

从 Hotelling 到选举制度：多维偏好下候选人策略与社会选择扭曲

2026 年 7 月 8 日

摘要

Hotelling 空间竞争和 Downs 中位选民逻辑给出一个清楚的基准：在一维政策空间、两名候选人、单峰偏好和简单多数制下，候选人会向中位选民靠拢。这个结论依赖很强的条件。放松这些条件以后，极化、多党竞争、弃保、联盟政治和名单政治都可以从模型里推出。全文先把政策空间扩展为经济、文化和治理能力等多维属性，并用加权二次距离、议题满意度、CES、Cobb-Douglas 和 Stone-Geary 等形式讨论选民如何聚合多维偏好；随后用二维 Condorcet 循环和 Ostrogorski / Anscombe 悖论说明，个体偏好的理性不保证社会选择稳定。选举制度也不只是计票工具，它会压缩、保留或扭曲偏好信息。相对多数制、两轮制、排序投票、比例代表、并立制和联立制分别改变选民表达方式、候选人目标函数和政党进入策略。最后的小型仿真和席位公式算例显示，同一组选民在不同规则下会产生不同胜者、席位结构和政策距离损失。制度设计要比较的，是代表性、稳定性、问责性、比例性和策略可操纵性之间的取舍。

目录

1	从 Hotelling 出发：位置竞争怎样变成选举模型	4
1.1	一个五人例子：中位点为什么能赢	6
1.2	两名候选人为什么向中位点靠	7
1.3	偶数选民与中位区间	8
2	第一处问题：现实政治为什么不总是中间化	9
2.1	逐条放松假设：哪些结论会先断掉	10
2.2	从“偏离中位”到“制度诱导”	12

目录	2
3 多维政策空间与选民效用	12
3.1 从一维政策到多维政策	12
3.2 加权二次距离效用	13
3.3 满意度聚合与底线议题	14
3.4 议题权重如何改变同一理想点的选择	15
3.5 方向模型：选民有时关心“往哪边推”	15
3.6 Valence：政策距离之外的候选人质量	16
3.7 CES、Cobb-Douglas 与底线议题	17
3.8 多维效用如何导向制度问题	18
4 个体理性为何不能保证社会选择稳定	18
4.1 二维 Condorcet 循环	18
4.2 Ostrogorski / Anscombe 悖论	19
4.3 不可能性作为背景	20
4.4 May 定理：两选项多数决为什么特殊	20
4.5 二维距离如何生成 Condorcet 循环	21
4.6 议程控制：同一偏好，不同投票顺序	23
4.7 Ostrogorski / Anscombe 悖论的完整计算	23
4.8 从多数循环走向不可能性定理	24
4.9 从循环图看多数关系的问题	24
4.10 Arrow 不可能定理	25
4.11 Gibbard-Satterthwaite 定理	27
5 选举制度如何读取偏好	28
5.1 同一组选票下的单赢家制度比较	29
5.2 FPTP：第一偏好最多者胜	30
5.3 两轮制：先表达，再整合	31
5.4 IRV：把淘汰者的选票往后转	31
5.5 Borda：把排序位置转成分数	32
5.6 Condorcet：逐对比较是否有稳定赢家	33
5.7 Approval：可接受集合如何改变结果	33

目录	3
5.8 不同规则保留了哪些偏好信息	34
6 从赢家到席位：比例代表和混合制度	34
6.1 从单一赢家到席位向量	35
6.2 D'Hondt：除数法怎样偏向大党	35
6.3 Sainte-Lague：让中小党更容易拿到下一席	37
6.4 最大余数法：先按配额取整，再比分数尾巴	38
6.5 MMM 与 MMP：混合制度的差别在哪里	38
6.6 门槛例子：为什么 4.9% 和 5.1% 差很多	39
7 候选人和政党如何适应制度	40
7.1 FPTP：最大化成为第一名的概率	41
7.2 策略投票：选民也在解最优化问题	42
7.3 两轮制：第一轮进决赛，第二轮扩联盟	43
7.4 IRV：不要只看第一名，还要避免被排斥	44
7.5 PR：小党不必赢第一名也能有收益	44
7.6 联盟议价：席位小不等于影响小	45
7.7 MMP：地方候选人与全国品牌分开优化	46
7.8 策略比较的结论	46
8 小型仿真：从选民坐标到制度结果	47
8.1 先算一个选民：排序从效用来	47
8.2 第一步：从效用得到第一偏好	49
8.3 第二步：两轮制和 IRV 怎样把票转给 B	49
8.4 第三步：Borda 和 Condorcet 为什么也选 B	50
8.5 第四步：PR 把集群保留为席位	50
8.6 第五步：政策距离损失	51
9 制度设计权衡	52
9.1 把指标代入同一个例子	52
9.2 把制度评价写成目标函数	54
9.3 指标从哪里来	54

9.4 前沿制度：把 tradeoff 换一种形式	55
9.4.1 Approval-based multiwinner voting	55
9.4.2 Method of Equal Shares	56
9.4.3 Quadratic voting	57
9.4.4 计算复杂性也是制度约束	58
10 结论与制度选择建议	58
10.1 没有脱离目标的最优制度	58
10.2 分析主线的最终表述	59

1 从 Hotelling 出发：位置竞争怎样变成选举模型

Hotelling 空间竞争模型¹ 提供了一个便于分析的起点。设消费者分布在一条线段上，企业选择自己的位置；消费者购买离自己更近的产品，企业通过位置选择争取更多消费者。Hotelling 讨论的正是这种位置竞争为什么会把竞争者推向市场中心 [6]。

把这个模型搬到选举里，只需要换一组名字。企业变成候选人，产品位置变成政策位置，消费者的交通成本变成选民的政策距离损失，市场份额变成得票率。保留下来的核心逻辑是：候选人选择位置，选民按效用作选择，制度把这些选择转化成胜负。

图 1 给出全文路线。第一步建立一维基准模型：如果政策空间只有一条轴，选民偏好是单峰的²，两名候选人只关心胜选，那么中位选民会成为多数比较的稳定点。第二步放松这些假设：候选人有意识形态成本，选民有议题权重，第三党可以进入，制度也不一定是单一赢家。第三步进入社会选择问题：多维偏好在多数规则下可能循环，政党票也可能把多个议题捆绑在一起。最后才比较具体制度和策略，因为制度规则会反过来改变候选人和政党的最优反应。

设政策空间为区间 $[0, 1]$ 。选民 i 的理想点为 x_i ，候选人 j 的政策位置为 y_j 。最简单的距离效用可以写为

$$U_i(y_j) = -|y_j - x_i|,$$

也可以写成二次损失形式

$$U_i(y_j) = -(y_j - x_i)^2.$$

绝对距离和二次距离在这里给出同一个排序：政策点离选民理想点越近，选民越满意。二次形式的好处是后面容易扩展到多维空间。

¹ **Hotelling 空间竞争模型** (Hotelling spatial competition model; 霍特林空间竞争模型)：把消费者或选民放在一条空间线上，竞争者通过选择位置争取需求或选票的模型。

² **单峰的** (single-peaked preferences; 单峰偏好)：每个选民在一维政策线上有一个最理想点，离该点越远偏好越低的偏好结构。

全文论证路线：从一维中位选民到制度设计权衡



图 1: 全文的推导路线：先有一维基准模型，再逐步加入现实约束。

命题 1.1（一维中位选民逻辑）

若政策空间是一维的，选民偏好是单峰的，且集体选择用两两多数决进行，那么中位选民的理想点 m 能在多数比较中击败任何其他政策点。

证明

考虑任意政策点 $y > m$ 。由于 m 是中位数，至少一半选民的理想点不大于 m 。对这些选民来说， m 比 y 更接近自己的理想点，因此他们更喜欢 m 。所以 m 在与 y 的多数比较中获胜。若 $y < m$ ，则至少一半选民的理想点不小于 m ，他们也更喜欢 m 而不是 y 。因此 m 能击败所有其他政策点。 $\square$

Downs³ 把这种空间竞争直觉系统引入民主选举 [3]。如果只有两名候选人，而且候选人只关心胜选，偏离中位点的一方会给对手留下争夺多数的空间。于是最简单的均衡直觉就是 Downsian convergence⁴：两名候选人都靠近中位选民。

³ **Downs** (Downsian spatial voting model; 唐斯空间投票模型)：把候选人平台看成政策空间位置，并分析候选人为赢得选票如何选择位置的选举模型。

⁴ **Downsian convergence** (Downsian convergence; 唐斯式收敛)：在一维、两候选人、胜选最大化等条件下，候选人平台向中位选民靠拢的预测。

1.1 一个五人例子：中位点为什么能赢

先看一个离散例子。政策数值越大，表示越偏向高税收和高公共服务。五名选民的理想点分别为

$$0.10, \quad 0.30, \quad 0.50, \quad 0.70, \quad 0.90.$$

中位选民的理想点是 $m = 0.50$ 。现在比较三个政策方案：

$$L = 0.30, \quad M = 0.50, \quad R = 0.70.$$

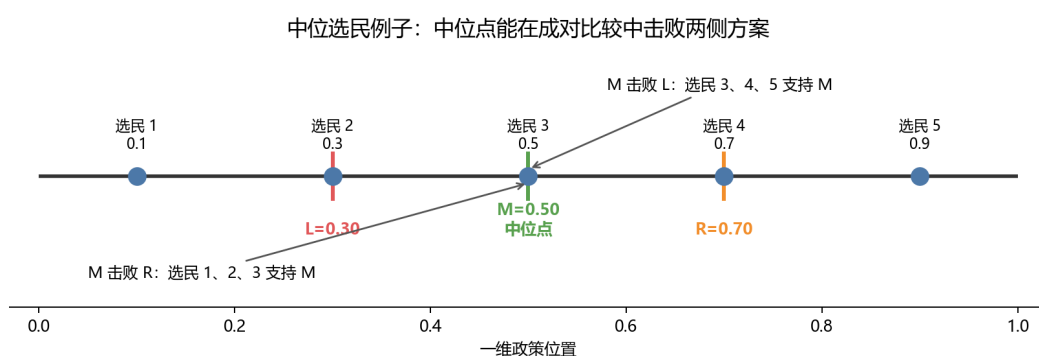


图 2: 五名选民的一维政策线。中位点 M 不让所有人最满意，但能分别击败左右两侧方案。

图 2 比表格更直接。比较 M 和 L ：选民 3、4、5 离 M 更近，选民 1、2 离 L 更近，所以 M 以 3:2 击败 L 。比较 M 和 R ：选民 1、2、3 离 M 更近，选民 4、5 离 R 更近，所以 M 又以 3:2 击败 R 。

表 1: 离散选民中位点的多数比较

选民	理想点 x_i	$d(x_i, L)$	$d(x_i, M)$	$d(x_i, R)$	最接近方案
1	0.10	0.20	0.40	0.60	L
2	0.30	0.00	0.20	0.40	L
3	0.50	0.20	0.00	0.20	M
4	0.70	0.40	0.20	0.00	R
5	0.90	0.60	0.40	0.20	R

命题 1.2（中位政策是 Condorcet winner）

设有奇数个选民，理想点按大小排列为 $x_1 \leq x_2 \leq \dots \leq x_n$ ，其中 $n = 2r + 1$ 。选民偏好是单峰的，且在一维线上离自己的理想点越近越好。令中位理想点为 $m = x_{r+1}$ 。则 m 在成对多数比较中击败任意 $y \neq m$ 。

证明

只证明 $y > m$ 的情形。因为 $m = x_{r+1}$ ，至少有 $r + 1$ 个选民满足 $x_i \leq m$ 。对这些选民来说，

$$|m - x_i| = m - x_i < y - x_i = |y - x_i|.$$

他们都偏好 m 而非 y 。 $r + 1$ 已经是多数，所以 m 击败 y 。若 $y < m$ ，至少有 $r + 1$ 个选民满足 $x_i \geq m$ ，同理他们也偏好 m 。命题得证。 $\square$

1.2 两名候选人为什么向中位点靠

中位选民定理说的是社会多数比较。候选人收敛还需要加入候选人的目标。设两名候选人为 A, B ，政策位置为 y_A, y_B 。候选人只关心胜选，收益为

$$\Pi_A(y_A, y_B) = \begin{cases} 1, & A \text{ 得票超过 } B, \\ \frac{1}{2}, & \text{平票}, \\ 0, & A \text{ 得票少于 } B. \end{cases}$$

