给 Alice.
- ▶ 不管你给了 Alice 多少钱, 组织都会再给 Alice 两倍于此的钱.

$$(12 - x, 12 + 3x)$$

- ▶ 至于 Alice 是否愿意与你分享一部分她的钱, 分享多少, 完全取决于她自己.

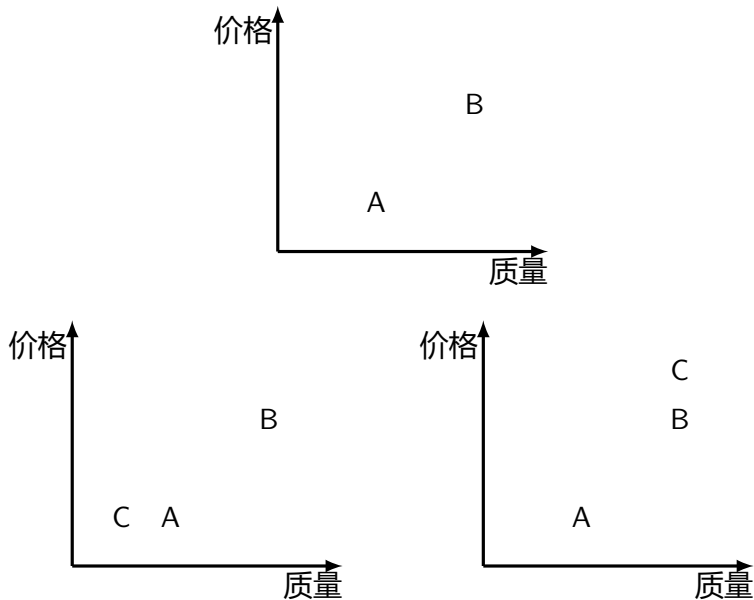
$$(12 - x + y, 12 + 3x - y)$$

- ▶ 你愿意给 Alice 多少钱?

效用有框架效应 — Tversky and Kahneman

1. 某种传染病预计将导致 600 人死亡, 现有两种救助方案:
 - 1.1 方案 A: 200 人获救
 - 1.2 方案 B: $1/3$ 的概率 600 人全都获救
 $2/3$ 的概率无人生还
2. 某种传染病预计将导致 600 人死亡, 现有两种救助方案:
 - 2.1 方案 C: 400 人死亡
 - 2.2 方案 D: $1/3$ 的概率没有人死亡
 $2/3$ 的概率 600 人全都死亡

偏好逆转



平等的世界是合理的吗？

权重的偏倚

- ▶ 设想有一个绝对平等的世界.
- ▶ 这里的公民不会有任何道德上的偏袒, 对所有人都一视同仁.
- ▶ 假如有一人必须面对一个痛苦的选择: 是救他的儿子还是救一个陌生人?
- ▶ 他只能用掷硬币的方式来决定.....
- ▶ 你愿意成为那个世界的公民吗?

平等 vs 公平

Problem (最后通牒博弈)

- ▶ 我出 100 块钱, 供两个人分.
- ▶ 一人负责提议分成比例, 另一人只能选择同意还是拒绝.
- ▶ 只有一次机会.
- ▶ 若同意, 则按比例分; 若拒绝, 则谁也得不到.

Remark: 匿名惩罚 → 利他主义

1. 平等原则: 平均分配
2. 公平原则: 按劳分配或按需分配等方式

Remark: 计划经济更注重平等, 市场经济更注重公平.

荒岛遗言 vs 张三的选择

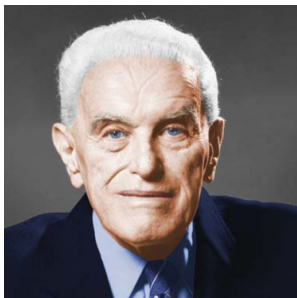
功利主义 vs 契约论

1. 功利主义的代表人物: 休谟 (Hume)、亚当·斯密 (Adam Smith)、边沁 (Bentham)、穆勒 (John Stuart Mill)、西奇威克 (Sidgwick)、埃奇沃思 (Edgeworth) 等...
 - 福利、效率
 2. 契约论的代表人物: 霍布斯 (Hobbes)、洛克 (Locke)、卢梭 (Rousseau)、康德 (Kant)、罗尔斯 (Rawls) 等...
 - 权利、自由
 - 目的的正当性不能证成手段的正当性.
- ▶ 张三和李四被困于一个荒岛上.
 - ▶ 垂死的李四对张三留下遗言: 倘若张三能活着回去, 就将自己的遗产用于建立一个野猫收容所.
 - ▶ 张三答应了.
 - ▶ 但当获救后, 张三认为: 如果把该遗产用于修建孤儿院, 将产生更大福利.

海萨尼⁴ PK 罗尔斯

期望效用最大化 vs 最大化最小原则

1. 假设有 A 和 B 两个感染了严重肺炎的患者, B 还是癌症晚期, 现有的抗体只能救一个人, 应该优先救谁?
2. 在分配某份教育资源时, 假设可以用它让有数学天赋和兴趣的 A 学习数学, 或者用它让严重弱智的 B 学会系鞋带, 应该优先考虑谁?



⁴ John C. Harsanyi: Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory. 1975.

海萨尼: 最大化最小原则能够成为道德的基础吗? — 对罗尔斯理论的批判.

哪一种情况更好?

功利主义 vs 美德伦理学

1. 想象宇宙中只存在一个有感觉的生物, 他错误地相信其他感觉生物正在遭受剧烈的折磨. 这种想法给他带来巨大的愉悦.
2. 想象宇宙中只存在一个有感觉的生物, 他错误地相信其他感觉生物正在遭受剧烈的折磨. 不过他会为受到折磨的同胞而感到悲伤.

功利主义 vs 美德伦理学

卡尔维诺《黑羊》

- ▶ 从前有个国家，人人是贼。晚上，每人都去邻居家行窃。
- ▶ 人们就这样幸福地生活在一起。
- ▶ 某天，有个君子到了该地定居。晚上，他不出门行窃，却呆在家里读书。
- ▶ 贼来了，见灯亮着，就没进去。
- ▶ 这样持续了一段时间。人们感到有必要向他挑明一下，纵使他自己想怎样就怎样，可他没理由妨碍到别人啊。他晚上不出门，就意味着有人第二天饿肚子。
- ▶ 从此君子也晚上出门，但他不行窃。他走到桥上看流水。
- ▶ 不到一星期，君子就被偷的家徒四壁了。
- ▶ 君子不从别人那里偷东西，总有人家里没被动过。不久，那些没有被偷过的人发现自己变富了。而那些跑到君子家里去行窃的人，却发现里面空空如也，于是就变穷了。
- ▶ 富人也想去桥上看流水，也不想再行窃了，他们想：“我们雇那些穷的去替我们行窃吧。”
- ▶ 以免因遭穷人行窃而返贫，富人又雇了穷人中的最穷者来帮助他们看守财富，这就意味着要建立警察局和监狱。
- ▶ 在君子出现后没几年，人们就不再谈什么偷窃或被偷窃了，而只说穷人和富人。
- ▶ 但他们个个都还是贼。
- ▶ 唯一正直的只有开头的那个君子，但他不久便死了，饿死的。

