

《行为经济学》复习笔记

VectorPikachu
08 January 2026

1. 课程逻辑与总结

行为经济学建立的逻辑: 从对传统经济学“理性人”假设的批判出发, 系统性地引入心理学洞见来解释现实世界中人类的非理性行为, 并最终探讨如何利用这些知识来设计更好的政策和干预措施 (即“助推”).

行为偏差 (Behavioral Deviations) 的分类:

- 非标准偏好 (Non-standard Preferences)
 - 现时偏见 (Present Bias)
 - Exponential vs. Quasi-Hyperbolic Discounting (β, δ)
 - $U_t = u_t + \beta\delta u_{t+1} + \beta\delta^2 u_{t+2} + \dots$
 - Sophisticates vs Naïve. ¹
 - Commitment Devices
 - 参考依赖 (Reference Dependence)
 - Expected Utility Theory (EUT): $EU = \sum_{i=1}^n p_i u(x_i)$
 - Prospect Theory: EUT + Initial Wealth + Reference Point + Risk Attitude. ²
 - 社会偏好 (Social Preferences)
 - Outcome-based, Image-based, Intention-based.
- 非标准信念 (Non-standard Beliefs)
 - 过度自信 (Overconfidence)
 - Overestimation, Overplacement, Overprecision.
 - 小数定律 (Law of Small Numbers)
 - 投射偏差 (Projection Bias)
- 非标准决策过程 (Non-standard Decision-Making)
 - 由于认知资源有限, 人们在做决策时会依赖启发式 (经验法则), 但这些启发式常常导致系统性错误.
 - 有限注意力 (Limited Attention)
 - 框架效应 (Framing Effects)
 - 菜单效应 (Menu Effects)

助推 (Nudge)

- 默认选项 (Default Options)
- 社会传染 (Social Contagion)
- 目标设置 (Goal Setting)
- 承诺装置 (Commitment Devices)
- 噪音 (Noise): 指在专业判断中存在的、不受欢迎的随机变异.
- 反助推 (Sludge): 指那些故意或无意地使有益行为变得困难的摩擦或障碍 (例如, 繁琐的申请表、复杂的条款).

行为金融 (Behavioral Finance).

Contents

1. 课程逻辑与总结	1
2. Non-standard Preferences	2
2.1. Present Bias	2
2.2. Reference Dependence	4
2.3. Social Preferences	6
3. Non-standard Beliefs	9
3.1. Overconfidence	9
3.2. Law of Small Numbers	11
3.3. Projection Bias	11
3.4. Application	12
4. Non-standard Decision-Making	12
4.1. Limited Attention	13
4.2. Framing Effects	15
4.3. Menu Effects	15
5. Nudge	16
5.1. Default	17
5.2. Social Contagion	18
5.3. Goal Setting and Commitment Devices	18
5.4. Noise and Sludge	20
6. Behavioral Topics	21
6.1. Behavioral Economics for Environmental Issues	21
6.2. Behavioral Finance	23
6.3. Health and Development Issues	25
6.4. Behavioral Bias in Relationship	25
Bibliography	27

¹Sophisticates ($\hat{\beta} = \beta < 1$), Fully naïve ($\beta < \hat{\beta} = 1$), Partially naïve ($\beta < \hat{\beta} < 1$)

² Prospect Theory 公式:

$$U_{PT} = \sum_{i=1}^n w(p_i) v(\Delta_i),$$
$$\Delta_i = (w_0 + x_i) - r,$$
$$v(\Delta) = \begin{cases} \Delta^\alpha, & \Delta \geq 0 \\ -\lambda(-\Delta)^\beta, & \Delta < 0 \end{cases}$$
$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{\frac{1}{\gamma}}}$$

2. Non-standard Preferences

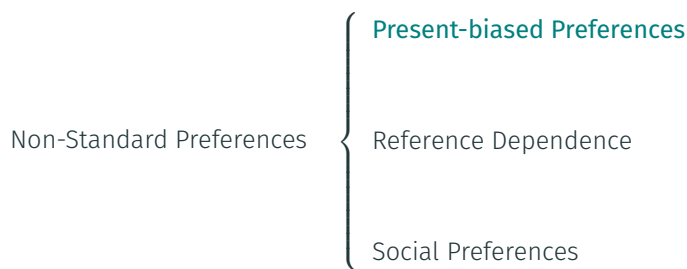


Figure 2.1: Non-standard Preferences

2.1. Present Bias

2.1.1. Present Bias

如今我们面临着很多跨时间决策 (Intertemporal Choices), 例如储蓄、锻炼、饮食等。传统经济学假设人们是理性的, 会根据贴现因子 (Discount Factor) 来权衡当前和未来的效用³。然而, 行为经济学发现人们往往表现出时偏见 (Present Bias), 即过度重视当前的利益而忽视未来的利益。

Exponentially discounted utility. $\Rightarrow$ Time consistent.⁴

Present Bias. 人们会对于“今天”有一个额外的偏好。Quasi-hyperbolic discounting.

- *Present bias refers to the tendency of individuals to place disproportionately high weight on immediate costs and benefits, while undervaluing outcomes that occur in the future.*

$$U_t = u_t + \beta \delta u_{t+1} + \beta \delta^2 u_{t+2} + \dots$$

Leisure goods: immediate rewards, delayed costs.⁵

$\Rightarrow$ People *over-consume* leisure goods relative to their long-run plans.

Investment goods: immediate costs, delayed benefits.⁶

$\Rightarrow$ People *under-consume* investment goods (like exercise) relative to their long-run plans.

下面是一个关于**动态不一致 (Dynamic Inconsistency)**的例子。

Amy's Assignment

Amy has to finish her assignment in one of three periods: $t = 0, 1, 2$. Parameters are $\beta = \frac{1}{2}, \delta = 1$. And the instantaneous dis-utilities (costs of doing the work):

$$u_0 = -1, u_1 = -\frac{3}{2}, u_2 = -\frac{5}{2}.$$

Case 1: Commitment Available:

- At $t = 0$, Amy can commit in advance which period to do the assignment.
 - ▶ 在 $t = 0$ 的时候做作业, $U_0^{t=0} = -1$;
 - ▶ 在 $t = 1$ 的时候做作业, $U_0^{t=1} = \beta u_1 = \frac{1}{2} \times \left(-\frac{3}{2}\right) = -\frac{3}{4}$;
 - ▶ 在 $t = 2$ 的时候做作业, $U_0^{t=2} = \beta u_2 = \frac{1}{2} \times \left(-\frac{5}{2}\right) = -\frac{5}{4}$;
- 所以 Amy 会选择在 $t = 1$ 的时候做作业。

Case 2: No Commitment:

³ $\delta \in (0, 1]$. u_t 表示今天做的事情获得的收益, u_{t+1} 表示明天做的事情获得的收益...

$$U_t = u_t + \delta u_{t+1} + \delta^2 u_{t+2} + \dots$$

⁴ 所谓的时间一致性 (Time Consistency), 指的是一个人在不同时间点做出的决策不会相互矛盾。也就是说, 如果一个人是在时间点 t 做出了某个决策, 那么在未来的时间点 $t+1$ 也会坚持这个决策。

⁵ 例子: 吃冰淇淋。

- Immediate utility benefit: $B_{\text{pleasure}} = 2$
- Delayed cost: $C_{\text{health}} = 3$
- 这里取 $\delta = 1, \beta = 0.5$

如果下个星期吃冰淇淋, 那么:

$$\begin{aligned} U_{\text{nextweek}} &= \beta(B_{\text{pleasure}} - C_{\text{health}}) \\ &= 0.5(2 - 3) = -0.5 < 0 \end{aligned}$$

所以不会在下个星期吃冰淇淋。

如果今天吃冰淇淋, 那么:

$$\begin{aligned} U_{\text{today}} &= B_{\text{pleasure}} + \beta(-C_{\text{health}}) \\ &= 2 + 0.5(-3) = 0.5 > 0 \end{aligned}$$

会选择今天吃冰淇淋。

⁶ 例子: 去健身房锻炼。

下个星期: $\beta(-C_{\text{effort}} + B_{\text{health}}) > 0$, 今天: $-C_{\text{effort}} + \beta B_{\text{health}} < 0$.

- At $t = 0$, Amy cannot commit in advance.
 - 根据上面的计算, 在 $t = 0$ 的时候 Amy 计划在 $t = 1$ 的时候做作业.
- However, at $t = 1$, she re-evaluates:
 - 在 $t = 1$ 的时候做作业, $U_1^{t=1} = u_1 = -\frac{3}{2}$;
 - 在 $t = 2$ 的时候做作业, $U_1^{t=2} = \beta u_2 = \frac{1}{2} \times \left(-\frac{5}{2}\right) = -\frac{5}{4}$;
- So at $t = 1$, Amy will choose to do the assignment at $t = 2$.

Amy's preferences are dynamically inconsistent.

2.1.2. Naïveté and Sophistication

Naïve people procrastinate again and again, while sophisticated people look for ways to commit their future selves.

我们记 $\hat{\beta}$ 为个体对自己未来贴现因子的预测值.

- Sophisticates: $\hat{\beta} = \beta < 1$.
- Fully naïve: $\beta < \hat{\beta} = 1$.
 - 这些人坚信自己未来会变得完全理性.
 - Over-optimistic: "This time I will exercise/save/work..."
 - End up procrastinating again and again.
- Partially naïve: $\beta < \hat{\beta} < 1$.

如何估计 δ, β 和 $\hat{\beta}$?

- **Task 1 (long-run patience):** 在 100 天获得 1 美元 vs 101 天后获得 2 美元.⁷
- **Task 2 (present bias):** 在今天获得 1 美元 vs 明天获得 2 美元.⁸
- 如果这两项的结果不同, 则说明存在现时偏见.
- **Task 3: Ask people to predict their future behavior.**
 - 明天你更喜欢获得 1 美元还是继续等待以获得 2 美元?⁹

⁷估计出 δ .

⁸估计出 β .

⁹估计出 $\hat{\beta}$.