如果两人都在 m ，任何一方单独偏离到 m 的左侧或右侧，都会被留在 m 的对手用多数击败。若两人都在中位点右侧，较靠近 m 的一方更有机会拿到左侧和中间选民；若两人都在左侧，对称结论成立。收敛直觉来自这个可抢占空间：离中位点太远，会把多数留给对手。

命题 1.3 (连续均匀分布下的 Downs 收敛)

设选民在 $[0, 1]$ 上均匀分布，效用为 $-|x - y|$ 或 $-(x - y)^2$ ，两名候选人只关心胜选，平票各得 $\frac{1}{2}$ 。若候选人可以选择同一位置，则唯一纯策略 Nash 均衡是

$$y_A = y_B = \frac{1}{2}.$$

证明

先证明 $(1/2, 1/2)$ 是均衡。若 B 留在 $1/2$ ， A 偏到 $y_A < 1/2$ ，无差异点是

$$z = \frac{y_A + 1/2}{2} < \frac{1}{2}.$$

A 得票为 $z < 1/2$ ，输给 B 。若 A 偏到 $y_A > 1/2$ ，它得到右侧选民，得票为

$$1 - \frac{1/2 + y_A}{2} < \frac{1}{2},$$

也会输。偏离没有收益， B 同理。

再证明没有其他均衡。设 $y_A < y_B$ 。无差异点为 $z = (y_A + y_B)/2$ 。若 $z < 1/2$ ，右侧候选人 B 获得多数， A 输。此时 A 可以移动到 $1/2$ 。若 $y_B > 1/2$ ，新无差异点为 $(1/2 + y_B)/2 > 1/2$ ， A 得到左侧超过一半选民，胜出；若 $y_B \leq 1/2$ ， A 移到 $y_B + \varepsilon$ 且仍小于 $1/2$ ，就会成为更

靠右的候选人，得到无差异点右侧的大多数。两种情况下 A 都能从输变赢，所以原位置不是均衡。

若 $z > 1/2$ ，对称地，右侧候选人 B 输， B 可以移动到 $1/2$ 或 $y_A - \varepsilon$ 取得多数。若 $z = 1/2$ 但 $y_A \neq y_B$ ，两人平票；左侧候选人向右靠近 $1/2$ 一点，或右侧候选人向左靠近 $1/2$ 一点，都能获得超过半数。因此也不是均衡。剩下的唯一可能是两人同在 $1/2$ 。□

这个证明比“候选人会抢中间”更精确。候选人不是因为喜欢中间政策才去中间，而是因为任何偏离都会把一半以上选民留给对手。中位点在这里像价格竞争中的 Bertrand 价格：只要对手没有站在中位点，自己就有一个可盈利的微小移动；两人都站到中位点以后，单方移动又会立刻丢掉多数。

例 1.1（连续均匀分布下的得票份额）

设选民在 $[0, 1]$ 上均匀分布，两名候选人位置满足 $y_A < y_B$ 。两者之间的无差异选民位于

$$z = \frac{y_A + y_B}{2}.$$

所有 $x < z$ 的选民更喜欢 A ，所有 $x > z$ 的选民更喜欢 B 。因此

$$V_A = z = \frac{y_A + y_B}{2}, \quad V_B = 1 - z = 1 - \frac{y_A + y_B}{2}.$$

若 $y_A = 0.40, y_B = 0.70$ ，则 $z = 0.55$ ， A 得到 55% 选票。此时 B 的好反应不是继续右移，而是向左靠近中位点。

表 2: 连续均匀分布下的位置与得票

y_A	y_B	无差异点 z	A 得票	B 得票
0.20	0.80	0.50	50%	50%
0.40	0.70	0.55	55%	45%
0.49	0.51	0.50	50%	50%
0.50	0.70	0.60	60%	40%

1.3 偶数选民与中位区间

如果选民人数为偶数，中位点可能不是一个点，而是一个区间。设四名选民理想点为

$$0.20, \quad 0.40, \quad 0.60, \quad 0.80.$$

中间两个理想点是 0.40 和 0.60。任一政策 $y \in [0.40, 0.60]$ 都有一种稳定性：若挑战者在区间左侧，至少后两名选民更喜欢区间内政策；若挑战者在区间右侧，至少前两名选民更喜欢区间内政策。在严格多数规则下，区间内不同点之间可能平票；在需要破平规

则时，破平规则会进一步选择具体点。

例 1.2（偶数选民下的平票与破平）

比较 $y = 0.45$ 与 $z = 0.55$ 。选民 0.20 和 0.40 更接近 0.45，选民 0.60 和 0.80 更接近 0.55，结果 2 : 2。若制度规定平票由现任者胜出，现任者位于 0.45 就能挡住 0.55；若规定随机破平，两个政策的期望胜率相同。这个小差别说明：即使在一维单峰偏好下，细节规则也会影响最终选择。

基准结论可以概括为一句话：在一维、单峰、两候选人、只关心胜选的环境中，候选人会被推向中位点或中位区间。后面的推导每次只改变其中一两个条件，观察结论怎样变形。

2 第一处问题：现实政治为什么不总是中间化

一维中位选民模型给出的预测很清楚：两名只关心胜选的候选人会向中间靠。现实政治却常常出现极化、多党、小党长期存在、弃保、联盟和名单政治。差异来自假设被放松以后，候选人的好反应改变了。

基准模型至少依赖几个条件：政策空间是一维的；选民偏好是单峰的；只有两名候选人；候选人只关心胜选；选民只看政策距离；制度是单赢家；选民真诚投票；候选人进入和退选不需要单独分析。

一旦改变这些条件，结论就会动。政策空间变成多维后，不一定存在一个同时代表所有维度的中位点；候选人有意识形态偏好时，胜选不再是唯一目标；选民关心 valence⁵ 时，能力强但策略远的候选人可能胜过政策更近但能力弱的候选人 [18]；选民按议题方向而不是距离投票时，强烈表达某个方向的候选人可能比温和候选人更有吸引力 [15]；制度从 FPTP⁶ 变成 PR⁷ 后，小党不必成为第一名也能获得席位。

下面把“为什么不收敛”拆成几个可计算的机制：意识形态成本、动员收益、小党进入收益和制度收益断点。

⁵ **valence** (valence；候选人质量或共同评价维度)：多数选民通常同向评价的候选人属性，例如治理能力、廉洁程度、执行力和危机处理能力。

⁶ **FPTP** (First-Past-The-Post；相对多数制)：每位选民投一票，票数最多者胜出，不要求过半的单赢家规则。

⁷ **PR** (Proportional Representation；比例代表制)：把政党或候选人得票尽量按比例转换为多个席位的多赢家制度。

2.1 逐条放松假设：哪些结论会先断掉

第一节的基准模型写清楚以后，可以更精确地问：现实政治为什么不总是中间化？一个直接答案是，模型结论依赖一组条件。只要改变其中一条，候选人的最优反应就会改变。表 3 把主要断点列出来。

表 3: 从 Hotelling-Downs 基准到现实政治的假设放松

基准假设	放松后的现实情形	可能出现的结果
一维政策空间	经济、文化、身份、治理能力同时存在	没有共同中位点，多数循环和议程控制出现
候选人只关心胜选	候选人或政党也关心政策本身	平台不会完全收敛，可能保持意识形态距离
只有两名候选人	候选人进入内生性，小党可以参选	票源分裂、弃保、联盟和多党竞争出现
选民真诚投票	选民考虑可胜性和弃票风险	策略投票、协调失败和“少输一点”的选择出现
制度是单一赢家	PR、MMP、排序投票等制度存在	小党收益、转移票、席位门槛和名单策略改变

表 3 给出后文的展开顺序。最先变化的是政策空间：一旦选民不只关心一条轴，中位点可能不存在。接着变化的是效用函数：选民可能给不同议题不同权重，候选人也可能有 valence 或意识形态。然后进入制度层面：同一组选票在 FPTP、IRV、Borda、PR 下会被压缩成不同结果。最后是策略层面：候选人和选民预期这些规则以后，会改变自己的行为。

例 2.1（例 3：意识形态候选人为什么不一定收敛）

设一维中位点为 $m = 0.5$ ，候选人 A 的个人理想政策为 $z_A = 0.2$ 。若候选人的收益不是单纯胜选，而是

$$\Pi_A(y_A, y_B) = P_A(y_A, y_B)R - \gamma(y_A - z_A)^2,$$

其中 P_A 是胜选概率， R 是职位收益， γ 是政策偏离成本。当 R 很大或 γ 很小时，候选人会靠近中位点；当 γ 很大时，靠近 0.5 的政策损失可能超过胜选概率提升。比如 $R = 1, \gamma = 8$ ，从 0.2 移到 0.5 的政策成本为 $8(0.3)^2 = 0.72$ 。如果移动只把胜选概率提高 0.3，则职位收益增加为 0.3，小于政策成本，候选人不会完全收敛。

这个例子说明，“候选人为什么不向中间靠”不必用非理性解释。只要候选人有政策偏好，他就在职位收益和政策损失之间做权衡。政治经济学中，候选人可以被看作既追求 office，又追求 policy 的行为人；政党则更明显，因为政党组织、党员、捐助者和长期品牌都会让平台偏离短期中位点。

例 2.2（例 4：动员模型下的极化）

设两名候选人在一维线上竞争，温和位置 0.5 能吸引较多中间选民，但核心支持者的投票率取决于候选人离本阵营理想点的接近程度。令左翼核心人数为 40，理想点 0.2；右翼核心人数为 40，理想点 0.8；中间选民人数为 20，理想点 0.5。若候选人 A 选择 0.5，他获得中间选民 20 人，但左翼核心投票率只有 50%，即 20 人，总票约 40。若 A 选择 0.25，他失去部分中间选民，只获得 8 人，但左翼核心投票率升到 90%，即 36 人，总票约 44。于是极化位置反而更优。

这个计算背后的机制是：候选人最大化的不是“偏好上支持自己的人数”，而是实际投票数。若参与成本、热情、组织动员和捐款都取决于偏好强度，候选人可能选择离中位点更远的位置，以提高核心群体的行动强度。

例 2.3（例 5：小党进入为什么在 FPTP 下困难）

设有三个政治产品：左翼 L 、中间 M 、右翼 R 。选民第一偏好比例为

$$L = 38\%, \quad M = 24\%, \quad R = 38\%.$$

若只有 L 和 R 参选，中间选民可能在两者之间分散，胜负接近。若 M 作为第三党进入 FPTP，它虽然有 24% 第一偏好，但不是第一名，席位收益为 0。若进入成本为 $F > 0$ ，它的短期收益就是 $-F$ 。相同得票在 10 席 PR 中， M 大约可获得 2 席；只要席位价值 $2r > F$ ，进入就是理性的。

这个例子把 Duverger logic 写成了一个简单收益比较。FPTP 的问题不只是“第一名通吃”，而是它把得票率函数变成一个不连续的收益函数：24% 与 0% 在席位上可能没有区别。PR 则把票数更连续地转化为席位，降低小党进入门槛。

例 2.4（弃保可以写成协调博弈）

继续使用 L, M, R 三个政治产品。设右翼 R 有稳定 40 票，左翼 L 有 35 票，中间 M 有 25 票。若 L, M 都参选且选民真诚投票，FPTP 结果是

$$R = 40, \quad L = 35, \quad M = 25,$$

R 胜出。对 M 选民来说，最喜欢 M ，其次是 L ，最不喜欢 R 。现在 M 选民要决定是否继续投 M ，还是策略性转投 L 。

设 M 选民中有 t 人转投 L 。票数变为

$$L = 35 + t, \quad M = 25 - t, \quad R = 40.$$

只要

$$35 + t > 40,$$

即

$$t > 5,$$

L 就能击败 R 。因此，至少 6 名 M 选民需要协调转投。若只有 3 人转投，结果是 $L = 38, R = 40$ ，他们既没保住 M ，也没阻止 R 。弃保不是单个选民的简单计算，而是一个协调问题：每个人都希望“足够多人一起转”，又担心自己单独转没有效果。