Contents

Introduction

逻辑学

形而上学

认识论

伦理学

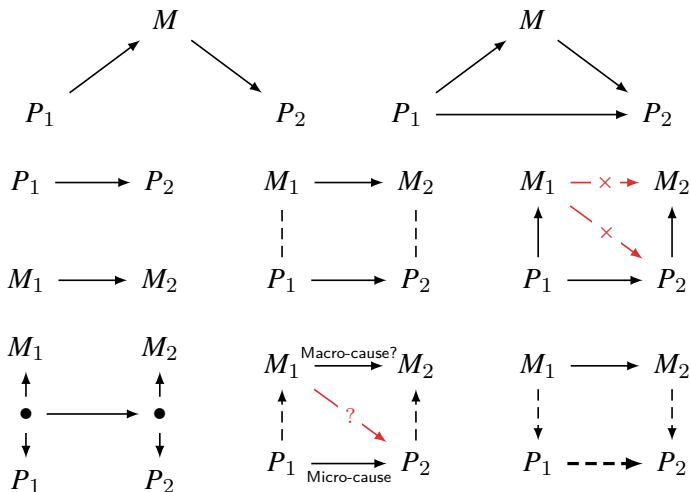
心灵哲学

语言哲学

科学哲学

政治哲学

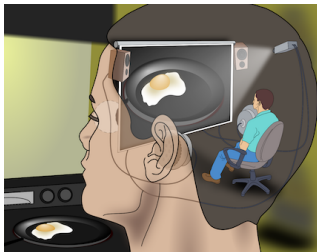
Dualism, (Materialism, Idealism, Neutral) Monism, Interactionism, Preestablished harmony, Pluralism, Epiphenomenalism, Emergentism ...



The Doubt Argument — Dualism

"Cogito, ergo sum."

— *Descartes*



1. I cannot doubt that my mind exists.
2. I can doubt that my body exists.
3. Leibniz's Law: x and y are distinct if they have at least one different property.
4. Therefore, my mind is distinct from my body.

Problem: How could they interact?

Is 'being doubtable' a property?

Shakespeare — Hamlet

What a piece of work is a man!
How noble is reason!
how infinite in faculty!
in form, in moving, how express and admirable!
in action how like an angel!
in apprehension how like a god!
the beauty of the world! 宇宙之精华!
the paragon of animals! 万物之灵长!
And yet, to me, what is this quintessence of dust?
man delights not me;
no, nor woman neither,
though, by your smiling, you seem to say so.

Contents

Introduction

逻辑学

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

语法 vs 语义

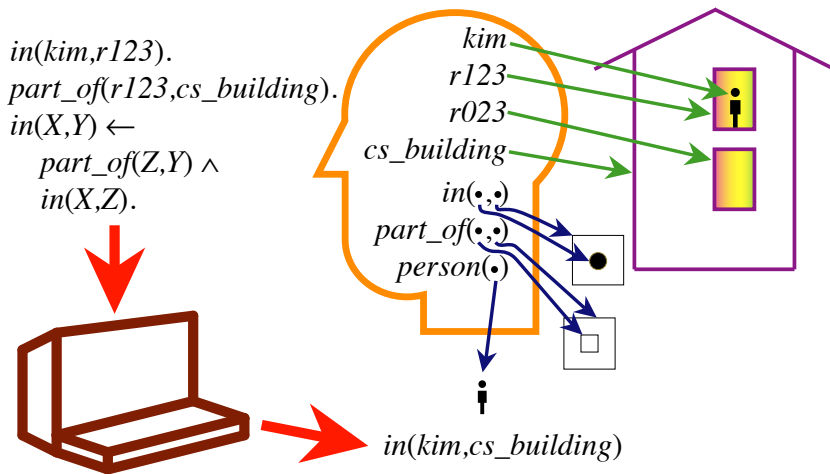


Figure: The computer takes in symbols and outputs symbols. The meaning of the symbols are in the user's head.

语言 — 思想 — 世界

- ▶ What is the meaning of 'meaning'? (symbol grounding problem)
- ▶ How do words relate to objects? thought?
- ▶ What makes a sentence true/false?

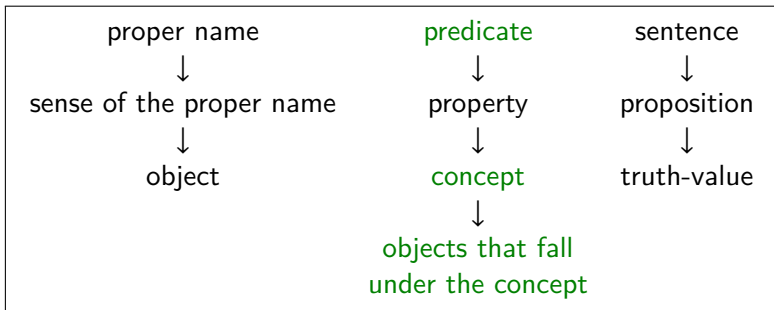


Table: Frege: symbol, sense & reference

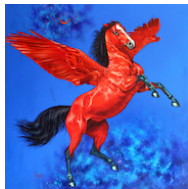
Frege: The meaning of a term is a function/algorithm which computes its denotation.

Wittgenstein: The limits of my language means the limits of my world.

What is the meaning of 'meaning'?

- ▶ Frege: The meaning of a term is a function/algorithm which computes its denotation.
- ▶ Carnap: The intension of an expression is a function from each possible world to the extension of the expression.
- ▶ Wittgenstein: Meaning as (context-dependent) use (family resemblance)

flying horse →



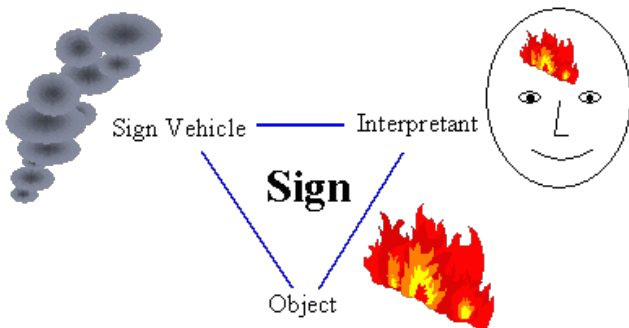
- ▶ The morning star is the evening star.
- ▶ Sherlock Holmes is a detective.
- ▶ The flying horse is not a horse.
- ▶ The round square is round.
- ▶ The golden mountain does not exist.

Charles Peirce 1839-1914

"Men and words reciprocally educate each other."

— Peirce

- ▶ "The mind is a sign developing according to the laws of inference."
- ▶ "Logic is formal semiotic."
- ▶ "Every thought is a sign, taken in conjunction with the fact that life is a train of thought, proves that man is a sign."
- ▶ "All this universe is perfused with signs."



Jerry Fodor 1935-2017 Language of Thought Hypothesis



- ▶ 心灵表征理论: 心理表征由可以赋值为真或假的因果序列构成. 命题态度是主体与心理表征之间的关系.
— $B_i p$ iff there is a 'mental representation' S such that i 'believes' S and S means that p .
- ▶ 心灵计算理论: 思维过程是心理表征的 token 序列组成的计算过程.
- ▶ 思维语言假设: 心理表征既有组合性的语法又有组合性的语义. 思维是基于思维语言的计算.

Remark: 大脑神经元可以表征 (符号) 概念吗? 比如怎么表征 “兔子”?

Logical Positivism

- ▶ Analytic-Synthetic Distinction
 - ▶ analytic sentence = a sentence that is true/false in virtue of its meaning.
 - ▶ synthetic sentence = a sentence that is true/false in virtue of its meaning and how the world actually is.
- ▶ Verifiability Theory of Meaning: The meaning of a sentence consists in its method of verification.
 - ▶ It's too weak! e.g. "All metals expand when heated and the Absolute Spirit is perfect" is verifiable.
 - ▶ It's too strong! e.g. "Superstrings exist" is not verifiable.
- ▶ Observational & Theoretical Languages
- ▶ The Role of Logic: analyze the language of science in terms of logic (Deductive & Inductive).

Problems:

- ▶ Hypotheses cannot be tested in isolation (Duhem-Quine Thesis).
- ▶ Nothing is immune to revision, not even logic (analytic sentences).
 - Move from classical to quantum physics requires analogous move from classical to quantum logic!