2.1.3. Applications: Commitment Devices

Type	Description	Example
Financial Incentives 财务激励	Monetary rewards or penalties tied to goal achievement	Performance-based bonuses or salary increments
Public Declarations 公开声明	Openly stating goals or intentions to create social pressure	Sharing team objectives in company-wide meetings
Contractual Agreements 合同协议	Formal documents outlining commitments and consequences	Signing a professional development plan with specific milestones
Technology-based Tools 基于技术的工具	Apps or software that track progress and enforce commitments	Project management tools with deadline alerts and progress tracking
Peer Accountability 同行问责	Partnering with colleagues to mutually support goal achievement	Forming accountability groups for skill development

Table 2.1: Commitment Devices

Fertilizer Puzzle

Puzzle: around the world, farmers often under-use fertilizer despite very high returns.

- 标准解释: credit constraints, risk, knowledge. 但是并不充分.
- 行为经济学的解释: farmers intend to buy fertilizer, but procrastinate due to present bias [1].

- 解决方案: 提供 commitment devices, e.g., SAFI Program = Savings and Fertilizer Initiative.
 - ▶ Offered right after harvest (when farmers have cash).
 - ▶ Farmers could buy fertilizer voucher at normal price, with free delivery later.
 - ▶ Immediate decision required $\Rightarrow$ reduces procrastination.

2.2. Reference Dependence

2.2.1. Expected Utility Theory

In economics, a *prospect* (or lottery) means a risky choice.

在标准模型中, 人们使用期望效用理论 (Expected Utility Theory, EUT) 来评估风险选择。¹⁰

Extension I: Risk Attitude. 效用函数的形状决定了个体的风险态度。

- risk neutral: $u(x) = x$;
- risk averse: concave $u(x)$, e.g., $u(x) = \sqrt{x}$;
- risk loving: convex $u(x)$, e.g., $u(x) = x^2$.

¹⁰EUT 公式: $EU = \sum_{i=1}^n p_i u(x_i)$, 其中 p_i 是结果 x_i 发生的概率, $u(x_i)$ 是结果 x_i 的效用。

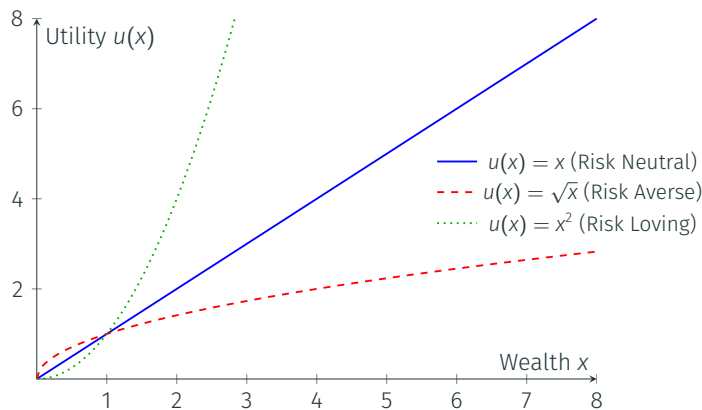


Figure 2.2: Risk Attitude

那么什么决定了风险态度呢?

1. 财富水平. 穷人更倾向于风险厌恶, 富人更倾向于风险偏好.
2. 偏好与个性. 有些人天生更喜欢冒险, 有些人则更喜欢稳定.
3. 情境与环境 (Context and stakes). 在不同的情境下, 人们的风险态度可能会发生变化.

Extension II: Risk Premium. 个人为规避风险而愿意支付的最高金额。¹¹

如何计算风险溢价:

1. 计算出期望收益 $EU = \sum_{i=1}^n p_i u(x_i)$.
2. 计算出 Certainty equivalent (CE), 即使个体无差异的确定金额. 计算方法: $u(CE) = EU$.
3. 计算风险溢价: $RP = \sum_{i=1}^n p_i x_i - CE$.

Extension III: Overweighting Small Probabilities. 人们往往会高估小概率事件的发生概率, 这被称为“小概率事件的过度权重化”。¹²

所以我们将 p_i 换成一个 probability weighting function $w(p_i)$. 那么:

¹¹保险 (Insurance): 人们购买保险是因为风险溢价大于保险费 $\Rightarrow$ 愿意支付额外费用以避免不确定的损失。

金融投资 (Financial investments): 高风险的资产必须提供更高的预期回报以吸引投资者, 这就是所谓的风险溢价。

劳动力市场 (Labor markets): 高风险的工作通常会提供更高的薪酬作为补偿, 这也是一种风险溢价。

¹²例子: 彩票. 尽管中奖概率非常低, 但人们仍然愿意花钱购买彩票, 因为他们高估了中奖的可能性。

$$U = \sum_{i=1}^n w(p_i)u(x_i),$$

其中 $w(p)$ 满足 $w: [0, 1] \rightarrow [0, 1]$, 且 $w(0) = 0, w(1) = 1$. 通常 w 是一个 inverse-S-shaped 函数, 也即:

$$\underbrace{w(p) > p}_{\text{overweight small } p} \quad \text{for small } p, \quad \underbrace{w(p) < p}_{\text{underweight large } p} \quad \text{for large } p.$$

Extension IV: Initial Wealth. 人们一开始有一个初始财富水平 w .

2.2.2. Prospect Theory

前景理论通过四个关键特征更准确地描述了人类在不确定条件下的真实决策行为:

1. **参考点依赖 (Reference Dependence).**
2. **损失厌恶 (Loss Aversion).** 损失比获益有大约 2 倍的心理影响.
3. **敏感性递减 (Diminishing Sensitivity).** 获益时, risk aversion; 损失时, risk seeking.¹³
4. **概率加权 (Probability Weighting).** Overweight small probabilities, underweight large probabilities.¹⁴

$$U_{PT} = \sum_{i=1}^n w(p_i)v(\Delta_i),$$

$$\Delta_i = (w_0 + x_i) - r,$$

$$v(\Delta) = \begin{cases} \Delta^\alpha, & \Delta \geq 0 \\ -\lambda(-\Delta)^\beta, & \Delta < 0 \end{cases}$$

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{\frac{1}{\gamma}}}.$$

其中, $\lambda > 1$, 因为 loss aversion. $0 < \alpha, \beta \leq 1$ 则说明 diminishing sensitivity. 其中:

$$\gamma^+ \approx 0.61 \quad (\text{for gains}), \quad \gamma^- \approx 0.69 \quad (\text{for losses}).$$

Certainty Effect: a strong preference for sure outcomes.

禀赋效应 (Endowment Effect): 由于损失厌恶, 人们对自己拥有的物品估值 (WTA, 愿意接受的最低价格) 远高于未拥有时的估值 (WTP, 愿意支付的最高价格). $WTA \gg WTP \Rightarrow \text{Endowment Effect}$.

2.2.3. Application

正常来说, 每小时的工资越高, 劳动力供给就越多.¹⁵

但是在纽约的出租车行业中, 研究人员发现: 驾驶员对每小时工资的反应并不完全如标准劳动力供给理论所预测的那样, 同时也受到每日收入目标的影响, 这与前景理论的建议相吻合 [2]

Observed Behavior:

- On high-wage days, many drivers quit early.
- On low-wage days, many drivers work longer.

¹³ 一个 value function 如下:

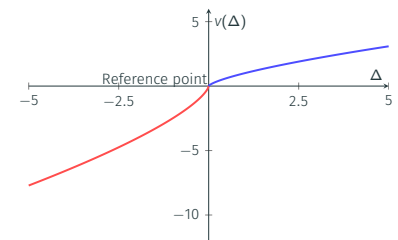


Figure 2.3: Value Function

¹⁴ 下面是一个 w :

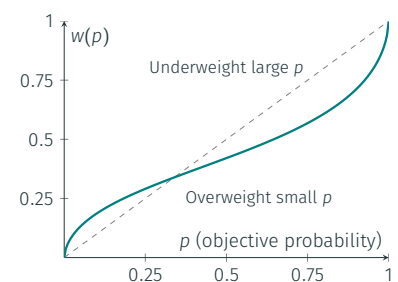


Figure 2.4: Probability Weighting

¹⁵ 也就是说, 劳动力供给和 wage per hour (w) 的关系如下:

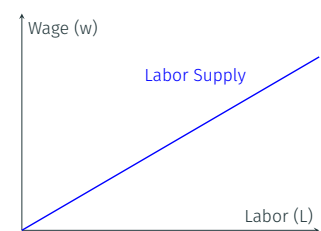


Figure 2.5: 劳动力供给和 wage per hour (w) 的关系

- Hours and wages are negatively correlated. (按照标准理论, 应该是正相关的.)

解释: 司机们设定了每日收入目标 (Daily income targets).

2.3. Social Preferences

经典模型: 人们只关心自己, 经济人 (*Homo Economicus*).

现实中却不尽如此, 人们会关心 fairness, reciprocity (互惠), altruism (利他), and envy.

Key question: How do social concerns shape economic choices and outcomes?

- Outcome-based** (distributional). 人们关心最终的分配结果是否公平, 而不仅仅是自己的绝对收益.
 - Model 1:** Fehr and Schmidt.¹⁶
 - Model 2:** Charness and Rabin.¹⁷
- Image-based** (face-saving / social image). 人们关心自己的行为如何被他人 (或自己) 所看待, 即关心自己的“社会形象”或“面子”.
 - Intrinsic + Extrinsic.¹⁸
- Intention-based** (reciprocity, procedural justice). 人们对结果的评价不仅看结果本身, 更看重对方行为背后的意图.
 - own payoff + reciprocity term.¹⁹

2.3.1. Outcome-based Preferences

Model 1: Fehr and Schmidt.

$$U_i = \pi_i - \alpha_i \max\{\pi_j - \pi_i, 0\} - \beta_i \max\{\pi_i - \pi_j, 0\}.$$

- π_i 是 Player i 的收益, π_j 是另一个人的收益.
- $\alpha_i > 0$ 表示不利的不平等的负效用 (envy).
- $\beta_i \geq 0$ 表示有利的不平等的负效用 (guilt).
- 通常, $\alpha_i > \beta_i$, 因为嫉妒比愧疚更强烈.

Extension I: 可以画出无差异曲线.²⁰

Extension II: Many-Player Model.

$$U_i = \pi_i - \frac{\bar{\alpha}_i}{n-1} \sum_{j \neq i} \max\{\pi_j - \pi_i, 0\} - \frac{\bar{\beta}_i}{n-1} \sum_{j \neq i} \max\{\pi_i - \pi_j, 0\}.$$

Normalization by $(n-1)$: each comparison is averaged over all other players.