表 4: 中间选民的弃保协调

M 选民转投人数 t	票数结果	最终赢家
0	$L = 35, M = 25, R = 40$	R
3	$L = 38, M = 22, R = 40$	R
6	$L = 41, M = 19, R = 40$	L
25	$L = 60, M = 0, R = 40$	L

这个表把 Duverger 所说的心理效应具体化了。机械效应是 FPTP 的席位转换：第三名没有席位。心理效应是选民和政党预期到机械效应以后改变行为：选民弃保，政党退选或合并，捐助者把钱投向可胜者。两者叠在一起，才会形成两大阵营竞争的压力。PR 下，同样的 25 票可以换成席位，协调压力就会弱得多。

2.2 从“偏离中位”到“制度诱导”

上面三个例子分别改变候选人目标函数、选民参与函数和席位转换函数。它们共同说明：选举制度会提前进入候选人和选民的预期。FPTP 下，第三党知道自己可能没有席位，于是可能不进入；选民知道第三党可能无胜算，于是可能弃保；大党知道小党容易被挤出，于是有更强动力争夺中间选民。PR 下，这些激励都变弱，政党可以围绕更窄的政策群体建立品牌。

后面的分析不能只问“哪种制度算出来谁赢”，还要问“候选人在这种制度下会怎样定位，选民会怎样表达偏好，政党会不会进入，联盟会在哪里形成”。制度先改变约束和收益函数，行为人再对新约束作最优反应，最后形成观察到的政治结果。

3 多维政策空间与选民效用

3.1 从一维政策到多维政策

现实选举很少只围绕一个议题展开。一个候选人的政策平台通常包含经济政策、文化身份、治理能力、地方利益、外交安全等多个维度。先把政策写成二维向量：

$$y_j = (E_j, C_j),$$

其中 E 表示经济政策维度，例如税收、福利、市场化和再分配； C 表示文化或身份维度，例如保守与自由、民族与全球、宗教、移民、地方认同等。

再进一步，可以加入治理能力或候选人质量：

$$y_j = (E_j, C_j, G_j),$$

其中 G 表示治理能力、反腐、行政效率、地方服务或候选人 valence。选民 i 的理想点也变成

$$x_i = (x_{iE}, x_{iC}, x_{iG}).$$

这一步只抓住一个问题：选民的投票选择来自多个属性的综合评价。

3.2 加权二次距离效用

最直接的多维效用函数是加权二次距离：

$$U_i(y_j) = - \sum_k w_{ik} (y_{jk} - x_{ik})^2 + \lambda_i q_j + \varepsilon_{ij}.$$

这里， w_{ik} 是选民 i 对议题 k 的权重，反映 issue salience⁸； x_{ik} 是选民的理想点； y_{jk} 是候选人的政策位置； q_j 是候选人的能力、廉洁、个人魅力或治理质量； λ_i 表示选民对 valence 的重视程度； ε_{ij} 表示信息误差、情绪或身份认同带来的随机项。

这个式子比一维模型多了两点。第一，不同选民对议题权重不同。一个选民可能把经济再分配看得最重，另一个选民可能更重视文化身份或地方治理。第二，候选人质量可以独立影响投票。即便某候选人的政策位置不是最近，如果其治理能力明显更强，仍可能获得支持。

可以用一个简单例子说明权重的作用。设两名候选人 A, B 在经济维度和文化维度上的位置分别为

$$A = (0.2, 0.8), \quad B = (0.8, 0.2).$$

某选民的理想点是 $x_i = (0.4, 0.4)$ 。如果他对两个议题同等重视，则二者距离相近；但若他对经济议题权重很高， $w_{iE} = 0.8, w_{iC} = 0.2$ ，则经济位置更接近的 A 可能更有吸引力；若他对文化议题权重很高，则结论可能反过来。这样，两个拥有相同理想点坐标的选民，只要议题权重不同，也可能做出不同投票选择。

这构成了现实选举中“交叉裂痕”的来源。一个选民在经济上支持福利国家，在文化上偏保守；另一个选民在经济上支持市场化，在文化上偏自由。若政党把经济和文化立场捆绑成 policy bundle，选民就必须在多个维度之间取舍。选举制度则决定这种取舍是被压缩成一张单一选票，还是可以通过排序、批准或比例代表表达得更细。

⁸ **issue salience** (issue salience; 议题显著性)：某个议题在选民决策中占多大权重，权重越高，该议题偏离理想点造成的效用损失越大。

3.3 满意度聚合与底线议题

Cobb-Douglas⁹、CES¹⁰ 和 Stone-Geary¹¹ 不能直接套到政策坐标上，因为政策坐标不是普通商品，不是越多越好。更自然的做法是先把每个议题转化为满意度：

$$s_{ik}(y_{jk}) = \max\{0, 1 - a_k|y_{jk} - x_{ik}|\}.$$

当候选人在某议题上接近选民理想点时，满意度接近 1；距离过远时，满意度下降到 0。

有了满意度之后，可以用 CES 形式聚合：

$$U_i(y_j) = \left(\sum_k \alpha_{ik} s_{ik}(y_{jk})^\rho \right)^{1/\rho}.$$

当 ρ 较高时，不同议题之间替代性较强，一个议题的不满可以被另一个议题的满意弥补；当 ρ 较低时，议题之间互补性更强，某一维度太差会严重拉低总体评价；当 $\rho \rightarrow -\infty$ 时，效用接近 Leontief¹²，最差的议题决定整体态度。

Cobb-Douglas 形式可以写成

$$U_i(y_j) = \prod_k s_{ik}(y_{jk})^{\alpha_{ik}},$$

表示各方面都重要，某个维度为零会显著压低总效用。Stone-Geary 形式则可以写成

$$U_i(y_j) = \prod_k [s_{ik}(y_{jk}) - \gamma_{ik}]^{\alpha_{ik}},$$

其中 γ_{ik} 是底线要求。若某候选人在移民、税收、腐败或地方利益等议题上低于最低可接受水平，选民可能直接排除该候选人。

多属性效用理论本来就是处理这类多目标选择的工具 [7]。放到选举中，它说明：选民不是简单地左或右，而是在多个议题之间进行权衡；不同议题之间可能替代、互补，也可能存在底线约束。

这种效用函数为分析候选人定位问题提供了工具。若大多数选民的议题之间替代性强，候选人可以在某些议题上偏离中位点，用另一些议题补偿；若许多选民存在底线议题，候选人就必须避免触碰这些底线，否则其他议题再接近也难以获得支持。于是，候选人的最优策略不再只是“靠近中位数”，而是选择一个能同时平衡距离、权重、valence 和底线约束的位置。

⁹ **Cobb-Douglas** (Cobb-Douglas utility function; 柯布-道格拉斯效用函数)：用各维度满意度的幂乘积表示总体效用，常用来刻画多个维度都重要的选择问题。

¹⁰ **CES** (Constant Elasticity of Substitution utility function; 常替代弹性效用函数)：用一个参数控制不同维度之间替代性的效用函数。

¹¹ **Stone-Geary** (Stone-Geary utility function; 斯通-吉尔里效用函数)：在 Cobb-Douglas 形式中加入最低需求或底线参数的效用函数。

¹² **Leontief** (Leontief utility function; 列昂惕夫效用函数)：由最短板维度决定总体效用的完全互补型效用函数。

3.4 议题权重如何改变同一理想点的选择

多维模型最容易被误解的一点，是把“理想点”当作唯一信息。实际上，理想点只是每个维度上的目标位置；选民还要决定每个维度有多重要。两个选民可以有相同理想点，却因为 issue salience 不同而投给不同候选人。

例 3.1（例 6：同一理想点，权重不同，投票不同）

设候选人位置为

$$A = (0.2, 0.8), \quad B = (0.8, 0.2),$$

其中第一维是经济 E ，第二维是文化 C 。选民理想点为 $x = (0.4, 0.4)$ 。加权损失函数为

$$L_j = w_E(E_j - x_E)^2 + w_C(C_j - x_C)^2,$$

损失越小越好。若选民更重视经济，取 $w_E = 0.8, w_C = 0.2$ ；若更重视文化，取 $w_E = 0.2, w_C = 0.8$ 。计算见表 5。

表 5: 议题权重改变候选人排序

权重情形	候选人	经济差距平方	文化差距平方	加权损失	选择
$w_E = 0.8, w_C = 0.2$	A	0.04	0.16	0.064	A
$w_E = 0.8, w_C = 0.2$	B	0.16	0.04	0.136	
$w_E = 0.2, w_C = 0.8$	A	0.04	0.16	0.136	
$w_E = 0.2, w_C = 0.8$	B	0.16	0.04	0.064	B

这个例子解释了为什么多维政策空间中不能简单说某个候选人“更接近中间”。如果中间是一个几何点，那么 A 和 B 到 x 的欧氏距离相同；但选民并不是用无权重欧氏距离投票。议题权重改变了不同维度的边际替代关系，于是同一个理想点也可能导向不同选择。

3.5 方向模型：选民有时关心“往哪边推”

距离模型假设选民最喜欢自己的理想点，候选人离理想点越远越差。但政治态度有时更像方向选择。选民未必要求候选人正好落在自己的理想点上，而是希望候选人把政策从现状往自己支持的方向推动。Rabinowitz 和 Macdonald 的 directional theory 即指出这一点 [15]。

设现状为原点 0，选民的方向向量为 x_i ，候选人的政策方向为 y_j 。一个简单方向效用可以写成

$$U_{ij} = x_i \cdot y_j - \kappa \|y_j\|^2,$$

其中 $x_i \cdot y_j$ 是方向一致性, $\kappa \|y_j\|^2$ 表示候选人过度激进带来的惩罚。若候选人方向与选民一致, 点积为正; 方向相反, 点积为负。

例 3.2 (距离近的人未必赢, 方向对的人可能赢)

设政策是一维, 现状为 0, 选民希望政策向右, 方向强度为 $x_i = 1$ 。候选人 A 位置为 0.2, 候选人 B 位置为 0.6。距离模型若选民理想点为 0.3, 会给出

$$L_A = (0.2 - 0.3)^2 = 0.01, \quad L_B = (0.6 - 0.3)^2 = 0.09,$$

所以 A 更好。方向模型取 $\kappa = 0.5$, 则

$$U_A = 1 \cdot 0.2 - 0.5(0.2)^2 = 0.18,$$

$$U_B = 1 \cdot 0.6 - 0.5(0.6)^2 = 0.42.$$

B 更好, 因为它更有力地把政策往选民支持的方向推动。若 B 进一步变成 1.5, 效用为

$$U_B = 1.5 - 0.5(1.5)^2 = 0.375,$$

仍高于 A, 但过度激进的惩罚已经开始压低收益。

方向模型对后文有两个用处。第一, 它解释为什么候选人不一定围绕中位点做距离最小化; 在某些议题上, 选民奖励的是“方向鲜明”。第二, 它能解释动员政治: 强方向候选人可能让核心支持者更愿意投票、捐款和传播。第七章讨论候选人策略时, 侧翼政党保持差异化就可以用这种方向收益理解。

3.6 Valence: 政策距离之外的候选人质量

多维效用的第二个关键变量是 valence。它指候选人被多数选民共同认为更好或更差的属性, 例如治理能力、廉洁程度、执行力、危机处理能力。valence 与普通政策维度不同: 税率高低可能有人喜欢有人讨厌, 但更廉洁通常不会被大多数选民讨厌。

例 3.3 (valence 如何弥补政策距离)

沿用例 6 的经济权重情形, 选民对 A 的政策损失为 0.064, 对 B 的政策损失为 0.136。若没有 valence, 选民选择 A。现在假设 B 的治理质量优势为 $q_B = 0.10$, A 的质量为 $q_A = 0$, 且选民对质量的权重 $\lambda = 1$ 。效用写为

$$U_j = -L_j + \lambda q_j.$$

则

$$U_A = -0.064, \quad U_B = -0.136 + 0.10 = -0.036.$$

尽管 B 在政策上更远, 加入治理质量以后 B 的总效用更高。

表 6: valence 弥补政策距离的计算

候选人	政策损失 L_j	valence q_j	总效用 $-L_j + q_j$	结果
A	0.064	0.00	-0.064	
B	0.136	0.10	-0.036	胜出