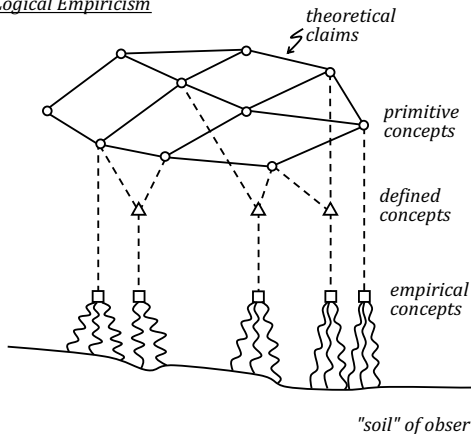
Popper — Falsificationism

- ▶ A theory is falsifiable if it is contradicted by an observation that is expressible in the language of the theory.
- ▶ However, it is models of theories, not the theories themselves, that are tested by experiments.
- ▶ In general, it is possible to falsify a parametric family of models, but impossible to falsify the class of all models of the theory, for it is too large.

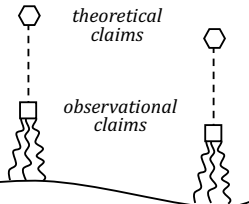
From Logical Positivism to Logical Empiricism

- ▶ Verifiability Theory of Meaning: The meaning of a sentence consists in its method of verification.
- ▶ Holistic Empiricist Theory of Meaning: Theoretical claims about unobservable phenomena gain meaning from their place in the structure of a given theory.

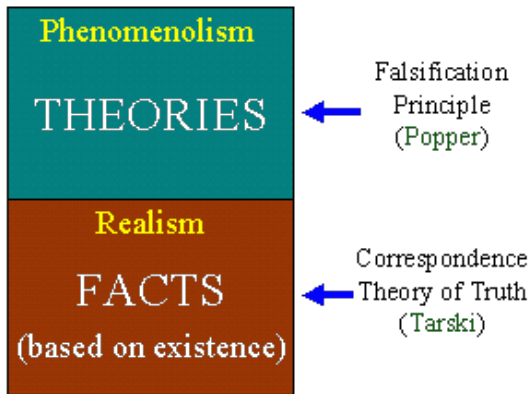
Logical Empiricism



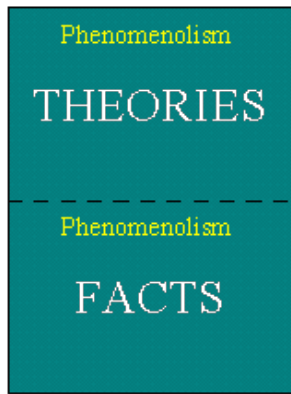
Logical Positivism



Logical Positivism



Pragmatism



- ▶ Pragmatism denies realism not only in the area of theories but also in the area of facts.
- ▶ There is no qualitative difference between facts and theories.

Wittgenstein: Philosophical Investigations

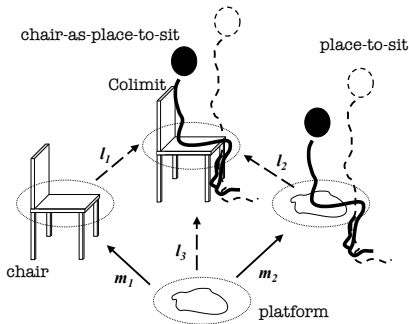
"I shall not today attempt further to define 'pornography'; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it."

— Potter Stewart

- ▶ We cannot define words.
- ▶ Language does not describe facts, it is used to communicate.
- ▶ The meaning of a word is its use in the language.
- ▶ The various uses of words can be best understood as family resemblance.
 - ▶ use *A* is similar to use *B*, because they share trait *X*
 - ▶ use *B* is similar to use *C*, because they share trait *Y*
- ▶ If a word is used in a new context, we draw on the various uses in other contexts.

Wittgenstein: Philosophical Investigations

- ▶ How are chairs identified?
- ▶ chair = 4 legs, back, place to sit, ...
- ▶ All chairs share a “family resemblance” in appropriate contexts of use.
- ▶ This family resemblance can't be formally encoded in a rule/definition.
- ▶ form of life = basic set of practices, behaviors, principles (No external justification.)
- ▶ language game = pattern of linguistic habits associated with a form of life.
- ▶ Language does not represent; rather, it is used by communities to communicate.
- ▶ Terms do not gain meaning by what they represent; rather, they gain meaning by how they are used.



盒子中的“甲虫”

- ▶ 设想每个人都有一个盒子，里面装着被称作“甲虫”的东西。
- ▶ 没有人能看见别人盒子里的东西。
- ▶ 每个人都说：他只是通过看自己盒子里的那只甲虫，才知道什么是“甲虫”。
- ▶ 盒子里的东西可能各不相同。甚至可以设想盒子里的东西在不断变化，或没有东西。
- ▶ 维特根斯坦试图提醒我们：虽然感觉经验是私有的，但不存在私人语言，否则日常交流就无法进行。语言的意义并不在于指向某个内在的私有实体，而在于它在公共语言游戏中的使用方式。

Quine 1908-2000 “Web of Belief”



What does “gavagai” mean?



- ▶ Scientific claims, common beliefs and opinions, are all interconnected in a single unified belief system.
- ▶ Changes in any part of the system can be accommodated by revision elsewhere.
(It confronts experience as a whole.)
- ▶ Indeterminacy of translation

- ▶ Holistic Theory of Meaning: A scientific term gets its meaning from the theory it appears in.
- ▶ There is no single set of standards entitled to govern the justification of beliefs.
- ▶ Justification of a belief system is internal to that system, not external.
- ▶ Scientific theories (facts) are social constructs.

What does “social construct” mean?

To construct X in the social world requires:

- ▶ Knowledge of X encourages behaviors that increase or reduce other people's tendency to act as though X does or does not exist.
- ▶ There is reasonably common knowledge of X
- ▶ There is transmission of knowledge of X .

Philosophy of Language

- ▶ “Classical” view (pre-1953): language consists of sentences that are true/false
- ▶ “Modern” view (post-1953): language is a form of action

Wittgenstein (1953), *Philosophical Investigations*

Austin (1962), *How to Do Things with Words*

Searle (1969), *Speech Acts*

Grice (1975), *Logic and Conversation*

situation

Speaker → *Utterance* → *Listener*

- ▶ Speech acts achieve the speaker's goals.
- ▶ Speech act planning requires knowledge of
 - Situation
 - Semantic and syntactic conventions
 - Listener's goals, knowledge base, and rationality

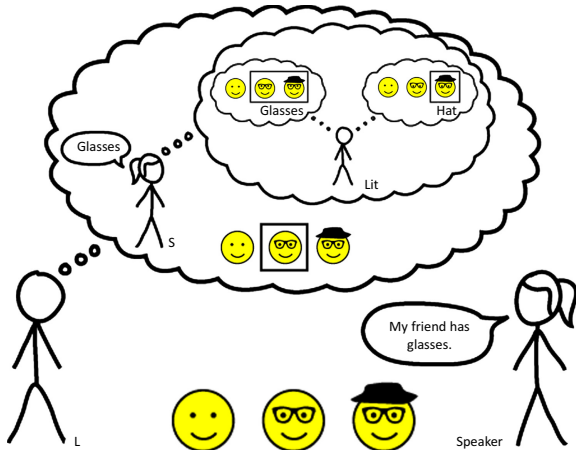
Stages in Communication

Intention	Speaker S wants to inform Listener L that meaning m
Generation	Speaker S chooses proposition p such that Listener L is most likely to infer m given p
Synthesis	Speaker S utters proposition p
Perception	Listener L perceives p
Analysis	Listener L infers possible meanings $m_1, \dots, m_n$
Disambiguation	Listener L infers intended meaning m_i
Incorporation	Listener L incorporates m_i into KB