Model 2: Charness and Rabin.

- 人们不仅关注不平等, 还关注总的社会福利 (overall social welfare) 和最少的收益 (the least well-of).

$$U_i = (1 - \lambda_i) \pi_i + \lambda_i [\delta_i \cdot \min\{\pi_1, \pi_2, \dots, \pi_n\} + (1 - \delta_i) \cdot \Pi],$$

where Π is the total payoff.

- λ_i : weight on social welfare (vs. own payoff).

¹⁶ 两个玩家的公式为:

$$U_i = \pi_i - \alpha_i \max\{\pi_j - \pi_i, 0\} - \beta_i \max\{\pi_i - \pi_j, 0\}.$$

¹⁷ 公式为:

$$U_i = (1 - \lambda_i) \pi_i + \lambda_i [\delta_i \cdot \min\{\pi_1, \pi_2, \dots, \pi_n\} + (1 - \delta_i) \cdot \Pi],$$

where Π is the total payoff.

¹⁸ 公式为:

$$U_i(a; \tau) = b(a) + r(\tau)a - c(a) + q\theta S(\mu(a)).$$

¹⁹ 公式为:

$$U_i(a_i, a_j) = \pi_i(a_i, a_j) + f_i(a_j) \cdot \pi_j(a_i, a_j),$$

²⁰ 无差异曲线如下:

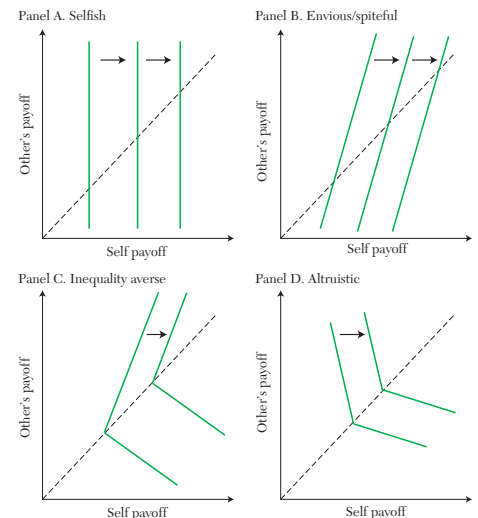


Figure 2.6: Indifference Curves for Different Types of Distributional Preferences

- δ_i : concern for the least well-off vs. total surplus.
- Explains altruism, charity, and solidarity in experiments.

2.3.2. Image-Based Social Preferences

	Self-oriented	Social-oriented
Intrinsic	• Inner satisfaction • Warm-glow, curiosity • Example: donating feels good even if nobody knows	• Caring directly about others' welfare • Altruism, empathy • Example: helping a stranger in need
Extrinsic	• Self-image: maintain positive identity • Avoid feeling selfish or immoral • Example: donate so I see myself as "good"	• Social image: reputation, face, esteem • Care how others perceive me • Example: donate more when name is published

Table 2.2: Intrinsic vs. Extrinsic × Self vs. Social

Basic Idea: Utility = material payoff + value of social image.

$$U_i = \pi_i + \theta_i \cdot S(a_i),$$

where $S(a_i)$ is the social esteem/reputation generated by action a_i .

More detailed mechanism:

$$U_i(a; \tau) = b(a) + r(\tau)a - c(a) + q\theta S(\mu(a)).$$

- $b(a), c(a)$: material benefit/cost from action a .
- Intrinsic motivation (depends on type τ): $r(\tau)a$.
- Extrinsic motivation: $q\theta S(\mu(a))$.
 - q : 被观察的概率.
 - $S(\mu(a))$: 社会形象函数, 取决于行为的社会认知.
 - θ : 个体对社会形象的敏感度.

行为的**可见性**(q)至关重要. 当行为公开($q = 1$)时, 人们会表现得更亲社会(如捐更多钱); 当匿名($q = 0$)时, 则主要受内在动机驱动.

应用: 通过公开表彰、同行比较(如家庭用电报告)等方式利用社会形象压力来促进亲社会行为.

挤出效应 (Crowding-out): 引入金钱激励有时会适得其反. 例如, 为献血付费可能降低献血量, 因为这会将“高尚的善举”解读为“为钱做事”, 从而损害了捐赠者的社会形象和内在动机.

2.3.3. Intention-Based Social Preferences

Core idea: People care about why others act, not just the outcomes.

Basic Idea: Utility = own payoff + reciprocity term (互惠项).

$$U_i(a_i, a_j) = \pi_i(a_i, a_j) + f_i(a_j) \cdot \pi_j(a_i, a_j),$$

- $\pi_i(a_i, a_j)$: Player i 's material payoff.
- $f_i(a_j)$: kindness of player j as perceived by i .²¹
- Reciprocity: j is kind, i rewards j ; j is unkind, i punishes j .

纳什均衡: 任何一位玩家在此策略组合下单方面改变自己的策略 (其他玩家策略不变) 都不会提高自身的收益。

²¹ Kindness of player j towards i :

$$f_i(a_j) = \frac{\pi_i(a_i, a_j) - \frac{1}{2}[\pi_i^{\max}(a_j) + \pi_i^{\min}(a_j)]}{\pi_i^{\max}(a_j) - \pi_i^{\min}(a_j)}.$$

- $\pi_i^{\max}(a_j)$: best payoff j could give i .
- $\pi_i^{\min}(a_j)$: worst payoff j could give i .

Ultimatum Game: Derive the Acceptance Threshold

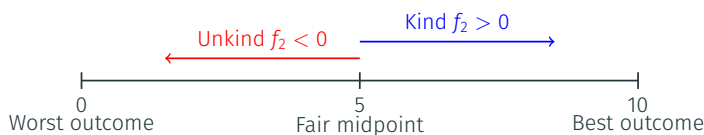
Two players split \$10.

- Player 1 offers \$ x to Player 2.
- Player 2 chooses Accept or Reject.
- If Reject: both get 0.

Find the minimum \$ x that will be accepted.

此处的 Player 1 的动作 a_1 可以记为 x (the amount offered to Player 2).
Player 2 的动作 a_2 可以记为 Accept (A) 或 Reject (R).

首先, $\pi_2^{\max} = 10$, $\pi_2^{\min} = 0$. 那么 Fair midpoint = 5.



那么我们就推导出了:

$$f_2(x) = \frac{x - 5}{10}.$$

If Player 2 accepts:

$$U_2(x, A) = x + f_2(x) \cdot (10 - x) = x + \frac{x - 5}{10} \cdot (10 - x).$$

If Player 2 rejects:

$$U_2(x, R) = 0.$$

那么 Player 2 会接受当且仅当 $U_2(x, A) \geq U_2(x, R)$:

$$x + \frac{x - 5}{10} \cdot (10 - x) \geq 0 \Rightarrow x \geq 2.2.$$

Lesson: Very low offers are rejected, not for money, but because they signal *unkind intentions*.

最后的结果如下:

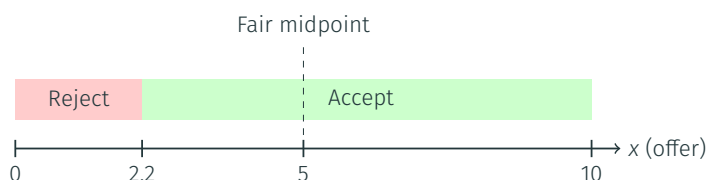


Figure 2.7: Ultimatum Game Results

2.3.4. Application

Public Shaming as a Pressure. (公众羞辱)

- 人们面临采取“负责任”或“爱国”行动的压力，因为他们的选择被放在了大众视野下。
- 例子：抵制、在线“点名文化”、募捐活动。
- 一战期间英国的“白羽女郎”向非军装男性分发白羽，将其污名化为懦夫。²²

3. Non-standard Beliefs

Beliefs = 关于不确定事件的主观概率。

Preferences vs. Beliefs:

- Preferences: 如何评价不同结果 (utilities). 我们喜欢什么样的结果。
- Beliefs: 不同结果发生的可能性 (probabilities). 我们预计什么样的结果会发生。

Non-standard beliefs: beliefs systematically deviate from reality.

过度积极的信念. positively biased beliefs.

3.1. Overconfidence

过度自信 (Overconfidence) 是指个体对自己能力、判断或控制结果的信心超过了实际情况. 主要有三种形式:

1. 高估 (Overestimation): 高估自己实际能力或绩效. E.g., 创业者低估创业失败率; 散户过度交易。
2. 优胜错觉 / 高于平均效应 (Overplacement): 认为自己比他人更优秀. E.g., 大多数司机认为自己比平均水平更好。
3. 过度确定性 (Overprecision): 对自己判断的准确性过于自信. E.g., 过窄的置信区间. 在风险决策中, 因低估了不确定性而做出错误选择 (如接受一个实际风险很高的项目)。

3.1.1. Overestimation

Why do too many people enter competitive markets?

标准理论: 在利润 = 0 时就会停止进入。

然而现实却是: average profit < 0.

- 设有两个动作: $a \in \{0, 1\}$, invest ($a = 1$) or not ($a = 0$).
 - $a = 1$ 时, Payoff: success $\Rightarrow$ prize $R > 0$; failure $\Rightarrow$ sunk cost $C > 0$. 这个成本是无论如何都要付出的。
 - $a = 0$ 时, Payoff = 0.
- 市场上真的成功的概率为 $p \in (0, 1)$.
- 投资者有过度自信: $\tilde{p} = p + \delta, \delta > 0$.

那么, 投资者认为的收益为:

$$\Pi^{belief}(a = 1) = \tilde{p}R - C, \quad \Pi^{belief}(a = 0) = 0.$$

从而投资者会选择投资当且仅当:

$$\tilde{p}R - C \geq 0 \Rightarrow \tilde{p} \geq \frac{C}{R}.$$

那么就会存在一段过度投资的区间:



Figure 2.8: The White Feather: A Sketch of English Recruiting

$$\tilde{p} \geq \frac{C}{R} > p.$$

这一段区间内, 投资者会错误地认为投资是有利可图的, 但实际上却是亏损的.

$$\Pi^{real}(a=1) = pR - C < 0.$$

这一段区间的福利损失为:

$$\mathcal{L} = C - pR > 0.$$

3.1.2. Overplacement: The “Better-than-Average” Bias

我们假设:

- 所有人的能力服从正态分布 $a_i \sim N(\mu, \sigma^2)$.
- 那么理论上来说, 一半的人应该高于平均水平, 一半的人应该低于平均水平.
- 但是我们发现:

$$\Pr(a_i > \mu)^{belief} > 0.5.$$

Most individuals believe they are above average.