这个小计算能解释许多看似“违反政策距离”的投票行为。选民并非忘记了政策，而是在政策距离和候选人质量之间做替代。对候选人来说，valence 会改变定位策略：高 valence 候选人可以承受更大政策偏离；低 valence 候选人则必须更靠近目标选民，才能补偿质量劣势。

3.7 CES、Cobb-Douglas 与底线议题

加权二次距离假设不同维度可以线性加总。但现实中，选民可能不是这样聚合议题。有些议题可以替代：候选人在经济上略远，但治理能力更强，选民可以接受。有些议题更像互补品：候选人在经济、文化、廉洁三方面都不能太差，否则某一短板会拖垮整体评价。CES 和 Cobb-Douglas 的价值在于，它们能把“替代”和“互补”写成参数。

例 3.4（例 8：CES 聚合下的均衡型候选人）

设选民对两个议题的满意度都在 $[0, 1]$ 。候选人 A 的满意度向量为 $(0.95, 0.35)$ ，候选人 B 为 $(0.70, 0.70)$ 。若选民使用简单加总，且权重相等，则

$$S_A = 0.95 + 0.35 = 1.30, \quad S_B = 0.70 + 0.70 = 1.40,$$

B 略优。若第一个议题权重非常高， A 可能胜出。但若选民采用 Cobb-Douglas，

$$U_A = \sqrt{0.95 \times 0.35} \approx 0.577, \quad U_B = \sqrt{0.70 \times 0.70} = 0.700.$$

此时 B 的优势明显增大，因为 Cobb-Douglas 惩罚单一维度过低的方案。

表 7: 不同聚合方式下的候选人比较

候选人	议题 1 满意度	议题 2 满意度	加总满意度	Cobb-Douglas 满意度
A	0.95	0.35	1.30	0.577
B	0.70	0.70	1.40	0.700

例 3.5（例 9：Stone-Geary 底线议题导致候选人被排除）

继续使用满意度向量。设选民对议题 2 有底线要求 $\gamma_2 = 0.50$ 。Stone-Geary 型评价可以写

成

$$U_j = (s_{j1} - \gamma_1)^{1/2}(s_{j2} - \gamma_2)^{1/2},$$

其中 $\gamma_1 = 0$ 。候选人 A 的第二维满意度是 0.35，低于底线 0.50，因此即使第一维高达 0.95，也被选民排除。候选人 B 两维都高于底线，成为可接受选择。

底线议题能解释为什么候选人有时不会在所有议题上做边际权衡。若某一议题触发“不可接受”，其他议题的优势很难补偿。这一点与普通消费者理论中的 Stone-Geary 最低消费要求类似：在最低要求没满足前，其他商品再多也不能完全替代。

3.8 多维效用如何导向制度问题

多维效用本身还不是选举制度问题。它只是说明选民的真实偏好结构比一维模型复杂。制度问题出现在下一步：这些复杂偏好要通过选票表达。FPTP 只问“你最喜欢谁”；Borda 要求完整排序；approval 允许表达多个可接受候选人；PR 则把不同群体的第一偏好转化为席位。也就是说，制度相当于一个信息压缩函数：

多维效用和排序 $\longrightarrow$ 选票格式 $\longrightarrow$ 胜者或席位。

如果偏好本来就是多维的，那么任何制度都会丢失一部分信息。差别在于：它丢掉的是第二偏好、偏好强度、比例结构，还是候选人之间的成对比较信息。下一节的社会选择悖论会说明，即使个体偏好完全理性，这个压缩过程也可能产生集体层面的不稳定。

后面几章会反复用到这里的函数。加权二次距离用于第八章仿真，决定每个选民怎样给候选人排序；valence 解释为什么 B 虽然不是最大第一偏好集团的候选人，却能凭质量项和中间位置获得广泛支持；底线议题则进入第七章的策略讨论，说明候选人有时不能只靠“靠近中间”解决问题，因为某些群体会先判断候选人是否越过最低可接受线。

4 个体理性为何不能保证社会选择稳定

4.1 二维 Condorcet 循环

多维偏好的第一个问题是，多数关系可能循环。这里的循环称为 Condorcet 循环¹³。考虑三名候选人 A, B, C ，政策位置分别为

$$A = (0, 1), \quad B = (-0.866, -0.5), \quad C = (0.866, -0.5).$$

三类人数相同的选民有如下理想点和排序：

¹³ **Condorcet 循环** (Condorcet cycle; 孔多塞循环)：三个或更多选项中，成对多数关系形成 A 胜 B 、 B 胜 C 、 C 又胜 A 的循环。

选民类型	理想点	排序
1	$(-0.1, 0.9)$	$A \succ B \succ C$
2	$(-0.7, -0.6)$	$B \succ C \succ A$
3	$(0.7, -0.3)$	$C \succ A \succ B$

逐对比较可得：类型 1 和类型 3 更喜欢 A 而不是 B ，所以 A 击败 B ；类型 1 和类型 2 更喜欢 B 而不是 C ，所以 B 击败 C ；类型 2 和类型 3 更喜欢 C 而不是 A ，所以 C 击败 A 。因此

$$A \succ_{\text{maj}} B, \quad B \succ_{\text{maj}} C, \quad C \succ_{\text{maj}} A.$$

每个选民自己的偏好都是传递的，但社会多数关系循环。循环的原因是，不同候选人对之间形成的是不同多数联盟。Plott 和 McKelvey 的多维投票研究说明，多维空间中的多数决稳定性需要很强条件；在一般情形下，议程控制和路径依赖会变得重要 [14, 12]。

4.2 Ostrogorski / Anscombe 悖论

多维偏好的第二个问题是，政党整体胜利未必代表逐项议题多数。设两个政党 P, Q ，三个议题 E, C, G ，五组同等规模选民。表中写的是每组在各议题上更支持哪个政党：

选民组	E	C	G	整体投票
1	P	P	Q	P
2	P	Q	P	P
3	Q	P	P	P
4	Q	Q	Q	Q
5	Q	Q	Q	Q

如果选民按“三个议题中支持较多者”选择整体政党，那么组 1、2、3 投 P ，组 4、5 投 Q ，所以 P 以 3 : 2 赢得政党选举。但逐项看，议题 E 上 Q 得到组 3、4、5 支持；议题 C 上 Q 得到组 2、4、5 支持；议题 G 上 Q 得到组 1、4、5 支持。因此 Q 在每个议题上都以 3 : 2 获得多数。

这就是 Ostrogorski 悖论¹⁴ 的核心思想 [16]。政党是 policy bundle¹⁵，选民投给一个政党，并不等于逐项授权该政党的所有政策。现实政治中的政党名单、联盟协议和议题捆绑，都可能产生类似问题。

¹⁴ **Ostrogorski 悖论** (Ostrogorski paradox; 奥斯特罗戈尔斯基悖论)：政党整体投票结果可能与逐项议题多数方向相反的多议题社会选择悖论。

¹⁵ **policy bundle** (policy bundle; 政策组合)：候选人或政党把多个议题立场捆绑成一个整体平台，选民只能对组合做选择。

4.3 不可能性作为背景

Arrow 不可能定理¹⁶从社会排序层面说明, 个人理性偏好不能简单汇总成满足一组自然条件的社会排序 [1]。Gibbard–Satterthwaite 定理¹⁷则从策略投票层面说明, 在至少三个候选人和 unrestricted domain¹⁸ 下, 非独裁、onto¹⁹、确定性且 strategy-proof²⁰ 的单赢家规则不存在 [5, 17]。

这些定理给出一个清楚背景: 选举制度不是中性容器。制度必须决定如何压缩多维偏好, 如何处理循环, 如何面对策略投票, 如何在代表性和稳定性之间取舍。

4.4 May 定理: 两选项多数决为什么特殊

在进入 Condorcet 循环之前, 可以先看一个正面结果。若只有两个备选项 x, y , 简单多数决有非常好的性质。May 定理说明, 在若干自然条件下, 简单多数决几乎是唯一规则 [11]。

定义 4.1 (两选项社会选择规则)

设每个选民在 x 与 y 之间选择一个, 社会规则输出 x, y 或平局。记支持 x 的人数为 n_x , 支持 y 的人数为 n_y 。

May 使用的三个核心条件很直观。

1. 匿名性: 规则不看选民身份, 只看多少人支持 x 、多少人支持 y 。
2. 中立性: 规则不偏向某个选项。若把所有选民对 x, y 的选择互换, 社会结果也应互换。
3. 正响应性: 如果 x 已经不输, 而某个原本不支持 x 的选民转而支持 x , 社会结果应变得更有利于 x 。

定理 4.1 (May 定理)

在两个备选项的情形下, 满足匿名性、中立性和正响应性的社会选择规则就是简单多数决: 若 $n_x > n_y$, 选 x ; 若 $n_y > n_x$, 选 y ; 若 $n_x = n_y$, 平局。

¹⁶ **Arrow 不可能定理** (Arrow impossibility theorem; 阿罗不可能定理): 在偏好域足够广时, 不存在同时满足若干自然公理的非独裁社会排序规则。

¹⁷ **Gibbard–Satterthwaite 定理** (Gibbard–Satterthwaite theorem; 吉巴德-萨特思韦特定理): 在至少三个备选项的一般偏好域上, 任何确定性、满射且防策略操纵的单赢家规则都是独裁的。

¹⁸ **unrestricted domain** (unrestricted preference domain; 无限制偏好域): 允许选民提交任意完整且传递排序的偏好域。

¹⁹ **onto** (onto social choice function; 满射性): 每个备选项都能在某些偏好剖面下成为规则输出的性质。

²⁰ **strategy-proof** (strategy-proofness; 防策略操纵性): 任一选民都不能通过虚报偏好让结果在其真实偏好下变得更好。

证明

匿名性先把问题大大压缩。规则不能看“谁”支持 x ，只能看票数差

$$d = n_x - n_y.$$

中立性要求规则对 x, y 对称：如果 $d > 0$ 时选 x ，那么 $d < 0$ 时必须选 y ；如果 $d = 0$ ，互换 x, y 后情况完全一样，所以不能偏向任何一方，只能平局。

还剩下要说明的是：当 $d > 0$ 时规则不能平局或选 y 。先看 $d = 1$ 。若 $n_x = n_y$ ，由中立性得到平局。现在让一个支持 y 的选民改为支持 x ，票数差从 0 变为 2；正响应性要求结果严格向 x 改善，所以必须选 x 。若选民总数为奇数，也可以从 $d = -1$ 的 y 胜状态切换一票到 $d = 1$ ，同样得到 x 胜。继续使用正响应性， d 越大越不可能从 x 胜退回平局或 y 胜。因此所有 $d > 0$ 都选 x ，所有 $d < 0$ 都选 y ， $d = 0$ 平局。此即简单多数决。□

例 4.1（两选项时，多数决没有循环）

设 5 名选民在 x, y 之间选择，3 人支持 x ，2 人支持 y 。匿名性意味着哪 3 个人支持 x 不影响结果；中立性意味着若票数改成 2 比 3，结果应对称地变成 y ；正响应性意味着若 x 从 3 票增加到 4 票，结果不能变差。简单多数决完全符合这些要求。

May 定理的意义在于，它把问题边界画清楚了。两选项多数决很干净：匿名、中立、响应性都能同时满足。麻烦从三个或更多选项开始。因为社会不再只需要比较 x 和 y ，还要同时处理 x, y, z 之间的传递关系。下面的 Condorcet 循环即这个转折点：每一次两两比较都像 May 定理那样合理，但三次比较合在一起却形成了环。

可以把三选项多数决想成一张有方向的图。每个候选人是一个点；若多数选民在 A 与 B 之间选择 A ，就画一条 $A \rightarrow B$ 的箭头。两选项时，图中只有两个点和一条箭头，不可能循环。三选项时，图可能变成

$$A \rightarrow B, \quad B \rightarrow C, \quad C \rightarrow A.$$

这张图没有“最高点”。若把 A 当赢家，马上有 C 击败它；若把 C 当赢家，又有 B 击败它；若把 B 当赢家，又有 A 击败它。

这里的困难不是某一次多数比较错了。每条箭头都来自普通多数决，都满足 May 定理背后的匿名性和中立性。困难出在组合以后：社会排序要求箭头不能成环，但多数比较只保证每一对候选人的局部结论。局部合理不自动推出整体传递：每个局部选择看起来都可辩护，拼在一起却可能没有全局最大值。

4.5 二维距离如何生成 Condorcet 循环

Condorcet 循环最常见的呈现方式是直接给出三类选民排序：

$$A \succ B \succ C, \quad B \succ C \succ A, \quad C \succ A \succ B.$$

这种写法很简洁，但容易让人以为循环是人为编出来的。事实上，循环可以从二维政策空间中的普通距离偏好自然生成。设三名候选人的位置为

$$A = (0, 1), \quad B = (-0.866, -0.5), \quad C = (0.866, -0.5).$$

三类同等规模的选民理想点分别为

$$v_1 = (-0.1, 0.9), \quad v_2 = (-0.7, -0.6), \quad v_3 = (0.7, -0.3).$$

按欧氏距离排序，得到表 8.