Engaging in complex language behavior requires various kinds of knowledge of language

- ▶ **Linguistic knowledge:** Phonetics, phonology, Morphology, Syntax, Semantics, Pragmatics, Discourse
- ▶ **World knowledge:** common knowledge, commonsense knowledge

庄子: 言者所以在意, 得意而忘言?



$$p = \arg \max_p P_L(m \mid p)$$

$$P_L(m \mid p) \propto P_S(p \mid m)P(m)$$

$$P_S(p \mid m) \propto \exp(\alpha \cdot U(p, m)) \quad \alpha \text{ is a parameter}$$

$$U(p, m) = \log P_{\text{Lit}}(m \mid p) - \text{Cost}(p)$$

$$P_{\text{Lit}}(m \mid p) = \chi_{m \in \llbracket p \rrbracket} P(m) \quad \text{"informative" to the Literal Listener}$$

格赖斯 (Paul Grice) 的语用会话的“合作原则”

1. 数量原则 (Quantity: be informative, don't undershare or overshare)
— 提供对方所需的信息, 不少也不多
2. 质量原则 (Quality: be truthful, don't say what you don't believe)
— 不说假话, 不说没证据的话
3. 关联原则 (Relation: be relevant)
— 不说无关的话, 不答非所问
4. 方式原则 (Manner: be clear)
— 避免晦涩、避免歧义、简洁、有条理

Remark: Grice 的“合作原则”体现在效用函数 $U(p, m)$ 里.

- ▶ 数量原则、关联原则 $P_{\text{Lit}}(m \mid p)$
- ▶ 质量原则、关联原则 $\chi_{m \in \llbracket p \rrbracket}$
- ▶ 方式原则 $\text{Cost}(p)$

Contents

Introduction

逻辑学

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

科学哲学对科学有用吗？

科学哲学对科学家的用处跟鸟类学对鸟的用处差不多.

— 费曼

The Unreasonable Ineffectiveness of Philosophy

- ▶ 费曼：“砖头算不算本质客体？”
- ▶ 哲学家甲：“一块砖是独特的砖，是怀海德所说的本质客体。”
- ▶ 哲学家乙：“本质客体的意思并不是指个别的砖块，而是指所有砖块的共有的普遍性质，换句话说，‘砖性’才是本质客体。”
- ▶ 哲学家丙：“不对，重点不在砖本身，‘本质客体’指的是，当你想到砖块时内心形成的概念。”
- ▶ 就像所有关于哲学家的故事一样，最终以一片混乱收场。好笑的是，在先前的那么多次讨论中，他们从来没有问过自己，像简单的砖块究竟是不是“本质客体”。

— 费曼

哲学旨在感动那些混淆晦涩与深刻的人。

— 温伯格

我们已经有了有一套最完美的基于证明的哲学，那就是数学。

— 邓煜

The Unreasonable Ineffectiveness of Philosophy

- ▶ 如果一个哲学家说的东西是对的, 那一定是微不足道的; 如果不是微不足道时, 那一定是错的. — 高斯
- ▶ 哲学难道不是用蜜写成的吗? 乍一看, 很精彩, 再一看, 除了一团浆糊, 什么都没留下. — 爱因斯坦
- ▶ 哲学之于科学, 就像手淫之于性: 它廉价、易实现, 甚至有些人也更喜欢它.
- ▶ 哲学家就像瞎子, 在漆黑的夜里, 举着熄灭的蜡烛, 跑到黑暗的地下洞穴, 找寻一只并不存在的黑猫, 然后大呼小叫: “我脖子被猫爪挠啦!”
- ▶ 哲学可以作为跨学科研究的“催化剂”或“调味品”. 但“哲学内部”问题像是智力“黑洞”, 吸收大量智力, 却无任何输出. — *Johan van Benthem*
- ▶ 哲学偶尔有用, 能让我们免受另外一些哲学立场的影响. — 温伯格

Contents

Introduction

逻辑学

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

科学划界

科学解释

政治哲学

如何明确划分“科学”与“非科学”(特别是“伪科学”、“形而上学”和“宗教”)的界限?

科学划界问题

问题: 如何明确划分“科学”与“非科学”(特别是“伪科学”、“形而上学”和“宗教”)的界限?

- ▶ 逻辑实证主义: 可证实性标准
- ▶ 波普尔: 可证伪性标准
- ▶ 库恩: 范式下的解谜
 - 占星术之所以不是科学, 不是因为它不能做出预言, 而是因为它没有一个能够指导从业者去分析失败原因、逐步解决谜题的范式和累积性研究传统.
- ▶ 拉卡托斯: 科学研究纲领
 - 硬核与保护带. 对保护带的修改能带来新预测就是进步的纲领, 为了应付反例而不断做事后解释, 无法预测新事实就是退化的纲领.
- ▶ 费耶阿本德: 怎么都行. 划界企图是一种“科学沙文主义”.
- ▶ 劳丹等人的消解论与多元主义. 一刀切是伪问题. 没有单一标准, 但有多元特征: 信仰权威教条, 缺乏自我纠正机制, 过度依赖事后解释, 选择性接受证据, 隔离于主流科学...

Contents

Introduction

逻辑学

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

科学划界

科学解释

政治哲学

预测 vs 解释

- ▶ 科学是为了预测还是为了解释?
- ▶ 预测本身就是目标? 亦或重要的是理论的解释力, 而预测仅仅是理论可证伪性的需要?
 - 一个能精准预测未来, 但无法解释 '为什么' 的 AI, 算不算真正的智能?

科学解释

1. Hempel 的覆盖律模型: 定律提供解释

定律 初始条件
现象

2. Salmon 的因果模型: 原因提供解释

原因 $\xrightarrow{\text{机制}}$ 结果

3. Kitcher 的统一性模型: 理论提供解释⁵

一个理论具有统一性, 是指它能最大化适应广度、简洁性、严格性.

- ▶ 适应广度衡量其可推导出的结论数量.
- ▶ 简洁性衡量的是作为前提假说的公理集合的大小.
- ▶ 严格性衡量的是其应用范围的边界.

⁵统一性模型用理论替代了覆盖律模型中的“定律”, 强调全局的系统整合性.

覆盖律解释

问题: 为什么滑冰运动员把双臂收向身体时会旋转得更快?

角动量守恒定律: $L_i = L_f$

运动员不与外物作用: $L_i = I_i \omega_i, L_f = I_f \omega_f$

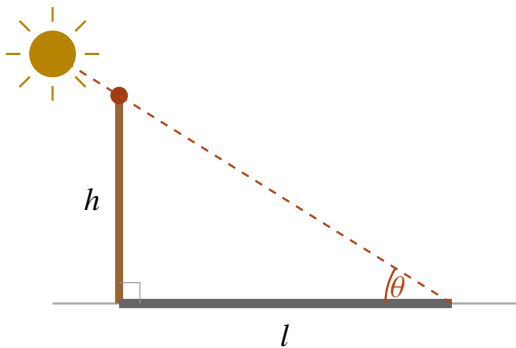
运动员有非零的初始角动量: $L_i > 0$

运动员收紧手臂时, 质点到旋转轴的距离 r 变小, 转动惯量 $I = mr^2$ 变小: $I_f < I_i$

运动员的旋转角速度变大: $\omega_f > \omega_i$

覆盖律解释的“对称性”问题

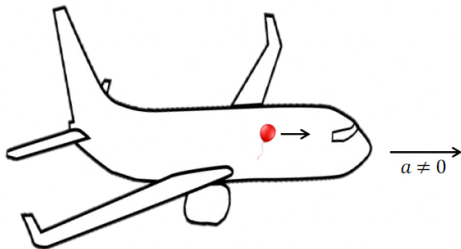
- ▶ **问题:** 为什么旗杆影子的长度为 l ?
- ▶ 因为太阳的高度角为 θ , 光线沿直线传播, 旗杆的高度为 h , 所以影子的长度为 $l = h \cot \theta$.