现在假设一个进入锦标赛的场景:

- 如果他成功了, 也即 a_i 超过了某个获胜的阈值 a^* , 那么他获得奖励 V ; 否则没有任何奖励.
- 进入锦标赛的成本为 C .
- 那么个体认为的收益为:

$$EU_i^{belief}(\text{Enter}) = \Pr(a_i > a^*)^{belief} \cdot V - C = \tilde{p}_i V - C,$$

$$EU_i^{belief}(\text{Out}) = 0.$$

- 个体会选择进入锦标赛当且仅当:

$$\tilde{p}_i V - C \geq 0 \Rightarrow \tilde{p}_i \geq \frac{C}{V}.$$

- 那么就会存在一段过度进入的区间:

$$\tilde{p}_i \geq \frac{C}{V} > p_i.$$

现在继续考虑 Performance 的情况, 因为表现和能力之间存在一定的随机性:

$$y_i = a_i + \varepsilon_i, \quad a_i \sim N(\mu, \sigma_a^2), \quad \varepsilon_i \sim N(0, \sigma_\varepsilon^2).$$

True success probability (exceed threshold t):

$$p_i = \Pr(y_i > t) = \Pr(a_i + \varepsilon_i > t) = 1 - \Phi\left(\frac{t - a_i}{\sigma_\varepsilon}\right).$$

Overplacement (“I am above average”): 认为自己的能力是 $a_i + \eta, \eta > 0$:

整个的步骤就是:

1. 计算出 true success probability p_i .
2. 计算出 overplaced success probability $\tilde{p}_i$.
3. 找出过度进入的区间 $\tilde{p}_i \geq \frac{C}{V} > p_i$.

$$\tilde{p}_i = 1 - \Phi\left(\frac{t - (a_i + \eta)}{\sigma_\varepsilon}\right).$$

3.1.3. Overprecision: Excessive Certainty

有一个 Uncertain variable $X \sim N(\mu, \sigma^2)$.

- An overprecise agent underestimates variance: $\tilde{\sigma}^2 < \sigma^2$.
- Beliefs are “too sharp” $\Rightarrow$ subjective distribution narrower.

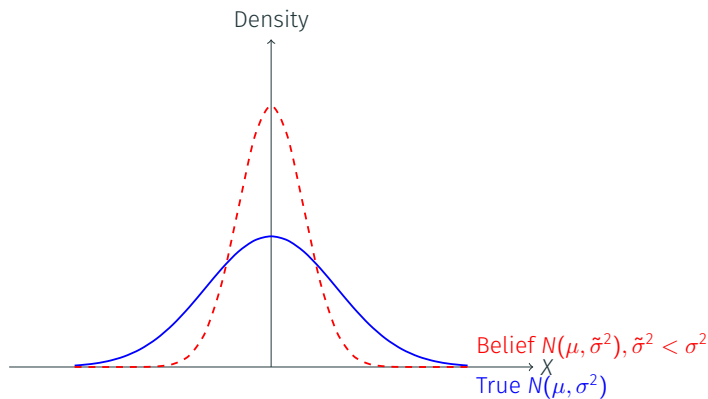


Figure 3.1: Overprecision: Subjective vs. Objective Distribution

Example: Overprecision with Risk Aversion

Preferences: $u(W) = \sqrt{W}$, Initial wealth $W_0 = 100$.

Project payoff: $X \sim N(\mu, \sigma^2)$, $\mu = 0.2$, $\sigma = 10$.

Agent is overprecise: believes variance is smaller, $\tilde{\sigma} = 5$.

Decision rule: accept if and only if $EU > u(W_0)$.

$$\mathbb{E}[u(W_0 + X)] \approx u(W_0 + \mu) - \frac{1}{2}u''(W_0 + \mu)\text{Var}(X), \quad u''(W) = -\frac{1}{4}W^{-\frac{3}{2}}.$$

然后计算²³ $EU^{\text{true}} < 10$, $EU^{\text{belief}} > 10$.²⁴

²³这样的话,

$$\text{Var}^{\text{true}}(X) = 100, \text{Var}^{\text{belief}}(X) = 25.$$

²⁴With the same mean μ , underestimating risk ($\tilde{\sigma} < \sigma$) makes the project look safe enough to accept, even though it should be rejected under the true variance.

3.2. Law of Small Numbers

小数定律: 人们错误地认为小样本也能完美地反映总体的统计特征.

- 根据一个小班级的短期高分就断定老师优秀.
- 认为小医院的高存活率代表其医疗质量更好.
- 期待随机序列能快速“回归均值”, 即相信独立事件之间存在“自我修正”的联系.

人们遵循启发 (Heuristics) 来简化问题.

- Anchoring $\Rightarrow$ Leads to over-precision
- Availability $\Rightarrow$ Connected to limited attention
- Representativeness $\Rightarrow$ Individuals expect random draws to be exceedingly representative of the distribution they come from

赌徒谬论 (Gambler's Fallacy): 认为独立随机事件之间存在负相关关系.

- 例如, 在连续抛出多次正面后, 认为下一次抛出反面的概率更大.

3.3. Projection Bias

人们在预测未来的偏好或感受时, 会系统性地偏向自己当前的状态.

- 在寒冷天气购买过多冬装, 天气转暖后又后悔退货.
- 饥饿时去超市会买太多食物.
- 戒烟时低估未来对尼古丁的渴望.
- 在危机 (如疫情、热浪) 期间高估公众对强力政策的长期支持度.

Loewenstein Model:

- 当前位于状态 s' , 消费 c 的效用为 $u(c, s')$.
- 未来位于状态 s , 消费 c 的效用为 $u(c, s)$.
- 预测未来效用时, 个体会高估当前状态的影响:

$$\hat{u}(c, s) = (1 - \alpha)u(c, s) + \alpha u(c, s'), \quad \alpha \in [0, 1].$$

3.4. Application

庇护法官在连续批准几个案件后, 会倾向于拒绝下一个案件, 仿佛批准和拒绝需要“平衡”一样, 尽管每个案件本应独立判决.

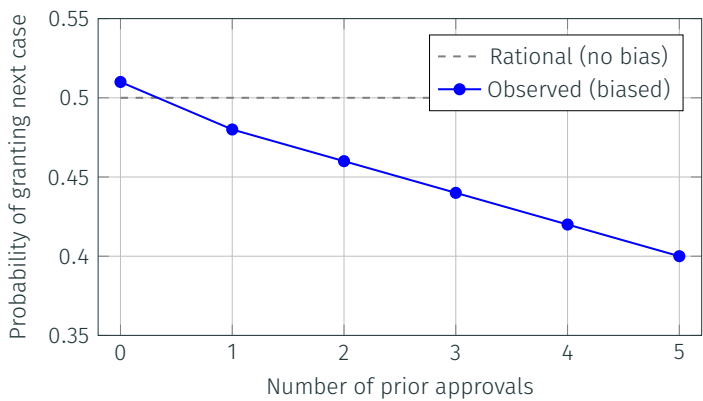


Figure 3.2: Judges Exhibit a “Gambler’s Fallacy” Pattern

4. Non-standard Decision-Making

Category	Meaning	Examples
Non-standard preferences	Preferences deviate from the standard expected utility framework (e.g., time consistency).	Present bias, loss aversion, social preferences.
Non-standard beliefs	People hold systematically biased beliefs or probability judgments about the world.	Representativeness heuristic, overconfidence, left-digit bias, anchoring effect.
Non-standard decision-making	Even with correct beliefs and preferences, decisions may be suboptimal due to limited attention, computation, or framing.	Framing effect, default bias, menu dependence.

Table 4.1: Three Non-standard Categories

在标准的经济学模型中, 个体通过下面的步骤做出决策:

1. 有良好定义的偏好;
2. 有关于不确定事件的正确概率信念;
3. 通过最大化期望效用来选择最优行动.²⁵

²⁵ 错误被视为随机“噪音”.

本章介绍偏离第 3 步的行为.

非标准决策 (Non-standard Decision-Making): 即使偏好和信念正确, 决策也可能因以下原因而次优.

- **有限注意力 (Limited Attention):** 人们倾向于关注显著或易于观察的属性, 或回避认知负荷较高的任务.
- **框架效应 (Framing Effects):** 决策受到信息呈现方式的影响, 即使实质内容相同.
- **默认/现状偏见 (Default/Status Quo Bias):** 人们倾向于维持现有状态, 即使改变更有利.

4.1. Limited Attention

Inattentive Fees: 比如你在携程买票, 北京-海口, 最便宜的是 ¥1298, 但是你付钱的时候一看变成了 ¥1416²⁶, 原来是因为:

- 携程默认给你加上一堆“全能保障”.
- 机场建设费, 燃油附加费等各种附加费用.

这些趁你不注意偷偷加上的费用, 就叫做 **Inattentive Fees**.

A simple model of limited attention.

一个产品有总的价值 V , 这个价值由两部分组成:

- visible component v (显而易见的部分);
- opaque component o (不易察觉的部分).

$$V = v + o.$$

人们只能部分的注意到 opaque component:

$$\hat{V} = v + (1 - \theta)o, \quad \theta \in [0, 1].$$

- $\theta = 0$: fully attentive;
- $\theta = 1$: fully inattentive.

How can we estimate θ ?

- 设置: 在美国, 税是不包含在标价中的, 消费者只能看到标价 p , 而不知道税后价格 $p(1 + t)$.
- 需求函数为:

$$D = D[v + (1 - \theta)tp].$$

Strategy 1: Manipulate the Tax Rate.²⁷ 这是一种利用自然实验 (natural experiment) 的方法, 通过分析历史数据中税收政策的外生变化来推断消费者的行为.