表 8: 二维空间中的距离与排序

选民类型	$d(A)$	$d(B)$	$d(C)$	距离排序
$v_1 = (-0.1, 0.9)$	0.141	1.596	1.701	$A \succ B \succ C$
$v_2 = (-0.7, -0.6)$	1.747	0.194	1.569	$B \succ C \succ A$
$v_3 = (0.7, -0.3)$	1.477	1.579	0.260	$C \succ A \succ B$

例 4.2 (例 10: 从距离表到多数循环)

三类选民人数相同。比较 A 与 B : 类型 v_1 和 v_3 都把 A 排在 B 前面，所以 A 以 2:1 击败 B 。比较 B 与 C : 类型 v_1 和 v_2 支持 B ，所以 B 以 2:1 击败 C 。比较 C 与 A : 类型 v_2 和 v_3 支持 C ，所以 C 以 2:1 击败 A 。于是

$$A \succ_{\text{maj}} B, \quad B \succ_{\text{maj}} C, \quad C \succ_{\text{maj}} A.$$

多数关系矩阵可以写成表 9。行候选人与列候选人相比，单元格表示行候选人得到的票数。

表 9: Condorcet 循环的多数矩阵

行对列	A	B	C
A	—	2	1
B	1	—	2
C	2	1	—

这个矩阵说明，社会多数关系不是一个可以最大化的单峰目标函数。每个个体的偏好都来自清楚的二维距离，所以个体层面没有“不理性”；问题出在多数聚合本身。多数联盟在不同候选人对之间切换：支持 A 击败 B 的联盟是 $v_1 + v_3$ ，支持 B 击败 C 的联盟是 $v_1 + v_2$ ，支持 C 击败 A 的联盟是 $v_2 + v_3$ 。没有一个稳定多数同时支持全部比较。

4.6 议程控制：同一偏好，不同投票顺序

多数循环一旦存在，投票顺序就不再是中性的。若议程制定者可以决定先比较哪两个方案，再让胜者与第三个方案比较，那么他可以选择最终赢家。

表 10: 议程顺序如何改变最终赢家

第一轮	第一轮赢家	第二轮	最终赢家
A 对 B	A	A 对 C	C
B 对 C	B	B 对 A	A
C 对 A	C	C 对 B	B

例 4.3 (例 11: 为什么议程制定者可以“安排”赢家)

若议程制定者想让 C 赢，就先安排 A 对 B 。由于 A 击败 B ，第一轮后 B 被淘汰；第二轮 C 击败 A ，所以 C 成为最终赢家。若他想让 A 赢，就先安排 B 对 C ；若想让 B 赢，就先安排 C 对 A 。这不是因为选民偏好变了，而是因为多数循环让“路径”本身有了权力。

议程控制相当于先选择博弈树。参与者偏好固定，但博弈树不同，子博弈完美均衡结果可以不同。由此也能理解为什么委员会、议会程序、修正案顺序和表决规则并不是细枝末节：它们决定了社会选择函数的实际输入顺序。

4.7 Ostrogorski / Anscombe 悖论的完整计算

Ostrogorski / Anscombe 悖论关注另一种信息压缩问题：选民在多个议题上有偏好，但制度要求他们投给一个整体政党。设两党为 P, Q ，议题为经济 E 、文化 C 、治理 G ，有五类同等规模选民。表 11 中每一格表示该类型选民在某议题上支持哪个党。

表 11: 政党整体投票与逐议题多数的冲突

选民类型	经济 E	文化 C	治理 G	整体投票
1	P	P	Q	P
2	P	Q	P	P
3	Q	P	P	P
4	Q	Q	Q	Q
5	Q	Q	Q	Q

命题 4.1 (Ostrogorski / Anscombe 悖论的多数反转)

在表 11 的偏好结构中，若选民按整体政党投票， P 以 3 : 2 胜出；但若三个议题分别表决， Q 在每个议题上都以 3 : 2 胜出。

证明

整体投票中，类型 1、2、3 投给 P ，类型 4、5 投给 Q ，所以 P 得 3 票， Q 得 2 票。逐议题表决时，经济议题上支持 P 的是类型 1、2，支持 Q 的是类型 3、4、5，所以 Q 以 3:2 胜出。文化议题上支持 P 的是类型 1、3，支持 Q 的是类型 2、4、5，所以 Q 以 3:2 胜出。治理议题上支持 P 的是类型 2、3，支持 Q 的是类型 1、4、5，所以 Q 仍以 3:2 胜出。故整体政党胜者与逐议题多数胜者相反。□

这个悖论和前面的 Condorcet 循环不同。Condorcet 循环指出成对比较不传递；Ostrogorski / Anscombe 悖论指出多议题被捆绑成政党票以后，整体授权可能与逐项多数相反。二者共同说明：社会选择的困难不是“选民不懂”，而是多维信息被制度压缩时，不同压缩方式保留的信息不同。

4.8 从多数循环走向不可能性定理

Condorcet 循环和 Ostrogorski / Anscombe 悖论已经说明，多维偏好经过多数规则或政党捆绑以后会出现扭曲。Arrow 和 Gibbard-Satterthwaite 定理把这个观察推进一步：当候选人或政策方案超过两个，且偏好域足够广时，不存在一个同时满足所有好性质的确定性规则。Arrow 讨论的是从个人排序到社会排序的聚合；Gibbard-Satterthwaite 讨论的是单赢家规则的防操纵性。前者关心“社会能不能形成一致排序”，后者关心“选民有没有动机虚报排序”。

例 4.4（例 12：一个简单操纵例子）

三名选民、三个候选人，使用 plurality，平票按 $A \succ B \succ C$ 破平。真实偏好为

$$1: A \succ B \succ C, \quad 2: B \succ A \succ C, \quad 3: C \succ B \succ A.$$

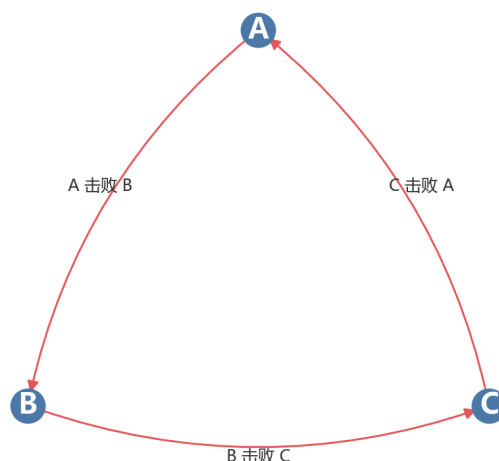
若真诚投票，三人分别投 A, B, C ，形成 1:1:1，按破平规则 A 胜出。对选民 3 来说， A 是最差结果。若选民 3 策略性地投给 B ，票数变成 $A = 1, B = 2, C = 0$ ， B 胜出。因为 $B \succ_3 A$ ，选民 3 通过不真诚报告改善了结果。

这个例子不是 Gibbard-Satterthwaite 定理的完整证明，但它展示了定理的直觉。只要制度把复杂排序压缩成一个胜者，而某些候选人“可胜”与“不可胜”的预期不同，选民就可能不把第一偏好作为最优行动。第五节开始比较具体制度时，策略投票将反复出现。

4.9 从循环图看多数关系的问题

图 3 把前面的多数矩阵画成了一个环。只要社会多数关系形成环，制度就必须额外决定怎样破坏。议程控制、Borda 积分、Condorcet 系破解规则、随机化规则，都可以

Condorcet 循环：成对多数关系形成环

图 3: Condorcet 循环：多数关系可以形成 $A \succ B \succ C \succ A$ 。

看作不同的破坏办法。Arrow 和 Gibbard–Satterthwaite 定理则说明：如果偏好域足够广，想同时保留所有看起来合理的性质，会碰到硬限制。

4.10 Arrow 不可能定理

Arrow 研究的是从个人排序到社会排序的规则。设候选人集合至少有三个备选项。每个选民给出一个完整且传递的排序，社会福利函数²¹把这些个人排序聚合成一个社会排序。

常见公理有四个。

1. Pareto：如果所有人都认为 $x \succ y$ ，社会也应认为 $x \succ y$ 。
2. IIA²²：社会对 x, y 的排序只取决于个人对 x, y 的相对排序，不应受第三个选项 z 的位置影响。
3. 非独裁：不存在某个选民总能把自己的严格偏好强加为社会偏好。
4. 完备传递：社会排序本身要能比较任意两个选项，并且不能循环。

²¹ 社会福利函数（social welfare function；社会福利函数）：把每个选民的完整排序聚合为一个社会排序的规则。

²² IIA（Independence of Irrelevant Alternatives；无关选项独立性）：社会对 x, y 的排序只应取决于个人对 x, y 的相对排序，不应受第三个选项的位置影响。

定理 4.2 (Arrow 不可能定理)

当备选项至少有三个，且允许所有可能的个人理性排序时，不存在一个社会福利函数同时满足 Pareto、IIA、非独裁和社会排序的完备传递性 [1]。

证明思路

完整证明较长，但核心路线可以分三步看。第一步，定义“决定性联盟”：如果某组选民都把 x 排在 y 前面，并且无论其他人怎样排序，社会都把 x 排在 y 前面，那么这组选民对 x, y 是决定性的。第二步，用 Pareto 保证全体选民是决定性联盟，再用 IIA 和传递性证明：如果一个联盟能决定某一对选项，它的决定力会扩展到其他选项对。第三步，考虑最小决定性联盟。若这个联盟有两个或更多成员，可以构造一个排序剖面，把其中一个人从联盟中分离出来，利用 IIA 和传递性得到更小的决定性联盟，和“最小”矛盾。最后只剩单个选民决定所有选项对，即独裁者。非独裁与此矛盾。□

为了把这个证明路线讲得更直观，可以把“决定性”理解为一种极强的市场势力。如果联盟 S 对 x, y 是决定性的，那么只要 S 内所有人都说 $x \succ y$ ，社会排序就必须给出 $x \succ y$ ，其他人的排序无效。Pareto 告诉我们，全体选民 N 对任意 x, y 都是决定性的，因为若所有人都偏好 x ，社会不能偏好 y 。问题在于，IIA 和传递性会把这个全体决定性不断向更小集合压缩。

例 4.5 (为什么 IIA 会放大大局局部决定力)

考虑三个选项 x, y, z 。假设某个联盟 S 已经能决定 $x \succ y$ 。构造一个偏好剖面：联盟 S 中的选民都排

$$x \succ y \succ z,$$

其他选民都排

$$y \succ z \succ x.$$

由于 S 对 x, y 决定，社会必须有 $x \succ y$ 。同时，所有人都认为 $y \succ z$ ，所以 Pareto 给出 $y \succ z$ 。传递性推出 $x \succ z$ 。现在看 x, z 这一对：在这个剖面中，联盟 S 全部认为 $x \succ z$ ，其他人全部认为 $z \succ x$ ，社会结论是 $x \succ z$ 。IIA 说，社会对 x, z 的判断只取决于每个人在 x, z 之间怎么排，不取决于 y 放在哪里。因此只要保持 S 认为 $x \succ z$ 、其他人认为 $z \succ x$ ，社会都必须给出 $x \succ z$ 。这就把 S 的决定力从 x, y 推到了 x, z 。