- ▶ 但是, 同理, 我们也可以用影子的长度 l 解释旗杆的高度 $h = l \tan \theta$.
- ▶ 然而, 我们通常只会用原因解释结果, 而不会用结果解释原因.
- ▶ 所以, 旗杆高度对影子长度的解释是一种因果解释.

因果机制解释

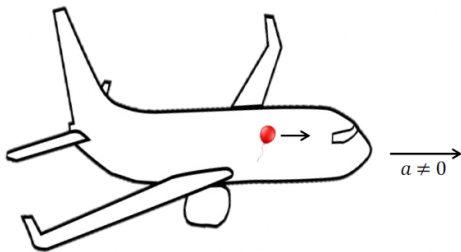
- ▶ **问题:** 飞机起飞时, 机舱内的氦气球往哪个方向飘?



- ▶ 飞机加速, 空气分子由于惯性往后积聚, 后方气压高, 前方气压低, 氦气球往前飘.

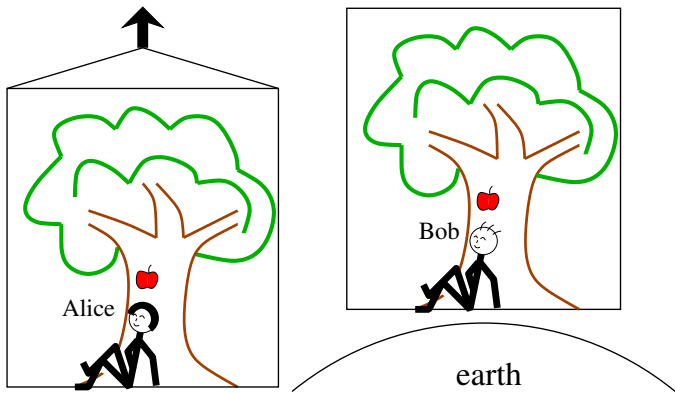
统一性理论解释

- ▶ **问题:** 飞机起飞时, 机舱内的氦气球往哪个方向飘?



- ▶ 根据广义相对论等效原理, 一个局域匀加速参考系在物理效果上与处于反向均匀引力场中的静止参考系不可区分. 飞机向前加速时, 加速度产生的向后等效引力场与地球原本向下的重力场叠加, 在机舱内部形成了一个指向斜后下方的合成等效引力场. 因此, 氦气球会向斜前上方飘浮.

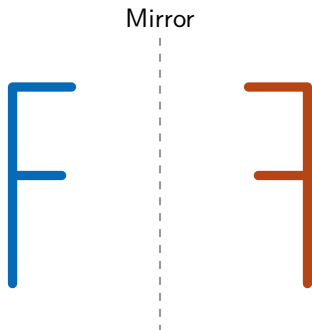
爱因斯坦的匀加速电梯思想实验



- ▶ 通过思想实验模拟的加速度感官体验与地球上的重力感官体验不可区分, 爱因斯坦溯因出等效原理: 引力场与惯性力场局域不可区分.

某些统一性理论解释未必显式的涉及因果机制

- ▶ **问题:** 为什么左手的手套无法戴到右手上?
- ▶ 因为, 在三维欧式空间中, 左手手套无法通过刚性运动, 变换到右手.⁶



⁶An object is an incongruent counterpart of another if they cannot be made to occupy the same place by rigid motions in a local region of space. Whether or not a mirror image is an incongruent counterpart depends on the properties of the space it is located in. If the space is 3-dimensional, or 2-dimensional but non-orientable, then "F" and its mirror image are congruent counterparts (e.g. a Möbius strip).

Contents

Introduction

逻辑学

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

Contents

Introduction

逻辑学

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

机制设计与道德风险

不完全信息博弈与声誉

演化博弈与自发秩序

民主投票与群体聚合

零知识证明与隐私保护

Cake Cutting

- ▶ Two Children:
 1. One cuts, the other chooses.
 - ▶ Three Children (A,B,C):
 1. A cuts a piece of the cake.
 2. B can leave it, or make it smaller.
 3. C can leave the remaining piece, or make it smaller.
- Whoever cuts the cake last, gets the remaining piece.

- ▶ 19 世纪中期法国的火车有头等座、二等座和三等座三种不同的车厢.
- ▶ 三种车厢票价相差较大.
- ▶ 头等车厢非常舒适, 二等车厢与三等车厢的区别是前者有顶盖, 后者没有. 这样, 坐三等车厢的旅客就要忍受日晒雨淋的痛苦.
- ▶ 给车厢加一个顶盖的成本微乎其微, 为什么公司不这样做呢?
- ▶ 为了吓唬富人, 故意让穷人受罪.

- ▶ 问题: 有时候, 拥有私人信息的一方有积极性通过一定的行动向另一方传递自己的私人信息, 但有时候他们没有积极性或没有有效的办法传递自己的私人信息.
- ▶ 机制设计: 没有私人信息的一方通过设计不同的分配方案使得有私人信息的一方通过自我选择揭示自己的私人信息.

Example: 保险市场中的混同均衡和分离均衡.

1. 保险金 2 万元; 只有第二年得病才可得到赔偿金 10 万元, 第一年得病得不到赔偿.
2. 保险金 7 万元; 无论第一年得病还是第二年得病, 都可得到赔偿金 10 万元.

高风险者选择合同 2; 低风险者选择合同 1.

逆向博弈论

已知理性人会选择最优策略, 为了诱导出某种类型的行为, 应该设计什么样的博弈?

博弈 vs 伦理

- ▶ 康德之流的义务论者可能会认为, 博弈对伦理无益. 因为, 伦理规范是约束你做那些你不想做的事, 而博弈论却是讲怎么得到你想要的东西.
- ▶ 休谟等效用主义者却认为, 除非伦理理论建议的行为符合每个人的实际利益, 否则毫无用处. 有效的制度设计, 就是通过对纳什均衡的选择实现帕累托最优.

拍卖

1. 高价格公开拍卖
2. 降价公开拍卖
3. 高价格密封拍卖
4. 次价格密封拍卖: 与高价格密封拍卖一样, 中标者仍然是出价最高者, 但中标者实际支付的是第二高报价. 次价格密封拍卖能让竞标者有积极性说真话.

Theorem

次价格密封拍卖中, 说真话是竞标者的占优策略.

Remark: 比如: 对某古董, 你的实际评价是 1 万, 如果出价 1 万, 第二高出价 9 千, 你赚 1 千; 如果你出价低于 9 千, 你什么也得不到.

VCG 机制 Vickrey-Clarke-Groves Mechanism

- 公共产品的偏好显示: 不同的人有不同的偏好, 是私人信息. 如何让每个人报告自己的真实偏好?
- 每个人可以任意地报告自己的偏好, 但可能要纳缴一定数量的“税”. 计算办法: 先将其他人的偏好加总, 给出总价值最大的项目; 然后将第一个人的偏好加上, 如果不影响结果, 不征税; 否则, 应纳税等于改变结果给其他人带来的损失.

$$p_i = \underbrace{\max_x \sum_{j \neq i} v_j(x)}_{i \text{ 不参与}} - \underbrace{\sum_{j \neq i} v_j(x^*)}_{i \text{ 参与}} \quad \text{where} \quad x^* = \operatorname{argmax}_x \sum_{i=1}^n v_i(x)$$

- 同学聚会去吃川菜还是粤菜? 每人报告自己的真实偏好是占优策略.