- 假设: 如果消费者是完全理性的, 他们会同等对待这两种税, 因为它们最终都增加了购买啤酒的总成本.
- 如果消费者对不显著的信息不敏感, 那么他们对消费税 (显著) 变化的反应会远大于对销售税 (不显著) 变化的反应.
- 酒精需要缴纳两种税: 销售税 τ^S 和消费税 τ^E . 所以其价格 $q = p(1 + \tau^E)(1 + \tau^S)$.

第一步, 从数据中计算出两个弹性:

1. $\varepsilon_{x, 1 + \tau^S}$: 税率变化对需求的弹性 (Sales tax);
2. $\varepsilon_{x, p}$: 价格变化对需求的弹性.

²⁶ 在携程上买票:



Figure 4.1: 在携程上买票

²⁷ 大约是比较销售税 (Sales Tax) 和消费税 (Excise Tax).

- Sales Tax: 普通商品, 不显著
- Excise Tax: 汽油, 酒精, 烟草等, 显著, “Sin taxes”, 通过税来抑制对这些商品的消费.

第二步, 比较这两者:

$$\frac{\varepsilon_{x,1+\tau^S}}{\varepsilon_{x,p}} = (1 - \theta_\tau).$$

- 如果消费者完全注意到了税 ($\theta_\tau = 0$), 那么比率等于 1.
- 如果消费者没有注意到税 ($\theta_\tau > 0$), 那么比率应该是小于 1 的.

现在有了销售税 (Sales Tax) 和消费税 (Excise Tax) 的数据:

- 销售税 (Sales Tax) $\rightarrow \varepsilon_{x,1+\tau^S} \cdot q = p_{\text{标价}}(1 + \tau^S)$.
- 消费税 (Excise Tax) $\rightarrow \varepsilon_{x,p}$. 因为这个税被含在标价里了, 也就是 $p_{\text{标价}} = p(1 + \tau^E)$.

第一步, 计算得到:

1. 销售税 (Sales Tax) 的弹性: $\varepsilon_{x,1+\tau^S} \approx 0.03$;
2. 消费税 (Excise Tax) 的弹性: $\varepsilon_{x,1+\tau^E} \approx 0.84$.

第二步, 先调整销售税的弹性: $\varepsilon_{x,1+\tau^A} \approx (1/0.6)\varepsilon_{x,1+\tau^S}$.²⁸

最后, 计算 θ_τ :

$$\theta_\tau = 1 - \frac{\varepsilon_{x,1+\tau^A}}{\varepsilon_{x,1+\tau^E}} \approx 0.06.$$

Strategy II: Manipulate Tax Salience.

在一个受控但真实的环境中直接测试消费者的反应.

地点: 在美国加州北部一家年收入约 2500 万美元的大型超市进行.

设计

- 控制组: non-salient tax condition. 商品价格标签上不显示税.
- 处理组: salient tax condition. 商品价格标签上显示税.

$$\theta_\tau = 1 - \frac{\text{observed demand response}}{\varepsilon_{x,p} \times \text{tax rate}}.$$

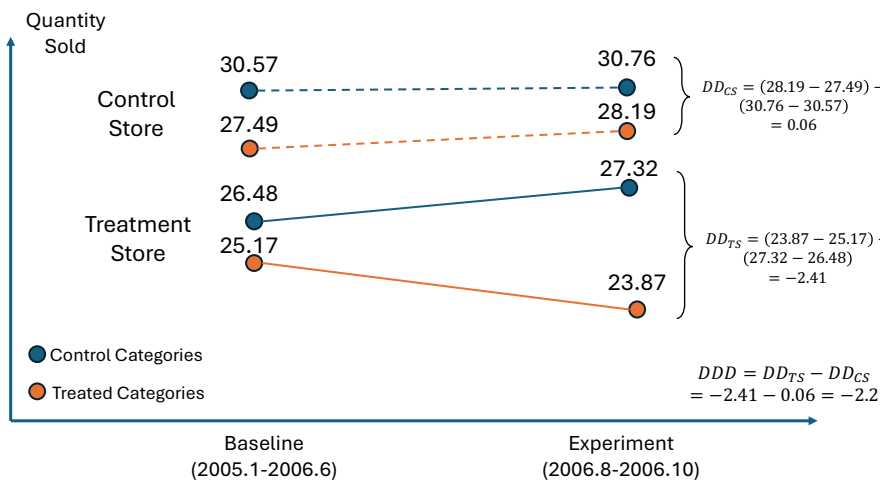


Figure 4.2: Difference-in-Difference-in-Differences (DDD)²⁹

同时, 本次实验的基准售出商品的量为 29 单位, 因此:

²⁸ 在比较销售税与消费税效应时, 一个重要考量是销售税的适用范围比酒精更广, 有约 40% 的消费需缴纳销售税.

设酒精的税为 τ^A . 下面所说的含税价格指的是含 τ^A 的税.

设酒精 (x) 的含税价格 $(1 + \tau^S)$ 增长 1%, 把除酒精外的所有商品视为一个整体 y , 其中 40% 的商品需要缴纳销售税, 那么 y 的价格变化为 $1 + 0.004$. 从而 x 相对于 y 的价格增长 $\frac{1.01}{1.004} - 1 \approx 0.006$.

那么: $0.6\Delta(1 + \tau^S) = \Delta(1 + \tau^A)$. 那么:

$$\varepsilon_{x,1+\tau^A} = \frac{\Delta D}{\Delta(1 + \tau^A)} = \frac{1}{0.6} \frac{\Delta D}{\Delta(1 + \tau^S)} = \frac{1}{0.6} \varepsilon_{x,1+\tau^S}.$$

²⁹ 为什么要使用 DDD? DDD 估计量能够同时排除两类冲击的影响: 一是商店特有的冲击 (例如客流量的暂时性增加), 二是商品特有的冲击 (例如某些商品需求的波动). 因此, DDD 估计量具有一致性的识别假设是: 在我们的实验干预期间, 不存在任何仅对处理商店中的处理商品的销售产生差异化影响的冲击. 鉴于本次干预是经过计划且外生的, 我们认为这一条件很可能是满足的.

$$\text{observed demand response} = \frac{2.2}{29} = 7.6\%.$$

估算出来的 $\varepsilon_{x,p} = 1.59$, the sales tax rate = 7.375 percent, 那么我们可以计算出:

$$\theta_{\tau} = 1 - \frac{0.076}{1.59 \times 0.07375} \approx 0.35.$$

核心问题: 人们之所以不注意, 是因为他们犯了错误, 还是因为集中注意力是有成本的?

- 注意力是一种稀缺的认知资源;
- 某些信息过于复杂、充满噪声, 或其重要性太低;
- 获取和处理信息是需要付出成本的;
- 体会在权衡信息获取成本与潜在收益的基础上, 最优地选择自己要了解多少信息.

4.2. Framing Effects

框架效应 (Framing Effects): 人们的决策受到信息呈现方式的影响, 即使实质内容相同.³⁰

例子: 约翰正在当地药店选购消毒湿巾. 他看到多种产品, 其中两款湿巾正在促销. 一款名为“Bleachox”, 另一款叫“Bleach-it”.

- 两款湿巾价格相同, 容量也一致.
- 唯一区别在于: Bleachox 宣称“能杀灭 95% 的细菌”, 而“Bleach-it”则表示“仅有 5% 的细菌存活”.
- 经过对比后, 约翰最终选择了 Bleachox 湿巾. 他实在无法接受厨房台面上残留细菌的念头.

例子: 当治疗效果被负面框架 (死亡概率) 描述, 且生存概率低于 50% 时, 患者更倾向于选择效果较差但无毒的方案, 因为此时“生活质量”比“生存时间”更重要.

4.3. Menu Effects

菜单效应: 人们的选择不仅取决于选项本身, 还取决于整个选择集 (菜单) 的构成和呈现方式. 这违反了标准经济学中的“无关选项独立性”原则.

- **context-dependent choice.**

影响的机制:

1. 比较框架 (Comparison framing): 选项的价值通过与其他选项比较得出 (如“折中”选项).
2. 吸引力/显著性 (Attention/salience): 增加一个极端选项可以凸显其他选项的某些属性.
3. 决策难度 (Decision difficulty): 过多选项可能导致认知过载, 影响选择过程.
4. 自我控制 (Self-control): 某些选项可能作为“诱惑”存在, 影响自我控制行为.

经典例子: 诱饵效应 (Decoy Effect)³¹

³⁰ 如图, 人们更喜好 80% Fat-Free 的表述, 而不是 20% Fat 的表述.

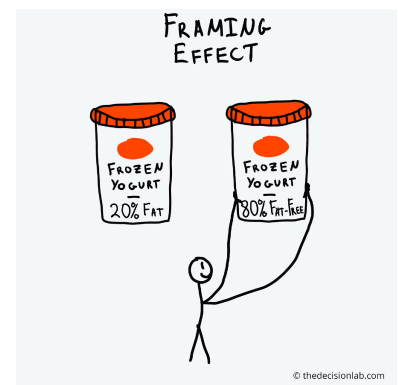


Figure 4.3: Framing Effects

31

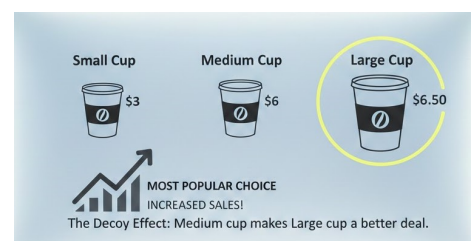


Figure 4.4: Decoy Effect

- 考虑如果本来只有: 小杯(\$3) + 大杯(\$6.5). 消费者可能会选择小杯, 因为大杯价格较高且容量过大.
- 现在引入一个中杯(\$6), 容量介于小杯和大杯之间. 这个中杯作为“诱饵”存在, 因为它在价格上接近大杯, 但容量明显小于大杯.
- 结果, 许多消费者现在选择大杯, 因为它在价格上相对于中杯更具吸引力, 从而增加了大杯的销量.

5. Nudge

人们并非总是理性的决策者, 他们的选择会受到决策环境 (即“选择架构”) 的深刻影响. “助推”是一种温和的干预方式, 旨在通过巧妙地设计选择环境, 引导人们做出更有利于自身和社会的选择, 同时不剥夺其自由选择权.