这个例子只展示了扩张的一小步。Arrow 证明要做的是把这一步反复使用：从某一对选项扩展到所有选项对，从全体决定性压缩到最小决定性集合。若最小集合有至少两个人，就把它拆成两块，用一个三选项剖面让其中一块在一对比较上变成决定性，从而得到更小决定性集合，矛盾。于是最小决定性集合只能是单人。单人决定所有选项对，就是独裁。

证明思路的重点是 IIA。IIA 看起来很自然：比较 x 和 y 时，为什么要管 z ？但它和传递性放在一起，会让局部的决定力扩散成全局决定力。也正因为如此，很多投票制

度会主动放弃 IIA。Borda 的结果会受其他候选人影响；IRV 的淘汰顺序也会受第三候选人影响。它们不是“不懂”IIA，而是在用牺牲 IIA 换取其他性质。

4.11 Gibbard-Satterthwaite 定理

Gibbard-Satterthwaite 研究的是单赢家社会选择函数²³。规则直接从个人排序选出一个胜者，而不是给出完整社会排序。

定义 4.2 (Strategy-proofness)

一个单赢家规则是 strategy-proof 的，指任一选民都不能通过虚报偏好，使最终结果在自己的真实偏好下变得更好。

定理 4.3 (Gibbard-Satterthwaite 定理)

若备选项至少有三个，社会选择函数是确定性的、onto 的，并且定义在所有可能的个人排序上，那么任何 strategy-proof 的规则都是独裁的 [5, 17]。

证明思路

证明可以借 Arrow 定理。给定一个 strategy-proof 且 onto 的单赢家规则，可以构造一个社会排序：若在只允许 x, y 竞争的情形下规则会选 x ，就把 x 排在 y 前。Strategy-proofness 保证这种成对社会排序不会被某个选民通过改变无关排序操纵，因此得到类似 IIA 的性质；onto 保证每个候选人都有可能被选出；再结合 Pareto 型条件，可以把这个规则嵌入 Arrow 框架。Arrow 定理推出社会排序必须由某个选民决定，对应回单赢家规则，就是该选民的最高偏好总会被选出。因此规则是独裁的。□

命题 4.2 (防操纵规则的单调性)

若一个单赢家规则是 strategy-proof 的，且某个剖面下 x 胜出，那么某位选民把 x 在自己的排序中上移、其他候选人的相对顺序不变， x 不能因此落选。

证明

反设 x 原来胜出。某位选民把 x 上移后，胜者变成 $y \neq x$ 。把上移后的排序看成这位选民的真实偏好。由于 x 被上移， x 在真实偏好中仍排在 y 前面。若他如实报告上移后的排序，结果是 y ；若他虚报原来的排序，结果是 x 。因为 $x \succ y$ ，虚报使他严格受益，违反 strategy-proofness。□

这个小命题展示了 Gibbard-Satterthwaite 证明里的基本压力：防操纵会迫使规则满足许多类似单调性、无关性和路径稳定的性质。可是只要候选人至少三个，且所有偏

²³ 社会选择函数 (social choice function; 社会选择函数)：把选民偏好剖面直接映射为一个或多个胜者的规则。

好排序都可能出现，这些性质很容易把规则逼向“某个人说了算”。

例 4.6 (Borda 的策略排序)

三名候选人 A, B, C ，使用 Borda 分数 $2, 1, 0$ 。有三类选民：

$$3 : A \succ B \succ C, \quad 2 : B \succ C \succ A, \quad 2 : C \succ B \succ A.$$

若所有人真诚排序，分数为

$$S_A = 3 \cdot 2 = 6, \quad S_B = 3 \cdot 1 + 2 \cdot 2 + 2 \cdot 1 = 9, \quad S_C = 2 \cdot 1 + 2 \cdot 2 = 6.$$

B 胜出。现在 A 支持者知道主要对手是 B ，便把 B 排到最后，报成 $A \succ C \succ B$ 。新分数为

$$S_A = 6, \quad S_B = 2 \cdot 2 + 2 \cdot 1 = 6, \quad S_C = 3 \cdot 1 + 2 \cdot 1 + 2 \cdot 2 = 9.$$

赢家变成 C 。如果 A 支持者认为 C 比 B 更接近自己，这种“压低主要对手”的排序就是有利的。Borda 读取完整排序，所以它能奖励广泛可接受性；同一个特点也给了选民调整排序位置的空间。

这个定理解释了为什么第五章的制度都会留下策略空间。FPTP 有弃保，approval 有批准门槛，Borda 容易受排序操纵，IRV 也不能在一般偏好域上完全防操纵。制度设计能改变操纵的形式和难度，但不能在一般情形下彻底消除操纵。

5 选举制度如何读取偏好

多维偏好本身不能直接变成政治结果。制度必须规定选民能提交什么信息，以及这些信息怎样被计算。FPTP 只读取第一偏好；两轮制²⁴ 读取前两名之间的后续选择；IRV²⁵ 读取排序转移；Borda²⁶ 读取排序位置；Condorcet 方法²⁷ 读取成对多数；approval voting²⁸ 读取可接受集合。

下面固定同一组选票，逐个计算这些制度的结果。这样差异不会被选民分布的变化掩盖。规则一变，保留的信息变了，胜者也可能变。

²⁴ **两轮制** (two-round runoff; 两轮决选制)：第一轮按第一偏好选出前两名，第二轮在前两名之间重新聚合多数的规则。

²⁵ **IRV** (Instant-Runoff Voting; 即时决选投票)：选民排序候选人，每轮淘汰当前第一偏好最少者并转移选票的单赢家规则。

²⁶ **Borda** (Borda count; 博尔达计数法)：按排序位置给候选人计分，通常第一名得 $m - 1$ 分、第二名得 $m - 2$ 分的积分规则。

²⁷ **Condorcet 方法** (Condorcet method; 孔多塞方法)：根据候选人之间的成对多数比较寻找胜者的一类规则。

²⁸ **approval voting** (approval voting; 赞成投票)：选民可以批准任意多个候选人，批准票最多者胜出的规则。

5.1 同一组选票下的单赢家制度比较

固定同一组偏好，再更换计票规则，制度差异会看得更清楚。设有 100 名选民和四名候选人 A, B, C, D ，偏好分布如表 12。

表 12: 单赢家制度比较的共同偏好剖面

人数	排序
34	$A \succ B \succ C \succ D$
29	$B \succ C \succ D \succ A$
22	$C \succ B \succ D \succ A$
15	$D \succ B \succ C \succ A$

这组偏好的特点很鲜明： A 的第一偏好最多，但 B 是几乎所有人的可接受候选人。只保留第一偏好的规则会选出 A ；保留转移、排序或成对比较信息的规则会更偏向 B 。

同一组选票，不同单赢家规则保留的信息不同

规则	保留的信息	同一剖面下的结果
FPTP	只读取第一偏好	A
两轮制	读取前二名后的转移票	B
IRV	逐轮淘汰并转移排序票	B
Borda	把完整排序换成积分	B
Condorcet	读取全部成对多数矩阵	B
Approval	读取可接受候选人集合	取决于批准门槛

规则差异不是只体现在赢家姓名上，首先体现在它允许选民报告多少偏好信息。

图 4: 同一组选票经过不同单赢家规则后，保留的信息和胜者不同。

表 13 先把“制度读取了什么”列出来。很多制度差异并不来自价值判断，而是来自信息格式。

选票不是偏好的完整复制，而是一种报告技术。FPTP 的报告空间只有“选一个”；Borda 的报告空间是完整排序；approval 的报告空间是一个集合；Condorcet 方法需要从完整排序中恢复所有成对比较。报告空间越小，投票成本越低，但信息损失越大；报告空间越大，表达更充分，也更容易产生策略排序、认知成本和计票复杂性。

还可以把选票格式看成一种交易成本。FPTP 的交易成本最低：选民只需知道自己最喜欢谁，计票者只需数第一偏好。IRV 和 Borda 要求选民能给候选人排序，这在候选人很多、信息很少时会变得困难；approval 看似简单，但选民要自己决定批准门槛；Condorcet 方法最接近“逐对比较”的多数概念，却需要处理循环和破坏规则。制度设

表 13: 单赢家规则读取的信息

制度	选票中被读取的信息	被弱化或丢掉的信息	典型后果
FPTP	第一偏好	第二偏好、成对多数、可接受性	最大核心群体可能胜出
两轮制	第一偏好和前两名间的转移	未进前二候选人的直接竞争力	鼓励进入决赛与扩联盟
IRV	排序和淘汰路径	不同淘汰路径下的成对比较	奖励可转移性
Borda	所有排序位置	偏好强度和多数门槛	奖励平均排序高
Condorcet	所有成对比较	循环时的破坏原则	奖励多数可接受性
Approval	可接受集合	可接受程度和完整排序	门槛选择会改变结果

计因此不只是在“更多信息”和“更少信息”之间选择，也是在“表达充分性”和“使用可理解性”之间选择。

例 5.1（选票成本也会改变参与）

设一次学生会选举有 8 名候选人。若使用 FPTP，选民只要圈出 1 人；若使用完整排序，选民要给 8 人排序，理论上可能的排序有

$$8! = 40320$$

种。普通选民当然不会逐一比较四万多种排序，但这个数字说明完整排序比单选票要求更多信息。若一部分选民只了解前两三名候选人，强制完整排序可能带来随意填写；若允许只排序前几名，又要决定未排序候选人的处理方式。规则越精细，越需要配套的选民教育和清楚的计票说明。

候选人也会利用这种成本。FPTP 下，竞选信息可以压缩成“我能赢，别浪费票”；approval 下，候选人会努力把自己放进更多选民的可接受集合；IRV 下，候选人会争取“把我排第二也可以”；Borda 下，候选人不只怕被排最后，也怕对手支持者策略性地压低自己。于是，选票格式不仅改变计票结果，还改变竞选语言本身。第七章讨论候选人策略时，这一点会重新出现：制度先规定选民怎样表达，候选人再围绕这种表达方式设计宣传、联盟和退选策略。

5.2 FPTP：第一偏好最多者胜

FPTP 只读取每张选票的第一名。表 12 中的第一偏好票为

$$A = 34, \quad B = 29, \quad C = 22, \quad D = 15.$$

因此 A 当选。这里的关键是，A 并没有得到多数，只得到 34%。FPTP 奖励最大第一偏好集团，而不是奖励多数可接受性。

5.3 两轮制：先表达，再整合

两轮制第一轮也看第一偏好，前两名 A, B 进入第二轮。第二轮只比较 A 和 B 。计算见表 14。

表 14: 两轮制的第二轮转移

选民组	在 A, B 之间选择	票数去向
34: $A \succ B \succ C \succ D$	A	$A + 34$
29: $B \succ C \succ D \succ A$	B	$B + 29$
22: $C \succ B \succ D \succ A$	B	$B + 22$
15: $D \succ B \succ C \succ A$	B	$B + 15$
合计		$A = 34, B = 66$

C 和 D 的支持者在 A, B 之间都选择 B ，所以 B 以 $66 : 34$ 胜出。两轮制的作用不是“多投一次票”这么简单。第一轮保留小群体的表达，第二轮迫使社会在可行的两名候选人之间做多数选择。

5.4 IRV：把淘汰者的选票往后转

IRV 使用完整排序，但不是一次性给分。它每轮淘汰当前第一偏好最少的候选人，再把这些选票转给下一位仍然存活的候选人。RCV²⁹ 常被用来泛指排序投票。

表 15: IRV 淘汰转移过程

轮次	A	B	C	D
第 1 轮	34	29	22	15, 淘汰
第 2 轮	34	44	22, 淘汰	—
第 3 轮	34	66, 胜出	—	—

第一轮 $D = 15$ 最少，被淘汰；这 15 票的下一存活偏好是 B ，所以 B 变成 44。第二轮 $C = 22$ 最少，被淘汰；这 22 票也转给 B 。最终 $B = 66$ ，超过半数。

IRV 和两轮制在这个例子中都选出 B ，但它们不等价。两轮制只允许第一轮前两名进入最终比较；IRV 允许淘汰过程逐步释放低票候选人的后续偏好。

²⁹ **RCV** (Ranked-Choice Voting; 排序选择投票)：泛指让选民给候选人排序的一类投票形式，IRV 是其中一种单赢家算法。

例 5.2（两轮制和 IRV 可以不同）

设有三名候选人 A, B, C ，100 名选民偏好为

$$40 : A \succ C \succ B, \quad 35 : B \succ C \succ A, \quad 25 : C \succ B \succ A.$$