	川菜	粤菜	税额 p_i
A	30	10	0
B	0	40	30
C	20	10	0
合计	50	60	30

公平与夏普利值

- ▶ A,B,C 合作开了一家公司. 公司任何决策只有超过 50% 的赞成票才能通过.
- ▶ 根据投资额, A 拥有 50% 的票力, B 拥有 40%, C 拥有 10%.
- ▶ 年终净赚 150 万, 根据票力之比, A,B,C 分别 75 万, 60 万, 15 万吗?
- ▶ C 提议: A 拿 80 万, 自己 70 万, B 一分没有.
- ▶ B 提议: A 拿 85 万, 自己 65 万, C 一分没有.
- ▶ 怎么分才公平?
- ▶ 三人中, 任何一人都不是决定性的, 都得与他人结盟才有决策权.
- ▶ 夏普利值的基本假设是: 各投票顺序联盟形成的可能性相同, 玩家的夏普利值为其对联盟的边际贡献之和除以各种可能的联盟组合总数.

1	A	A	B	B	C	C
2	B	C	A	C	A	B
3	C	B	C	A	B	A
关键加入者	B	C	A	A	A	A

Table: 所有可能的投票顺序

- ▶ 作为关键加入者的次数. A: 4 次; B: 1 次; C: 1 次.
- ▶ 夏普利值. A: $\frac{4}{6}$; B: $\frac{1}{6}$; C: $\frac{1}{6}$.
- ▶ 根据影响力, A: $150 \times \frac{4}{6} = 100$ 万; B: 25 万; C: 25 万.

讨价还价与纳什讨价还价解

- ▶ 市场上有一个画家 A 和一个画廊 B.
- ▶ 若画家 A 自己卖画, 可得 $a = 1500$.
- ▶ 若交给画廊 B 卖, 可得 $V = 6000$.
- ▶ 若画廊 B 干别的事, 可得 $b = 1000$.
- ▶ AB 该怎么分利?
- ▶ $S := V - a - b = 3500$ 是合作带来的剩余价值.
- ▶ 我们用 x 表示 A 应分得的价值, y 表示 B 应分得的价值. α, β 表示 A 和 B 剩余价值 S 的分配比例 (与边际贡献有关).

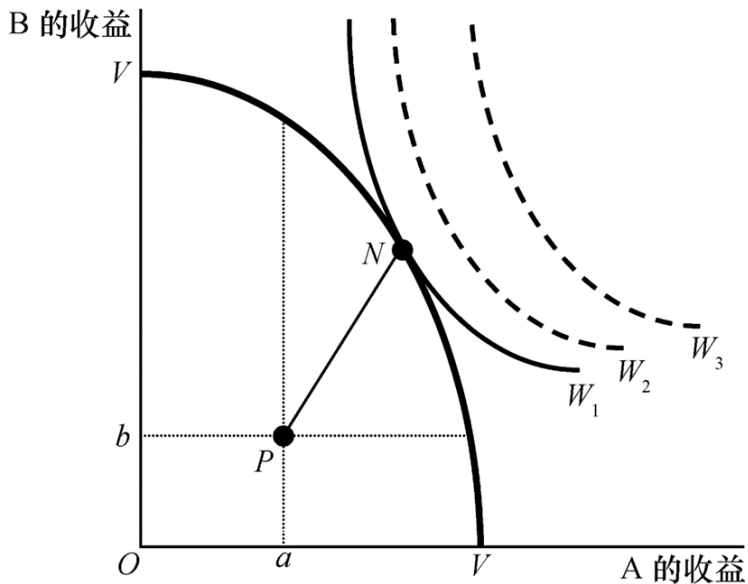
$$x = a + \alpha S, \quad y = b + \beta S$$

- ▶ 纳什讨价还价解即优化福利函数 $W(x, y) = (x - a)^\alpha (y - b)^\beta$

$$(x^*, y^*) = \operatorname{argmax}_{x, y: x+y=V} (x - a)^\alpha (y - b)^\beta$$

- ▶ 因为若缺少任何一方, 这 3500 的剩余价值都得不到, 所以每一方的边际贡献都是 3500. 总边际贡献是 $3500 + 3500 = 7000$. 每一方的边际贡献占比都是 $\frac{3500}{3500+3500} = \frac{1}{2} = \alpha = \beta$.
- ▶ A 应分得 $1500 + \frac{1}{2} \times 3500 = 1500 + 1750 = 3250$; B 应分得 2750.

纳什讨价还价解



谈判中的影响力 — 不可替代性

- ▶ 如果市场上还有一家画廊 C, 若 A 交给画廊 C 卖, 可得 $V' = 5000$.
- ▶ 若画廊 C 干别的事, 可得 $c = 500$.
- ▶ AB 又该怎么分利?
- ▶ AB 合作中 A 的边际贡献是 $V - a - b = 3500$.
- ▶ 若 AB 不合作, 则 AC 合作, 因为 B 的存在, C 的边际贡献为 0, A 拿走所有剩余 $V' - c$. 即 A 的市场价值不是 a 而是 $V' - c$.
- ▶ 所以, AB 合作中 B 的边际贡献是
$$V - (V' - c) - b = 6000 - (5000 - 500) - 1000 = 500.$$
- ▶ A 的边际贡献占总边际贡献的 $\frac{3500}{3500+500} = \frac{7}{8}$, B 占 $\frac{1}{8}$.
- ▶ A 应分得: $1500 + \frac{7}{8} \times 3500 = 4562.5$.
- ▶ B 应分得: $1000 + \frac{1}{8} \times 3500 = 1437.5$.

Remark: 大牌影星天价片酬.

腐败 vs 信息

- ▶ 腐败的深层根源: 信息不对称;
- ▶ 特别是事后的信息不对称: 一方当事人的行为不能被另一方观察到;
- ▶ 研究行为信息不对称的理论叫“委托-代理”理论或道德风险理论.

委托-代理关系

- ▶ 法律上的委托 - 代理关系: 如果甲乙两人达成一个协议, 甲将做某事的权利交给乙, 甲为委托人, 乙为代理人.
- ▶ 代理人对委托人的责任: (1) 没有许可, 不能再代理; (2) 不能把自己放在与委托人利益冲突的地位; (3) 保密责任和诚信责任.
- ▶ 委托人对代理人的责任: (1) 补偿责任. 委托人给代理人补偿报酬; (2) 免除法律责任. 委托人要为代理人的行为承担法律责任; (3) 留置权. 如果委托人给代理人的补偿没有到位, 代理人有权滞留委托人的财产.
- ▶ 经济上的委托 - 代理关系要宽松很多: 只要一方行为影响另一方的利益, 就有委托 - 代理关系. 有私人信息的一方是代理人, 没有私人信息的称为委托人.