助推 (Nudge) 是任何在不禁止任何选项或不显著改变其经济激励的前提下, 以可预测的方式改变人们行为的选择架构特征.

- 保持选择自由 (不强制、不禁止);
- 作用于心理机制: 通过认知偏差、启发式等心理因素起作用, 而非通过价格或法规;
- 旨在改善福利: 帮助人们做出更符合其长期利益的选择.

Category	Description
Information & Salience	Highlight key information to make it more noticeable or understandable (e.g., energy use feedback, calorie labels).
Reminders & Prompts	Send timely cues that help people remember to take action (e.g., SMS reminders for appointments).
Default Options	Set a beneficial option as the pre-selected choice, relying on inertia(惯性) (e.g., automatic enrollment).
Framing & Presentation	Change how choices or outcomes are presented without altering actual incentives (e.g., survival vs. mortality framing).
Social Norm Nudges	Show what others like you typically do to influence behavior (e.g., “Most households recycle”).
Simplification & Choice Ease	Reduce complexity or effort within the decision process (e.g., pre-filled forms, one-click actions).

Table 5.1: Types of Nudges

但是每个助推措施可能同时使用了不同的种类的方法, 如: SMS reminder for medical appointment; the hospital sends you an SMS the day before your appointment.

Category	How the SMS Functions
Information & Salience	Brings the appointment back to attention at the right moment.
Reminders & Prompts	Helps you follow through on an intention you already had.
Simplification & Choice Ease	Reduces the effort of remembering or re-checking details.

Table 5.2: How the SMS Functions

自由主义 (Libertarianism): 人们是理性的, 可以自己做出最优选择.

家长制 (Paternalism): 人们不是理性的, 需要政府干预来保护他们免受自己的错误决策的伤害.

Nudge → 自由主义家长制 (Libertarian Paternalism): 通过轻微的干预来帮助人们做出更好的选择, 同时保持选择自由.

适用场景: 助推在以下情境中尤为有效.³²

- 投资品 vs. 罪恶品 (Sinful goods): 如锻炼 (现在付出, 未来受益) vs. 吸烟 (现在享受, 未来受害).
- 困难的选择: 如购房、投资、选专业.
- 罕见但关键的选择: 如结婚、换工作, 缺乏练习机会.
- 缺乏反馈的重复选择: 如节食, 效果滞后.
- 结果陌生的选择: 如尝试新菜系、新课程.

助推的几种方法:

- Default
- Social Contagion
- Goal Setting and Commitment Devices
- Noise
- Sludge

5.1. Default

Default: 预先设定的、若无主动更改就会自动生效的选项.

浣熊先生就是一个默认选项的例子, 一般它设置的默认选项是 45 分钟超长洗, 人们通常都会选择这个选项.³³

影响力渠道 (Channels) (七大机制):

1. **转换成本 (Switching Costs):** 改变默认选项需要时间、精力或金钱, 这些成本可能阻碍人们做出改变.
 - 转换成本可分为机制性 (如填写表格、调整设置) 和信息性 (如理解替代方案、评估后果) 两类.
2. **暗示性认可 (Implied Endorsement):** 默认选项被视为权威或专家的推荐, 增强了其可信度和吸引力.
3. **参照点与损失厌恶 (Reference Point & Loss Aversion):**
 - 默认选项通常成为决策者衡量替代方案的参照基准.
 - 默认选项被视为“已有之物”, 放弃它会被感知为一种损失.
4. **显著性 (Salience):** 默认选项通常被设计得更醒目、更突出.
5. **现时偏见 (Present Bias):** 当决策需要当前付出努力但收益将在未来获得时, 人们倾向于拖延——即使他们认为未来收益更大.
 - 默认选项减少了即时付出努力的需求.
 - 个体无需自行“启动”或“发起”该行为.
 - 这有助于规避源于当前偏好的拖延行为.
 - 储蓄计划的自动注册机制可增加长期储蓄, 因个体无需立即完成注册流程.
6. **注意力 (Attention):** 默认选项更容易被注意到.
7. **注意力不集中 (Inattention):** 个体并非总能完整阅读选择界面的每个部分或充分评估所有可用选项. 当注意力有限时, 决策者可能仅因未察觉其他选项的存在而选择默认选项.

³² When Nudge is a good idea?

- Investment Goods and Sinful Goods.
- Difficult Choices.
- Infrequent but Crucial Choices.
- Repeatable Choices without Feedback.
- Choices with Unfamiliar Outcomes.

33



Figure 5.1: 浣熊先生

Diana Sosa 的文章 *A Default Life*³⁴ 将此概念扩展到社会领域. 从出生起, 我们的性别、阶级、种族等因素就为我们设定了社会“默认路径”(如职业、婚姻、生活方式). 虽然我们可以偏离这些路径, 但这通常需要克服巨大的社会摩擦. 这提醒我们要有意识地审视自己的选择, 思考“我们真正想要的是什么?”.

5.2. Social Contagion

社会传染 (Social Contagion): 态度、行为和决策通过观察、模仿和社会互动在人与人之间传播的过程.

作为助推的应用: 利用社会规范来影响行为.

- 酒店提示“本房间 75% 的客人都重复使用毛巾”;
- 电费账单显示“您比邻居多用 XX% 电”.
- 微信运动
- 学生排名
- 社会标语, “生男生女都一样”, “今年上门, 明年上坟”.
- 光盘行动
- “为祖国健康工作五十年”, “完全人格, 首在体育”.

影响力渠道 (Channels):

1. **信息性影响 (Informational Influence)**: 在陌生、模糊或高风险环境时, 通过他人的行为来判断何为合理/有效的捷径.
2. **规范性影响 (Normative Influence)**: 遵循社会规范以获得归属感和社会认可, 避免排斥或惩罚.
3. **情绪传染 (Emotional Contagion)**: 情绪和态度通过非语言线索 (如面部表情、语调) 在个体间传播, 影响决策.
4. **阈值动态 (Threshold Dynamics)**: 个体对行为参与的阈值存在差异.
 - 部分个体会早期采纳, 而多数个体则需观察他人行为后才参与.
 - 当参与度达到特定阈值时, 该行为可迅速扩散.
5. **向上比较 (Upward Comparison)**: 通过与社会中地位更高或更成功的个体比较, 激励个体提升自身行为以符合更高标准.
 - 当差距可追赶时, 个体更有动力改变行为.
 - 当差距过大时, 个体可能会感到气馁而放弃努力.
6. **向下比较 (Downward Comparison)**: 通过与地位较低或不如自己的个体比较, 提升自尊和满意度, 从而减少改变行为的动机.
 - 个体可能会通过贬低他人来提升自我价值感.
 - 这种比较有时会导致对现状的满足, 减少改进的动力.
7. **同伴比较 (Peer Comparison)**: 通过与同龄人或类似背景的个体比较, 激发竞争和模仿行为. 个体更容易受到与自己相似的人的影响. 同伴压力可促使个体调整行为以符合群体标准.
8. **地位信号 (Status Signaling)**: 个体通过采纳某些行为来传达其社会地位或身份. 这些行为可能被视为地位的象征, 激励个体模仿以提升自身形象.

5.3. Goal Setting and Commitment Devices

目标设定与承诺机制 (Goal Setting and Commitment Devices): 通过设定明确的目标和使用承诺机制来增强自我控制和实现长期目标的策略.

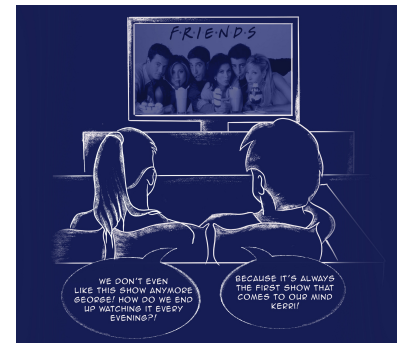


Figure 5.2: A Comic Drawing from GPT

针对**现时偏见 (Present Bias)**, 即人们过度看重眼前的成本/收益, 而低估未来的成本/收益.

5.3.1. Goal Setting

如何设计一个好的目标 → The SMART Criteria:

S	Specific	What will be accomplished? What actions will you take?
M	Measurable	What data will measure the goal? (How much? How well?)
A	Achievable	Is the goal doable? Do you have the necessary skills and resources?
R	Relevant	How does the goal align with broader goals? Why is the result important?
T	Time-Bound	What is the time frame for accomplishing the goal?

Table 5.3: The SMART Criteria

5.3.2. Commitment Devices

一种能改变未来选择集或收益的工具, 使得当下的最优行动在未来也成为私人最优.

- Price-based: 储蓄, 罚款, ...
- Quantity/choice-based: 锁定某些选项, 比如 site blockers, deadlines.
- Social: 公布承诺, 社交监督, ...

Hard vs. Soft, Internal vs. External, Monetary vs. Non-monetary.

Commitment in a $\beta - \delta$ World

Suppose an action $a \in \{0, 1\}$ yields cost $c > 0$ today and benefit $b > 0$ at $t + 1$.

$$U_0 = -ca + \beta\delta ba, \quad U_1 = -ca + \delta ba.$$

长期来看, 最优为 $\delta b > c$. 但是现在的 Present-biased self 会选择 $a = 1$, 只有 $\beta\delta b > c$ 时.

有一段区间: $\delta b > c \geq \beta\delta b$, 这里个体不愿意采取行动, 需要一个机制帮助个体采取行动.

Deposit contract: pay a refundable deposit D forfeited(违反) if $a = 0$.

- 选择 D 使得 $\beta\delta b + \beta D \geq c, -c + \delta b \geq 0$.

常见的失败原因:

- Goal dilution: too many goals $\Rightarrow$ attention spread thin.
- Moving goalposts: redefining success ex post.
- Crowding out: extrinsic penalties erode intrinsic motives.
- Slippage: ambiguous measurement leads to rationalization.
- Reactance: overly rigid commitments trigger avoidance.

5.3.3. Application

参与者将钱存入一个特殊账户, 6 个月通过尼古丁尿检即可拿回全部存款, 否则钱被没收. 这种机制有效利用了人们对未来损失的厌恶, 帮助其克服当下的诱惑.

5.4. Noise and Sludge

5.4.1. Noise

噪声: 指在相同情况下, 不同决策者 (或同一决策者在不同时间) 做出的不必要且有害的变异. 它不同于系统性偏差 (Bias).