两轮制第一轮前两名是 $A = 40$ 和 $B = 35$ ， C 的 25 票在 A, B 之间转给 B ，所以第二轮 $B = 60, A = 40$ ， B 胜出。

IRV 第一轮淘汰的是 $C = 25$ ，这 25 票也转给 B ，得到 $B = 60, A = 40$ ，仍是 B 。这个例子还没产生差异，因为淘汰者正好是第三名。把偏好改成

$$36 : A \succ C \succ B, \quad 34 : B \succ A \succ C, \quad 30 : C \succ B \succ A.$$

两轮制前两名是 A, B ， C 票转给 B ， $B = 64$ 胜出。IRV 第一轮淘汰 C ，也得到 B 。要让二者真正不同，需要四名或更多候选人，因为 IRV 的逐轮淘汰路径会释放多个候选人的后续偏好，而两轮制在第一轮后直接锁定前两名。设

$$31 : A \succ C \succ B \succ D,$$

$$29 : B \succ C \succ A \succ D,$$

$$25 : C \succ B \succ A \succ D,$$

$$15 : D \succ C \succ B \succ A.$$

两轮制前两名是 A, B 。 C, D 阵营在 A, B 之间都选 B ，所以 $B = 69$ 胜出。IRV 第一轮淘汰 D ，15 票转给 C ，此时

$$A = 31, \quad B = 29, \quad C = 40.$$

第二轮淘汰 B ，29 票转给 C ，于是 $C = 69$ 胜出。两轮制选 B ，IRV 选 C 。差异来自一个简单事实：IRV 允许原本第三名的 C 先吸收 D 的转移票，再超过 B ；两轮制第一轮结束时已经把 C 排除。

这个例子适合课堂展示。它说明“排序投票”等同于“两轮制自动化”的说法过于粗略。IRV 的机制不是直接模拟第二轮，而是把每一轮的最弱候选人删掉。谁先被删、谁吸收转移票，会改变可行赢家集合。

5.5 Borda：把排序位置转成分数

Borda count 对四名候选人按排序给分：第一名 3 分，第二名 2 分，第三名 1 分，第四名 0 分。完整计算见表 16。

B 的第一偏好不是最多，但它从四组人那里都得到分数，且没有被大量选民排在最后。Borda 因此奖励“平均排序高”的候选人。它的弱点也在这里：若新增或退出一个候选人，排序位置和分数都会改变。

表 16: Borda 分数的完整计算

候选人	来自 34 人的分数	来自 29 人的分数	来自 22 人的分数	来自 15 人的分数
<i>A</i>	$34 \cdot 3 = 102$	0	0	0
<i>B</i>	$34 \cdot 2 = 68$	$29 \cdot 3 = 87$	$22 \cdot 2 = 44$	$15 \cdot 2 = 30$
<i>C</i>	$34 \cdot 1 = 34$	$29 \cdot 2 = 58$	$22 \cdot 3 = 66$	$15 \cdot 1 = 15$
<i>D</i>	0	$29 \cdot 1 = 29$	$22 \cdot 1 = 22$	$15 \cdot 3 = 45$
总分	102	229	173	96

5.6 Condorcet: 逐对比较是否有稳定赢家

Condorcet winner³⁰ 的想法是：如果一个候选人在一对一比较中能击败所有人，就选它。在表 12 中，*B* 与其他候选人的比较为

$$B \text{ 对 } A : 29 + 22 + 15 = 66 > 34,$$

$$B \text{ 对 } C : 34 + 29 + 15 = 78 > 22,$$

$$B \text{ 对 } D : 34 + 29 + 22 = 85 > 15.$$

所以 *B* 是 Condorcet winner。这个结果和 FPTP 正好相反：第一偏好最多的是 *A*，成对多数最强的是 *B*。

5.7 Approval: 可接受集合如何改变结果

Approval voting 不要求排序，而是让选民标出所有可接受候选人。设三名候选人 *A, B, C*，偏好分布为

$$35 : A \succ B \succ C, \quad 33 : B \succ C \succ A, \quad 32 : C \succ B \succ A.$$

如果每人只批准第一名，票数是 $A = 35, B = 33, C = 32$ ，*A* 胜出。如果每人批准前两名，票数变成

$$A = 35, \quad B = 35 + 33 + 32 = 100, \quad C = 33 + 32 = 65,$$

B 胜出。

Approval 的计算很简单，难点在选民如何选择批准门槛。门槛一变，制度读取到的信息就变了。

³⁰ **Condorcet winner** (Condorcet winner; 孔多塞赢家)：在一对一多数比较中击败所有其他候选人的候选人。

表 17: Approval 门槛改变胜者

批准规则	A 批准数	B 批准数	C 批准数
只批准第一名	35	33	32
批准前两名	35	100	65

例 5.3 (approval 门槛也是策略选择)

设某位选民真实效用为

$$U(A) = 10, \quad U(B) = 8, \quad U(C) = 1.$$

如果他确信 A 与 B 竞争激烈，而 C 没有胜算，他可能只批准 A ，因为批准 B 会帮助次优候选人击败最优候选人。如果他相信 A 已经无望，而 B 与 C 接近，他会批准 A, B ，用批准 B 阻止最差的 C 。同一组真实效用下，最优批准集合随胜选概率预期改变。可以写成

$$\Delta EU(\text{批准} B) = p_{B>C}[U(B) - U(C)] - p_{B>A}[U(A) - U(B)],$$

其中 $p_{B>C}$ 表示这一批准票帮助 B 击败 C 的关键概率， $p_{B>A}$ 表示它帮助 B 击败 A 的关键概率。当前者收益大于后者成本时，批准 B 合理；反之则不批准。

5.8 不同规则保留了哪些偏好信息

图 4 可以作为本章的总结。FPTP 只读取第一偏好；两轮制读取前两名之间的后续选择；IRV 读取排序中的转移路径；Borda 读取每个候选人的排序位置；Condorcet 读取成对多数矩阵；Approval 读取“可接受集合”。

本章的共同剖面给出一个清楚结论：第一偏好最多者、两两多数最强者、平均排序最高者、可接受集合最大者，可能不是同一个人。选举制度的实际作用，就是在这些候选人概念之间选一个作为胜者定义。下一章转向多席位制度。单赢家制度最终必须把社会偏好压成一个人；比例代表制则把多个群体转化成席位结构。

6 从赢家到席位：比例代表和混合制度

单赢家制度把整个选区压缩成一个胜者。比例代表制换了一个输出格式：它把票数转化为席位。这样，小党不必成为第一名也能进入结果，选民也不必总在“最喜欢”和“最可能赢”之间二选一。

不过，比例代表制不是一个单一算法。D'Hondt³¹、Sainte-Lague³²、最大余数法³³、MMM³⁴ 和 MMP³⁵ 都会改变票数到席位的转换方式。下面用同一组票数把它们算出来。

6.1 从单一赢家到席位向量

第五章的所有规则都必须选出一个赢家。比例代表制不同，它的输出不是 A 或 B ，而是一个席位向量：

$$(s_A, s_B, s_C, s_D).$$

这一步改变了问题。单赢家制度问“谁代表整个选区”；比例代表制问“不同选民群体各得到多少正式代表”。上一章已经看到，同一组偏好可以在不同单赢家规则下选出不同的人；现在换一个输出格式，社会偏好不再被压缩成一个名字，而是变成席位向量。

设四个政党得票为

$$A = 32, \quad B = 28, \quad C = 21, \quad D = 19,$$

总席位为 7。三个常见公式给出的结果如图 5。

6.2 D'Hondt：除数法怎样偏向大党

D'Hondt 使用除数 $1, 2, 3, \dots$ 。每个政党的得票依次除以这些除数，然后取全体最高的 7 个商数。

星号标出的 7 个商数对应席位，得到

$$A = 3, \quad B = 2, \quad C = 1, \quad D = 1.$$

A 的第三席来自商数 $32/3 = 10.67$ 。这个商数仍高于 C 的第二商数 $21/2 = 10.50$ ，所以 A 多拿一席。D'Hondt 的除数增长较慢，已经拿到席位的大党仍可能保留较高商数。

³¹ **D'Hondt** (D'Hondt method; 顿特法)：一种最高平均数除数法，使用 $1, 2, 3, \dots$ 作为除数，通常略有利于大党。

³² **Sainte-Lague** (Sainte-Lague method; 圣拉古法)：一种最高平均数除数法，使用 $1, 3, 5, \dots$ 作为除数，相比 D'Hondt 更有利于中小党。

³³ **最大余数法** (largest remainder method; 最大余数法)：先按配额分配整数席位，再按余数大小分配剩余席位的方法。

³⁴ **MMM** (Mixed-Member Majoritarian; 并立制)：单一选区席位和比例名单席位分别计算后相加的混合制度。

³⁵ **MMP** (Mixed-Member Proportional; 联立制)：先用政党票决定总席位，再用名单席补偿选区席位造成的不比例的混合制度。

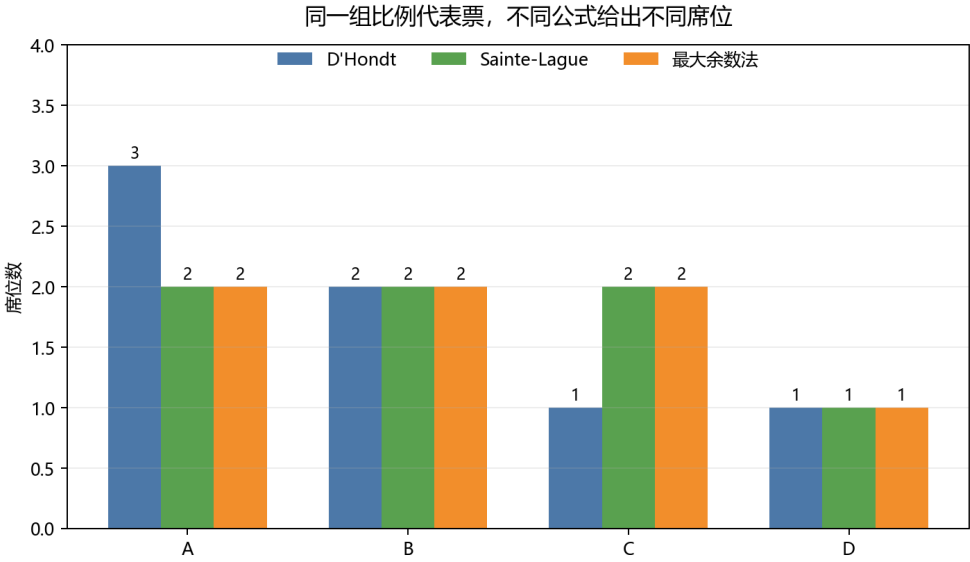


图 5: 同一组选票在不同 PR 公式下得到不同席位。

表 18: D'Hondt 商数表, 7 席

除数	$A = 32$	$B = 28$	$C = 21$	$D = 19$
1	32.00*	28.00*	21.00*	19.00*
2	16.00*	14.00*	10.50	9.50
3	10.67*	9.33	7.00	6.33
4	8.00	7.00	5.25	4.75

把这个过程写成“边际席位价格”会更直观。政党已经拿到 s_j 席时，下一席的竞争商数为

$$q_j^{\text{next}} = \frac{v_j}{s_j + 1}.$$

谁的 q_j^{next} 最大，下一席就给谁。D'Hondt 下，大党 A 得票高，所以即使已经拿到 2 席，下一席商数 $32/3 = 10.67$ 仍然高于 C 的 $21/2 = 10.50$ 。这说明 D'Hondt 的“大党优势”不是一个口号，而是来自边际商数的比较。只要大党的票数足够高，它在已经拿席后仍能用较高边际商数争取下一席。

6.3 Sainte-Lague：让中小党更容易拿到下一席

Sainte-Lague 使用奇数除数 $1, 3, 5, \dots$ 。同样取最高的 7 个商数。

表 19: Sainte-Lague 商数表，7 席

除数	$A = 32$	$B = 28$	$C = 21$	$D = 19$
1	32.00*	28.00*	21.00*	19.00*
3	10.67*	9.33*	7.00*	6.33
5	6.40	5.60	4.20	3.80
7	4.57	4.00	3.00	2.71