委托-代理问题

- ▶ 委托人与代理人可能的利益冲突：对委托人最优的选择不一定是对代理人最优的选择。
- ▶ 如果委托人能观察到代理人的行为，则可以通过强制性合同约定约束代理人的行为。
- ▶ 即使不能完全观察到代理人的行为，如果代理人不害怕风险，“承包制”可以使代理人成为完全的风险承担者。
- ▶ 但如果代理人承担责任的能力有限，“承包制”没有可行性。

对代理人的激励机制的设计

- ▶ 保险与激励的冲突
 - ▶ 最优风险分担: 如果委托人是风险中性的, 代理人是风险规避的, 风险应该完全由委托人承担, 代理人拿固定报酬.
 - ▶ 最优激励: 代理人应该承担完全风险.
 - ▶ 这就产生了保险与激励的冲突.
 - ▶ 最优激励合同要在保险与激励之间求得平衡.
 - 汽车的部分保险 (如 80%).
- ▶ 激励的强度
 - ▶ 代理人的边际生产率越高, 激励应该越强.
 - ▶ 代理人的产出的不确定性越大, 或测度越困难, 激励应该越弱.
 - ▶ 代理人越风险规避, 激励应该越弱.
 - ▶ 代理人对激励的反应程度: 反应越强, 激励应该越强.
- ▶ 相对业绩比较有助于改进激励, 但也可能导致代理人之间合谋或相互拆台.
- ▶ 论功行赏 vs 任人唯贤
 - 论功行赏可能导致每个人都被晋升到他不能胜任的位置
- ▶ 多重任务下的激励: 大学教学 + 科研

政府官员的激励

- ▶ 很多很多委托人;
- ▶ 多项任务;
- ▶ 业绩难以度量, 投入也不容易度量;
- ▶ 所以, 政府官员难以激励, 只能以监督为主.
- ▶ 但监督是不完全的, 是有成本的.
- ▶ 权利越大, 越难监督.
- ▶ “把权力关进笼子”. 民主、法治.

官员腐败问题

- ▶ W : 官员工资.
- ▶ $B(q)$: 权力租金, 即官员在位置上可能收受的贿赂, 也可看作不腐败的机会成本. 一般来说, 权力 q 越大, 权力租金越高.
- ▶ p : 腐败行为被发现暴露的概率.
- ▶ F : 对腐败的处罚.
- ▶ U : 政府外的保留效用.
- ▶ α : 官员的羞耻感或脸皮厚度.
- ▶ 官员腐败的期望收益

$$(1 - p)(W + B(q)) + p(U - \alpha F)$$

- ▶ 官员不腐败的条件:

$$W \geq \frac{1 - p}{p} B(q) + U - \alpha F$$

- ▶ 如何控制腐败? 提高 p, F, W, α , 减少 q .

Contents

Introduction

逻辑学

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

机制设计与道德风险

不完全信息博弈与声誉

演化博弈与自发秩序

民主投票与群体聚合

零知识证明与隐私保护

狐狸、权力与不完全信息

- ▶ 小女孩提着一篮子鸡蛋穿过树林.
- ▶ 喜欢抢鸡蛋吃的狐狸跳了出来: “我是狐狸!”
- ▶ 小女孩: “怎么证明你是狐狸?”
- ▶ 狐狸: “我有厚厚的皮毛.”
- ▶ 小女孩: “兔子也有厚厚的皮毛.”
- ▶ 狐狸: “我有毛茸茸的尾巴.”
- ▶ 小女孩: “松鼠也有毛茸茸的尾巴.”
- ▶ 狐狸一直试图证明自己, 不知不觉跟着小女孩走出了树林.
- ▶ 一只猎犬经过吓跑了狐狸.
- ▶ 狐狸边跑边喊, “不信你问猎犬, 我真的是狐狸!”
- ▶ 小女孩: “我知道, 我早就知道.”

	攻击	不攻击
防卫	-2, -2	0, 0
不防卫	-1, 1	0, 0

Table: 与狐狸博弈

	攻击	不攻击
防卫	0, -2	0, 0
不防卫	-1, 1	0, 0

Table: 与兔子/松鼠博弈

Remark: 权力需要被认可, 如果得不到认可, 权力也将不复存在. 如果有人试图用权力吓唬你, 你应该提出质疑, 让对方承担举证的责任.

声誉的积累

$$\begin{aligned} P(\text{君子} \mid \text{好事}) &= \frac{P(\text{君子} \wedge \text{好事})}{P(\text{好事})} \\ &= \frac{P(\text{好事} \mid \text{君子})P(\text{君子})}{P(\text{好事} \mid \text{君子})P(\text{君子}) + P(\text{好事} \mid \neg \text{君子})P(\neg \text{君子})} \end{aligned}$$

- ▶ 人不知而不愠, 不亦君子乎
- ▶ 勿以善小而不为, 勿以恶小而为之
- ▶ 狼来了
- ▶ 烽火戏诸侯
- ▶ “伪君子” 为什么更可恨?

Contents

Introduction

逻辑学

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

机制设计与道德风险

不完全信息博弈与声誉

演化博弈与自发秩序

民主投票与群体聚合

零知识证明与隐私保护

演化博弈 — 婚姻博弈

	物质型	感情型
物质型	1, 1	0, 0
感情型	0, 0	2, 2

- ▶ 假定总人口中, 物质型的比例为 x , 感情型的比例为 $1 - x$.
- ▶ 对任何一个个体而言, 物质型的期望效用: $x1 + (1 - x)0 = x$.
- ▶ 感情型的期望效用: $x0 + (1 - x)2 = 2(1 - x)$.
- ▶ 如果 $x > 2/3$, 物质型更适合生存, 将演化成稳定均衡.
- ▶ 如果 $x < 2/3$, 感情型更适合生存, 将演化成稳定均衡.
- ▶ 如果 $x = 2/3$, 两类人有同样的适应性, 但此二元均衡是非稳定的.

演化博弈 — 鹰鸽博弈

	鹰	鸽
鹰	-1, -1	1, 0
鸽	0, 1	$\frac{1}{2}, \frac{1}{2}$

- ▶ 假定鹰派的比例是 x , 鸽派的比例是 $1 - x$.
- ▶ 鹰派的效用: $-1x + 1(1 - x) = 1 - 2x$.
- ▶ 鸽派的效用: $0x + \frac{1}{2}(1 - x) = \frac{1}{2}(1 - x)$.
- ▶ 如果 $x < 1/3$, 鹰派占优势, 不稳定.
- ▶ 如果 $x > 1/3$, 鸽派占优势, 不稳定.
- ▶ 如果 $x = 1/3$, 同样的适应性, 稳定.
- ▶ 如果初始人口由单一类型构成, 另一类型可以成功入侵, 直到均衡.
- † 两个纯策略均衡: (鹰、鸽), (鸽、鹰); 一个混合均衡: $(1/3, 2/3)$
- △ 假定存在某种显性的标记机制: 在博弈开始之前, 每个人收到一个信号: A 或 B ; 概率是 $1/2$; 信号完全负相关; 标记是公共知识.
- 1. 如果 A , 选择“鹰”; 如果 B , 选择“鸽”. 是 ESS.
- 2. 如果 A , 选择“鸽”; 如果 B , 选择“鹰”. 是 ESS.
- 3. 无论 AB , 以 $1/3$ 的概率选择“鹰”, $2/3$ 的概率选择“鸽”. 不是 ESS.

自发秩序与产权制度

- ▶ 社会秩序是所有人行为选择的结果, 但不是集体理性集中设计的结果, 而是自发演化的结果.
- ▶ 社会规范是演化稳定策略, 但不一定是帕累托最优的.
- ▶ 虽然不是集体理性集中设计的产物, 但“制度企业家”扮演了重要角色. (哲学王)
- ▶ 改变游戏规则, 进行制度创新. 创新意味着让人们用新的价值观念代替旧的价值观念、用新的行为方式代替旧的行为方式、用新的是非观和新的善恶观代替旧的是非观和旧的善恶观, 意味着我们要认同原来可能不认同的东西或不再认同我们原来认同的东西.
— 儒家: “仁、义、礼、智、信”, “己所不欲勿施于人”, “以德报德, 以直报怨”, “君君臣臣父父子子”.
- ▶ 制度企业家面临的不确定性和风险挑战异于常人. 他们不能以“利”和“名”为目的, 买他们账的“客户”可能在遥远的未来.
- ▶ “自私的”基因 & 迷因 (meme)

演化稳定策略 Evolutionarily Stable Strategy

- ▶ 静态: 一个特定的行为模式被称为是演化稳定的, 如果它的种群不能被变异所入侵, 或者说, 任何偏离行为模式的个体具有更低的生存能力, 种群将会恢复到原来的状态.
- ▶ 动态: 假定初始状态存在多样的行为模式, 随着时间的推移, 如果某个特定的行为模式能逐步主导整个种群, 这个特定的行为模式就是演化稳定的.