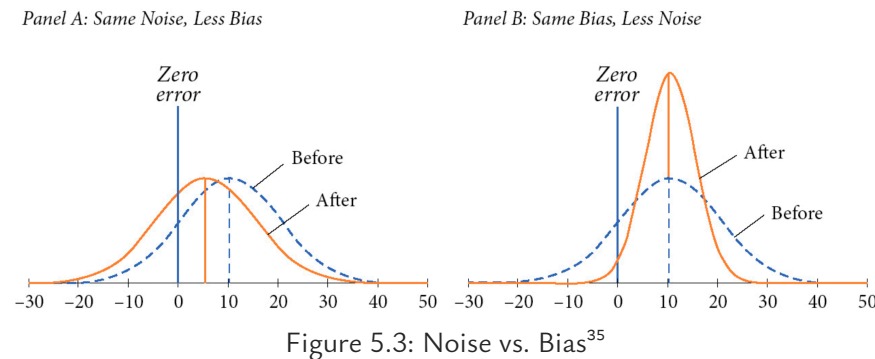
- A general property of noise is that you can recognize and measure it while knowing nothing about the target or bias.

Errors in a single measurement:

$$\text{Errors in a single measurement} = \text{Bias} + \text{Noisy Error}.$$

Overall Error (MSE):

$$\text{Overall Error (MSE)} = \text{Bias}^2 + \text{Noisy Error}^2.$$



³⁵ 请注意: 让 Noise 减少一半以后, 反而情况变差了?

The paradox here is that noise reduction seems to have made things worse. The forecasts are now **more concentrated** (less noisy) but **not more accurate** (not less biased).

例子: 温度和法庭判决的关系.

- 判决日温度每升高 10°F, 对申请者有利的决策减少 6.55%.

Linear probability model:

$$g_{it} = \beta_0 + \beta_1 temp_{it} + W_{it}\beta_2 + P_{it}\beta_3 + X_{it}\beta_4 + \gamma_i + \phi_{ct} + \theta_t + \epsilon_{it}.$$

- g_{it} : 第 i 个案子在时间 t 是否获得批准.
- $temp_{it}$: 判决日的温度 (6am to 4pm).
- W_{it} : 其他气候因素.
- P_{it} : 污染的因素.
- X_{it} : 法院和申请的其他特点.
- γ_i : 法官固定效应.³⁶
- ϕ_{ct} : 城市-月份固定效应, 用于控制城市层面随时间变化的不可观测因素影响.
- θ_t : 工作日固定效应, 用于控制一周内不同日期对研究结果的影响.
- ϵ_{it} : 误差项.

5.4.2. Sludge

Sludge: 指阻碍人们获得想要事物的摩擦和障碍, 如冗长的表格、漫长的等待时间、复杂的流程.

- Sludge is good when people are not reckless or impulsive.

³⁶ 固定效应: 一个简单的回归方程

$$y_{it} = \alpha + \beta x_{it} + \gamma_i + \epsilon_{it}.$$

其中的 γ_i 就是用来控制我们未观察到的个体效应, 当然, 要假设: 未观察到的个体效应随时间保持不变, 并与独立变量相关.

这里:

- 若某法官本身对庇护申请的批准率始终高于其他法官, 不控制法官固定效应会误将其个人倾向归因于温度等核心变量.
- 若某城市在夏季 (如 6 月) 因政策调整提高了庇护申请批准率, 而夏季温度本身较高, 不控制城市-月份固定效应会错误认为“高温导致批准率上升”.
- 若法官在周五普遍倾向于快速处理案件 (可能降低批准率), 而周五温度恰好较高, 不控制星期几固定效应会误将“周五效应”归因于温度.

- ▶ To get a driver's license, you must pass some exams and satisfy the health and safety requirements.
- Sludge is good in providing some small 'frictions'.
 - ▶ To get food or subsidies from the government, frictions can help justify that you really have the right to do it.

双重性:

- 坏的 Sludge: 阻碍人们获取应得的福利 (如医保、福利), 导致机会流失和资源浪费.
- 好的 Sludge: 帮助人们避免冲动决策 (如过度借贷、冲动购物), 促进理性选择和长期福利.

“黑暗模式”(Dark Patterns) 是企业故意设置 Sludge, 利用信息不对称诱导消费者做出不利选择 (如难以取消的订阅).

6. Behavioral Topics

6.1. Behavioral Economics for Environmental Issues

核心问题: 为什么环保政策常常失效?

- 传统政策的局限: 传统的环境政策 (如碳税、补贴、信息宣传) 基于“理性人”假设, 认为只要提供正确的激励或信息, 人们就会采取环保行动.
- “态度-行为鸿沟” (Attitude-Behavior Gap): 人们普遍表达出强烈的环保关切, 但在实际行动上却常常失败 (例如过度消费、不节约能源). 这表明“好意图 $\neq$ 好行为”.
 - ▶ Present-bias: undervaluing future environmental benefits.
 - ▶ Status quo bias: resistance to green technology adoption.
 - ▶ Limited attention & information overload: failure to notice energy-saving opportunities.

$\text{Gap} = \text{EC (environmental concern)} - \text{EB (pro-environmental behavior)}$

所以, Why Do People Lack Pro-Environmental Behaviors?

1. **Limited Attention.** Consumers struggle to evaluate long-term cost-benefit trade-offs.
 - LED bulbs: 前期费用高, 长期节省不明确
 - Energy-efficient appliances: 溢价与节电 (节电难量化)
 - Organic food: 价格高, 健康/环保效益不明确
2. **Self-control.** Present preferences vs. long-term goals.
 - Forgetting to turn off lights, AC.
 - Choosing to drive for short trips instead of walking.
 - Overconsumption, impulse buying.
3. **Underestimating Small Actions.** Misperception of individual environmental impact.
 - “My water conservation doesn't matter”
 - Waste sorting: “This little trash won't matter”
 - Small contribution of reducing plastic bag use

6.1.1. Detect Behavioral Issues

例子: 研究消费者对阶梯电价 (非线性定价) 的反应.

- 传统理论预测: 消费者应根据边际价格来调整用电量.

- 行为经济学发现: 大量消费者实际上根据平均价格来做决策.
- 如果消费者只关注平均价格, 那么旨在通过提高边际价格来抑制用电的阶梯电价政策可能效果不佳, 甚至适得其反. 这要求政策设计必须考虑人们的真实认知方式, 而非理想化的理性模型.

6.1.2. Applications

Single-Use Cutlery in Food Delivery.³⁷

- 中国面临着外卖中一次性餐具的过度使用问题.
- 一个询问餐具设置的弹窗会跳出来, 其中设置了默认选项为“无需餐具”, 且有提示可以获得绿色积分 (Default).

³⁷ 《科学》封面文章 | 全球健康发展研究院潘聿航与合作者报道绿色助推在减塑及企业 ESG 实践领域的重要价值

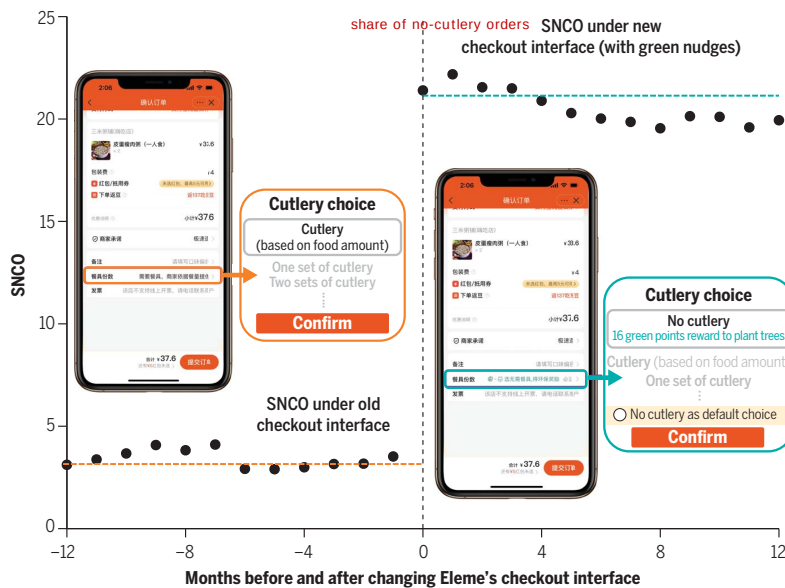


Figure 6.1: Nudging to Reduce Single-Use Cutlery in Food Delivery

The Buffet That Wasted Too Much.

- 酒店自助餐食物浪费严重.
- 做法: Swap to modestly smaller plates $\Rightarrow$ lower initial loads.
- 利用“份量规范”和“单位偏见”, 成功减少了约 20% 的食物浪费.
- Anchoring effect: smaller plates set a lower consumption norm.

what if we send you your neighbors' electricity bill?

- Social comparison
- 向家庭发送包含邻里用电对比的社会规范信息 (“您的用电量高于邻居”), 有效促进了节能行为.

Hotel with Sustainable Development.

1. Specific commitment (“I will reuse my towels”)
2. General commitment (“I will act eco-friendly”)
3. Each with / without a symbolic *Friend of the Earth* pin
4. Three controls: message-only, pin-only, no manipulation

让客人做出具体的环保承诺 (如“我将重复使用毛巾”) 并佩戴象征身份的徽章, 显著提高了毛巾重复使用率. 这利用了“一致性驱动”和“自我信号”效应.

Waste of Paper.

- No significant change under moral appeal message treatment.

- Change the Default Choice → Reduce paper use by 15%!

6.2. Behavioral Finance

标准的金融: modern portfolio theory, the efficient market hypothesis.

- Modern portfolio theory: choose a portfolio that maximizes expected return for a given level of risk.
- Efficient market hypothesis: asset prices fully and immediately reflect all available information.

然而,大量实证研究表明,投资者的行为和市场价格常常系统性地偏离这些理想化假设.行为金融学应运而生,旨在填补这一鸿沟.

1. Heuristic simplification: 依赖心理捷径做决策.

- **Representativeness**: Judging probabilities by similarity (e.g., "good company = good stock"). "好公司 ≠ 好股票."
- **Availability**: Overweighting recent or vivid events (e.g., recent crash = higher perceived risk).
- **Anchoring**: Sticking to initial numbers (e.g., purchase price) when updating beliefs. 过度依赖初始信息 (如买入价、52周高点), 导致"处置效应" (过早卖出赢家, 过久持有输家).

2. Affective short-circuiting: 情绪压倒理性分析.

- Greed in bull markets → chasing recent winners.
- Regret → holding losing stocks too long.

3. Self-deception: 持有过于乐观的信念以保护自尊.

- Overconfidence.
- Confirmation Bias: Seeking information that confirms existing beliefs.
- Self-Attribution Bias. 成功就说自己牛, 失败了就怪环境不行.

6.2.1. January Effect

January Effect: Stock returns, especially for small-cap stocks, tend to be abnormally high in January.

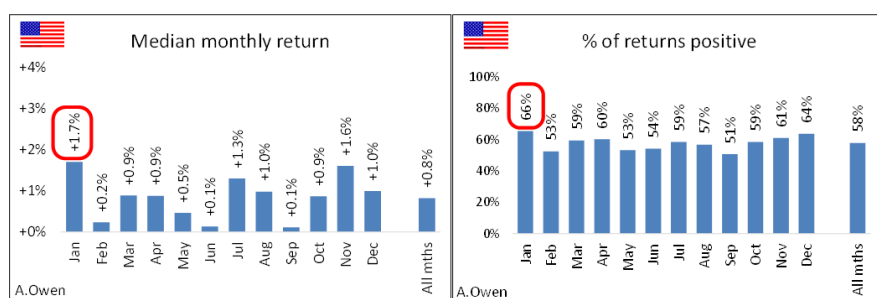


Figure 6.2: January Effect

传统的解释:

1. Tax-Loss Selling Hypothesis.
 - 投资者在 12 月抛售亏损股票以实现税务亏损.
2. Window Dressing by Fund Managers.
 - 基金经理在年底调整持仓, 以展示更好的业绩.
 - 一月又买回这些股票.
3. Liquidity Effects.
 - 12 月交易量下降. 1 月流动性趋于正常 → 收益率上升.

行为经济学解释:

- 年末投资者因亏损而情绪低落、过度反应, 新年伊始则充满乐观情绪, 愿意承担更多风险.
- Affective Short-Circuiting + Heuristic Simplification.

6.2.2. Panics

指信心突然崩溃, 导致价格急剧下跌、波动性飙升、流动性枯竭.

关键特征: 资产间相关性急剧升高, 表明投资者行为高度趋同.

6.2.3. Herding

羊群效应 (Herding): 投资者不是基于自己的独立信息, 而是有意模仿他人的交易行为.

- 导致拥挤交易、催生泡沫与崩盘、加剧市场波动.³⁸

Herding 的类型:

- Informational herding. 认为别人知道更多信息, 从而从他人的行为中推断信息.
 - ▶ Information Cascades: Agents ignore their own signal and follow predecessors.
- Reputational herding. 职业经理人为了避免“与众不同”的职业风险, 宁愿“和大家一起犯错”.
 - ▶ Reputational Herding: Managers follow others to protect their career.
- Institutional herding. 由于共同的基准、风控模型或监管要求, 导致机构投资者做出相似的交易决策.
 - ▶ Global Games and Crises: Strategic complementarities lead to coordinated actions (attacks, runs).

共同基金追逐热门板块、分析师预测趋同、散户跟风炒作“迷因股”.

Model 1: Information Cascades.

- 设当前资产的状态为 $\theta \in \{H, L\}$ (H 说明值得买), $P(\theta = H) = P(\theta = L) = \frac{1}{2}$.
- Agents $i = 1, 2, 3, \dots$ move sequentially.
- 每个 agent 自己有一个私人信号 $s_i \in \{h, l\}$.

Agent 1 选择买, Agent 2 结合了 s_2 和 a_1 , 也选择买; 此时 Agent 3 结合了 a_1, a_2 和 $s_3 = l$, 依然选择买. 剩下的人依然选择买...

Model 2: Reputational Herding.

$$U_i = \underbrace{E(\pi_i)}_{\text{investment performance}} - \underbrace{C(\text{deviation from peers})}_{\text{reputational cost}}$$

Model 3: Global Games and Crises.

- Fundamentals $\theta \in \mathbb{R}$ (e.g., strength of a currency or bank).
- 每个观察者有一个信号 $x_i = \theta + \varepsilon_i$, $\varepsilon_i \in N(0, \sigma^2)$.
- 有一个 threshold x^* , Attack if $x_i < x^*$, Not attack if $x_i > x^*$.

For fundamentals near the threshold, small changes in beliefs lead to large, coordinated move.

38



Figure 6.3: Herding Behavior in Financial Markets

Many investors attack at the same time, not just because of θ , but because they expect others to attack.

6.3. Health and Development Issues

6.3.1. Why are gyms so expensive?

Puzzle: Many people buy monthly or annual gym memberships but rarely attend.

标准理性模型: 每月费用 L , 单次访问计划 p , 每月去的次数为 v , $E[v]$ 每月去的期望.

$$\frac{L}{E[v]} \leq p,$$

才应该选择月度计划.

- 预测错误 (Forecasting Errors): 会员在办卡时高估了自己的使用频率.
- 现时偏见 (Present Bias): 办卡时对未来锻炼的承诺很坚定, 但到了具体行动的日子, 惰性占了上风.
- 惯性与自动续费 (Inertia & Auto-Renewal): 即使不再去健身房, 用户也会因为忘记或嫌麻烦而继续支付费用.

6.3.2. Why are healthcare prices so high?

Behavioral economics adds a new layer: **Doctors**, like patients, systematically deviate from fully rational decision-making.

- Behavior 1: Defensive Medicine (Loss Aversion). 要尽可能多做检查.
- Behavior 2: Heuristics and Cognitive Biases. 过度强调显性症状; 低估隐性风险. 将患者匹配至典型疾病模式.
- Behavior 3: Income Targeting and Financial Incentives. 医生可能会调整服务数量以达到收入目标.
- Behavior 4: Practice Style Inertia. 医生有自己的手术实践惯性.

如何衡量一个医生的行为?

1. 做手术的水平.
2. 诊断的水平.
3. Aggressiveness. 是否敢开刀. $\rightarrow \alpha$.
4. Representativeness. 这个医生有多依赖将患者匹配至典型疾病模式. $\rightarrow$ Responsiveness β .

这个框架超越了简单的“好医生/坏医生”二分法, 揭示了医疗决策中复杂的个体差异, 并为评估医疗质量和效率提供了新的分析工具.

6.4. Behavioral Bias in Relationship

Confirmation Bias: The tendency to interpret or recall information to align with preexisting beliefs about a partner or relationship.

- “He didn’t answer my text for hours – see, he never prioritizes me.”
- “He forgot my birthday, but he has been so stressed lately – he still loves me.”
- 争吵 $\rightarrow$ 健康的讨论 or 毁灭的征兆.

Fundamental Attribution Error: overemphasizing personality traits and ignoring situational factors. In relationships, this leads to blaming a partner's actions without considering external circumstances.

- 归因到个人性格, 而非当时情境.
- 基本归因错误

Halo Effect: the halo effect happens when a partner's positive qualities lead to unrealistic expectations, causing disappointment and strain when they inevitably fall short.

- 光环效应

Horn Effect: 和 Halo Effect 相对, 因为某个人坏品质而产生了偏见, 进而影响对其整体的负面评价.

Negativity Bias: Negativity bias is the tendency to focus on negative experiences over positive ones, causing partners to dwell on conflicts instead of celebrating successes.³⁹

Sunk Cost Fallacy: The sunk cost fallacy occurs when people continue investing based on past resources rather than the current situation. In relationships, this leads couples to stay together out of obligation or fear of wasted time, rather than true happiness.

Availability Heuristic: The availability heuristic is a mental shortcut where decisions are based on easily accessible information, which can lead to hasty conclusions about a partner's behavior, causing misunderstandings and conflict.

- 最近老看到别人出轨, 怀疑自己伴侣出轨.

Self-Serving Bias: The self-serving bias is the tendency to credit success to personal effort and blame failure on external factors. In relationships, this can lead to partners blaming each other for problems and taking credit for successes, fostering resentment and mistrust.⁴⁰

Anchoring Bias: Anchoring bias is the tendency to rely too heavily on initial information when making decisions. In relationships, this can lead to strong opinions based on first impressions, which may not reflect the full context or a partner's true character.

- 根据对方过去的错误来判断他现在的为人

False Consensus Effect: The false consensus effect is the tendency to overestimate how much others share our beliefs. In relationships, this can cause frustration when partners assume their spouse shares their views, only to find they don't.

- 虚假共识效应.

Empathy Gap: The empathy gap refers to the difficulty in understanding or predicting a partner's emotions, thoughts, or behaviors when in a different emotional state. In relationships, this bias can create disconnects and misunderstandings.

- 处于不同的情绪状态时就无法沟通

39

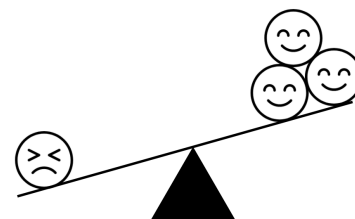


Figure 6.4: Negativity Bias in Relationships

40

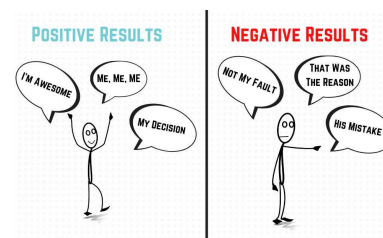


Figure 6.5: Self-Serving Bias in Relationships

Bibliography

- [1] E. Duflo, M. Kremer, and J. Robinson, "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya," *American Economic Review*, vol. 101, no. 6, pp. 2350–2390, Oct. 2011, doi: [10.1257/aer.101.6.2350](https://doi.org/10.1257/aer.101.6.2350).
- [2] C. Camerer, L. Babcock, G. Loewenstein, and R. Thaler, "Labor Supply of New York City Cabdrivers: One Day at a Time," *The Quarterly Journal of Economics*, vol. 112, no. 2, pp. 407–441, 1997, Accessed: Jan. 03, 2026. [Online]. Available: <http://www.jstor.org/stable/2951241>