Contents

Introduction

逻辑学

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

机制设计与道德风险

不完全信息博弈与声誉

演化博弈与自发秩序

民主投票与群体聚合

零知识证明与隐私保护

群体聚合 Group Aggregation

提案:

- A. 大学教育应该免费.
- B. 人人都应该有上大学的权利.
- C. 政府应该增加在大学教育上的投入.

$$\frac{A \quad B}{C}$$

民意调查:

	A	B	C
张三	1	1	1
李四	0	1	0
王五	1	0	0
群体聚合	1	1	1/0

个体理性: 直接对提案 C 进行投票, 个体自我一致.

集体理性: 先对理由 A, B 分别投票, 再推导结论 C.

Remark: 通过安排投票方式, 可以操纵选举结果.

孔多塞投票悖论

张三	$a > b > c$	$\xrightarrow{\text{多数原则}}$	$a > b > c > a$
李四	$b > c > a$		
王五	$c > a > b$		

Remark: 如果你可以制定投票规则, 那么, 你想让谁胜出, 你可以通过“两两 PK”的方式让其胜出.

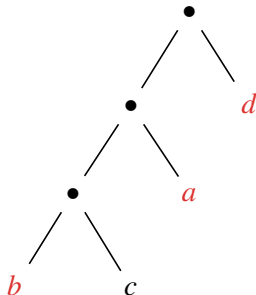
	$a > b$	$b > c$	$a > c$
张三	1	1	1
李四	0	1	0
王五	1	0	0
群体聚合	1	1	1/0

孔多塞投票的群体聚合

序贯多数投票

张三	$a > b > c > d$
李四	$b > c > d > a$
王五	$c > d > a > b$

- ▶ 所有人都认为 $c > d$
- ▶ d 还能胜出吗?



投票制度

1. **相对多数法 (plurality method)**: 选民每人投一个候选人
2. **双选投票法 (vote-for-two method)**: 选民每人投两个候选人
3. **反相对多数法 (anti-plurality method)**: 选民每人投 $n - 1$ 个候选人, 即投票剔除一个候选人
4. **绩点评分法 (GPA method)**: 选民对所有 n 个候选人进行排序
 - ▶ 第一名得 $n - 1$ 分
 - ▶ 第二名得 $n - 2$ 分
 - ▶ 倒数第二名得 1 分
 - ▶ 最后一名 0 分

GPA 得分最高者当选.

5. **排序选择法 (ranked-choice voting)**: 选民对所有 n 个候选人进行排序
 - 5.1 首轮计票时, 只计算选民的首选票. 如果有候选人获得过半数选票, 该候选人当选.
 - 5.2 如果没有候选人过半数, 就将票数最少的候选人淘汰, 把选择该候选人为首选的选票, 按次选票的候选人重新分配.
 - 5.3 重复第 2 步, 直到某个候选人获得过半数选票为止.

选举结果反映的究竟是民意还是选举方法的选择？

2 个选民	$a > b > c > d$
2 个选民	$a > d > c > b$
2 个选民	$c > b > d > a$
3 个选民	$d > b > c > a$

Table: 9 个选民, 4 个候选人

- ▶ 根据“相对多数法”， a 赢: $a4 > d3 > c2 > b0$
- ▶ 根据“双选投票法”， b 赢: $b7 > d5 > a4 > c2$
- ▶ 根据“反相对多数法”， c 赢: $c9 > b7 = d7 > a4$
- ▶ 根据“绩点评分法”， d 赢: $d15 > b14 > c13 > a12$

当有候选人退出时会怎样?

3 个选民	$a > c > d > b$
6 个选民	$a > d > c > b$
3 个选民	$b > c > d > a$
5 个选民	$b > d > c > a$
2 个选民	$c > b > d > a$
5 个选民	$c > d > b > a$
2 个选民	$d > b > c > a$
4 个选民	$d > c > b > a$

Table: 30 个选民, 4 个候选人

- ▶ 根据“相对多数法”, a 赢: $a9 > b8 > c7 > d6$
- ▶ 如果 d 退出: $c11 > b10 > a9$
- ▶ 如果 c 退出: $d11 > b10 > a9$
- ▶ 如果 b 退出: $d11 > c10 > a9$
- ▶ 如果 a 退出: $d12 > c10 > b8$
- ▶ 根据“绩点评分法”, d 赢: $d58 > c54 > b41 > a27$
- ▶ 根据“排序选择法”, c 赢: $c20 > b10$

Arrow's Impossibility Theorem

Theorem (Arrow's Impossibility theorem)

In an election with candidates ≥ 3 , any voting system that is unanimous and independence of irrelevant alternatives must be a dictatorship!

- ▶ unanimous: if everyone agrees, consensus decides the outcome
- ▶ independence of irrelevant alternatives: a voter can't move a above b by lying about how they feel about c
- ▶ nondictatorship: no single voter gets to decide the outcome

对无关选项独立性 IIA 的质疑

排序	张三	李四	王五	赵六
1	x	x	a	a
2	a	a	b	b
3	b	b	y	y
4	y	y	x	x

排序	张三	李四	王五	赵六
1	a	a	y	y
2	b	b	a	a
3	x	x	b	b
4	y	y	x	x

- ▶ 在两个表中, x 和 y 的相对排序是一样的. 张三和李四偏好 $x > y$, 王五和赵六偏好 $y > x$.
- ▶ 直观上, 表 1 中偏好 $x > y$ 的人比偏好 $y > x$ 的人的“感受更为强烈”; 表 2 则相反. 我们可能会说, 表 1 中 $x > y$ 是有道理的, 而表 2 中 $y > x$ 是有道理的.
- ▶ IIA 要求两表中社会对 x 和 y 的偏好一样: 要么 $x > y$, 要么 $y > x$.

社会选择理论 Social Choice Theory

- ▶ 计算社会选择理论: 用计算复杂性防止坏情况的发生
- ▶ 用逻辑做投票协议验证
 - ▶ 隐私性: 没有别人能知道你投的是谁
 - ▶ 无收据性: 你不能证明给别人你投了特定人的票
 - ▶ 可核查性: 你自己能检查你的票是不是被算进去了
 - ▶ 公平性: 之前投票的部分结果不会影响之后投票的结果

Contents

Introduction

逻辑学

形而上学

认识论

伦理学

心灵哲学

语言哲学

科学哲学

政治哲学

机制设计与道德风险

不完全信息博弈与声誉

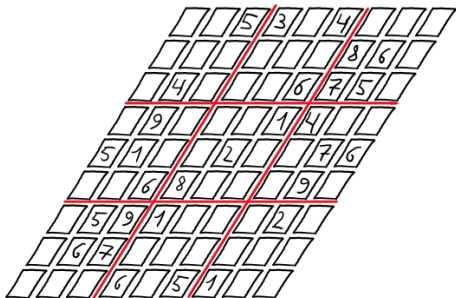
演化博弈与自发秩序

民主投票与群体聚合

零知识证明与隐私保护

零知识证明 Zero-Knowledge Proof

- ▶ 小帅给小美出了一道非常难的数独题.
- ▶ 小美怎么也解不出来, 怀疑小帅在耍她: “这道题无解!”
- ▶ 小帅说: “我会用零知识证明的方法给你证明这题有解. 我不会把解给你看, 却能让你信服我确实有这题的解.”



提示: 准备九个袋子 😊

Thank 