结果为

$$A = 2, \quad B = 2, \quad C = 2, \quad D = 1.$$

现在 C 拿到第二席，而 A 没有第三席。差异来自第二个除数：D'Hondt 的第二除数是 2，Sainte-Lague 的第二除数是 3。大党拿完第一席后被更强地“降权”，中小党更容易进入下一轮竞争。

同样用边际商数看，Sainte-Lague 中政党拿到第一席后的下一席商数是 $v_j/3$ ，而不是 $v_j/2$ 。 A 的第二席商数 $32/3 = 10.67$ 仍然很高，但第三席要看 $32/5 = 6.40$ ，已经低于 C 的第二席商数 $21/3 = 7.00$ 。所以 C 拿到第二席。制度差异在这里非常具体：它改变的是“已经拿到席位以后，再拿下一席要被打多大折扣”。

例 6.1（一票变化怎样改变最后一席）

仍用 7 席例子。D'Hondt 下，最后一席在 A 的第三席 $32/3 = 10.67$ 与 C 的第二席 $21/2 = 10.50$ 之间竞争，差距只有 0.17。若 C 的票从 21 增加到 22，则 C 的第二商数变成 $22/2 = 11.00$ ，高于 A 的第三商数，席位从

$$(3, 2, 1, 1)$$

变为

$$(2, 2, 2, 1).$$

这说明 PR 虽然比 FPTP 连续，但也不是完全平滑。最后一席附近仍然存在跳跃，政党会围绕这些跳跃集中资源。

6.4 最大余数法：先按配额取整，再比分数尾巴

最大余数法先计算配额。总票数为 100，总席位为 7，Hare quota³⁶ 为

$$q = \frac{100}{7} = 14.286.$$

每党先拿整数席，再按余数大小分配剩余席位。

表 20: 最大余数法的配额与余数

政党	得票	得票除以配额	初始整数席	余数票
<i>A</i>	32	2.240	2	3.43
<i>B</i>	28	1.960	1	13.71*
<i>C</i>	21	1.470	1	6.71*
<i>D</i>	19	1.330	1	4.71

初始席位合计 5，还剩 2 席。余数最大的两个政党是 *B* 和 *C*，所以最终为

$$A = 2, \quad B = 2, \quad C = 2, \quad D = 1.$$

最大余数法的直觉最容易讲：先问每党能装下几个完整配额，再看谁离下一席最近。它也会带来配额悖论等问题，席位分配理论中 Balinski 和 Young 对这些问题有系统讨论 [2]。

最大余数法和除数法的差别在于，前者先固定一个“每席需要多少票”的配额，后者让每一席在商数排名中竞争。最大余数法更贴近比例直觉：*B* 的 28 票几乎是两个配额，*C* 的 21 票也明显超过一个配额并接近第二个配额。可是配额法的尾数竞争会带来另一些悖论。例如总席位增加时，某个政党反而可能少一席，这类现象被称为 apportionment paradox³⁷。这里不用展开证明，但要知道，席位分配公式本身也是一套机制设计。

6.5 MMM 与 MMP：混合制度的差别在哪里

混合制度容易被一句“有选区票也有政党票”带过。区分 MMM 与 MMP 的关键，是名单席位是否补偿选区席位造成的不比例。MMM 是并立：选区席和名单席相加。MMP 是联立：政党票先决定总席位，名单席用于补偿。

³⁶ **Hare quota** (Hare quota; 黑尔配额)：总票数除以总席位数得到的每席票数配额。

³⁷ **apportionment paradox** (apportionment paradox; 席位分配悖论)：席位总数、配额或人口变化后，某个主体的席位反而朝反直觉方向变化的现象。

例 6.2 (同票下的 MMM 与 MMP)

总席位 10，其中 6 个单一选区席位、4 个名单席位。政党票为

$$A = 45\%, \quad B = 35\%, \quad C = 20\%.$$

单一选区结果为 $A = 5, B = 1, C = 0$ 。

MMM 中，4 个名单席位另按 D'Hondt 分配。最高商数为

$$45(A), \quad 35(B), \quad 22.5(A), \quad 20(C),$$

名单席为 $A = 2, B = 1, C = 1$ ，总席位为

$$A = 7, \quad B = 2, \quad C = 1.$$

MMP 中，先用政党票分配 10 个总席位，得到

$$A = 5, \quad B = 3, \quad C = 2.$$

再扣除选区席，名单补偿为 $A = 0, B = 2, C = 2$ 。最终席位就是 $A = 5, B = 3, C = 2$ 。

表 21: MMM 与 MMP 的同票差异

制度	选区席 A, B, C	名单或补偿席 A, B, C	总席位 A, B, C	制度含义
MMM	5, 1, 0	2, 1, 1	7, 2, 1	选区优势被保留
MMP	5, 1, 0	0, 2, 2	5, 3, 2	名单席补偿比例偏差

6.6 门槛例子：为什么 4.9% 和 5.1% 差很多

法定门槛³⁸ 把一个本来较平滑的席位函数重新切出一个断点。设总席位 100，法定门槛为 5%。小党 S 的席位收益近似为

$$s_S(v_S) = \begin{cases} 0, & v_S < 5\%, \\ 100v_S, & v_S \geq 5\%. \end{cases}$$

若 $v_S = 4.9\%$ ，席位为 0；若 $v_S = 5.1\%$ ，近似得到 5 席。只差 0.2 个百分点，席位从 0 跳到 5。

这就解释了门槛附近的小党策略。它可能降低某些议题的锋芒，争取边际选民；也可能和另一个 2% 或 3% 的小党组成联合名单。策略的目标不是抽象地“变温和”，而是跨过席位函数的断点。

³⁸ 法定门槛 (electoral threshold; 法定门槛): 政党必须达到的最低得票率或票数要求，低于门槛通常不能获得比例席位。

例 6.3（门槛如何改变联盟谈判）

设议会 100 席，右侧两个小党 S_1, S_2 的预期得票分别是 3.0% 和 2.4%，法定门槛为 5%。若分别参选，二者都低于门槛，席位为 0。若组成联合名单，合计得票为 5.4%，可得到约 5 席。设每席价值为 r ，联合名单需要支付协调成本 K 。分别参选的总收益为 0，联合参选收益为

$$5r - K.$$

只要 $5r > K$ ，联合名单就是理性的。这个计算解释了为什么门槛制度下经常出现小党联盟、选前协议、共同名单和“保门槛票”的动员话语。制度没有直接命令小党合并，但它把分散参选的收益压到零，把联合参选变成可能的最优反应。

7 候选人和政党如何适应制度

同一组偏好，在第五章已经产生了不同赢家。现在换一个角度：候选人和政党如果事先知道规则，会怎样调整策略？

继续使用同一个政治市场。四个政治阵营的第一偏好分别是

$$A = 34, \quad B = 29, \quad C = 22, \quad D = 15.$$

排序仍为

34	$A \succ B \succ C \succ D$
29	$B \succ C \succ D \succ A$
22	$C \succ B \succ D \succ A$
15	$D \succ B \succ C \succ A$

A 拥有最大第一偏好集团， B 拥有最强的第二偏好和成对多数优势， C, D 是较小但稳定的侧翼群体。这个例子适合贯穿本章，因为它能同时展示三种政治位置：最大核心盘、广泛可接受候选人、小党或侧翼候选人。

图 6 把这个例子放进不同规则。制度先改变收益函数，收益函数再改变候选人的好反应。下面逐项算。

表 22: 同一政治市场下的策略问题

制度	候选人最关心的变量	小党/侧翼的处境	典型策略
FPTP	是否成为第一名	有票但可能无席	弃保、退选、背书、压缩竞争者
两轮制	能否进前二、能否吸收转移票	第一轮可表达，第二轮要选择阵营	温和化、互不攻击、争取背书
IRV	淘汰顺序和转移路径	不一定要进前二，但不能过早被淘汰	提高可转移性，避免成为最后偏好
PR	得票率、门槛、联盟地位	得票可换席位	维持品牌、跨门槛、联盟议价
MMP	地方席与政党票的组合	地方与全国策略可分开	地方温和化，全国保持标签



图 6: 同一政治市场在不同制度下诱导不同策略。

7.1 FPTP：最大化成为第一名的概率

FPTP 下，候选人 j 的席位收益近似为

$$\Pi_j = R \cdot \mathbf{1}\{v_j > \max_{k \neq j} v_k\} - F_j,$$

其中 R 是赢得席位的收益， F_j 是参选成本。这个收益函数有一个尖锐断点：第二名和最后一名在席位上都等于 0。

在共同例子中，真诚投票给出

$$A = 34, \quad B = 29, \quad C = 22, \quad D = 15.$$

A 胜出。此时各方的策略不同。

表 23: FPTP 下各候选人的策略压力			
候选人	当前第一偏好	直接结果	主要策略
A	34	第一名	保持最大核心盘，防止 B 吸收 C, D
B	29	第二名	争取 C, D 退选或转投，形成反 A 协调
C	22	第三名	若无胜算，考虑和 B 协调或换取政策承诺
D	15	第四名	单独参选收益低，除非有长期品牌收益

B 的最优化问题很清楚。只要它能从 C, D 阵营中吸收 6 个百分点，票数就变成

$$B = 29 + 6 = 35 > A = 34.$$

所以 B 不一定要说服所有 C, D 选民，只需要找到足够多的边际转移票。 C, D 的问题则相反：如果参选只能让 A 获胜，它们可能用退选或背书换取 B 的政策让步。这就是弃保和退选协调的收益基础。

例 7.1（小党退选的收益比较）

把 C 看成一个小党。若 C 单独参选，FPTP 结果是 A 当选， C 无席。设 C 的长期品牌收益为 b_C ，参选成本为 F_C ，则

$$\Pi_C^{\text{enter}} = b_C - F_C.$$

若 C 退选并背书 B ，假设 C 的 22 票中有 18 票转给 B ，则

$$B = 29 + 18 = 47 > A = 34.$$

B 当选。设 B 给 C 的政策让步价值为 k ，但退选损失一部分品牌曝光 b_C ，则

$$\Pi_C^{\text{withdraw}} = k - \delta b_C,$$

其中 $\delta \in [0, 1]$ 表示品牌损失比例。退选条件为

$$k - \delta b_C > b_C - F_C,$$

即

$$k > (1 + \delta)b_C - F_C.$$

这条不等式解释了为什么小党不一定总退选。若长期品牌很重要，或者大党给的政策让步很少，小党会继续参选；若大党让步足够大，且选民高度愿意转移，退选或背书就成为理性策略。

例 7.2（ B 是否应该向 C, D 靠拢）

设 B 有两种平台。温和平台 b_M 保住原有 29 票，并吸收 C, D 中 8 票，得到

$$v_B(b_M) = 29 + 8 = 37.$$

更强烈的中间偏左平台 b_L 可以吸收 C, D 中 14 票，但会失去原有支持者中 7 票，得到

$$v_B(b_L) = 29 - 7 + 14 = 36.$$

若只看票数， b_M 更优，因为 $37 > 36$ 。这说明“向侧翼靠拢”不是越多越好。候选人在不同群体之间移动时，要比较边际获得票和边际流失票。FPTP 下的最优平台常常位于某个能刚好超过第一名的地方，而不一定等于政策空间的几何中心。

7.2 策略投票：选民也在解最优化问题

选民同样会看可胜性。设选民 3 的偏好为 $C \succ B \succ A$ 。如果他认为 C 没有胜算，而 A, B 接近，那么投给 C 的结果效用可能低于投给 B 。可以写成一个简单比较：

$$EU(\text{投}C) = p_C U(C) + p_B U(B) + p_A U(A),$$

$$EU(\text{投}B) = p'_C U(C) + p'_B U(B) + p'_A U(A).$$

若改投 B 明显提高 B 击败 A 的概率，且 $U(B) > U(A)$ ，策略投票就是理性的。

表 24: 策略投票的最小例子

选民 3 行动	A 票	B 票	C 票	胜者
真诚投 C	1	1	1	A (破平)
策略投 B	1	2	0	B

表 24 中，选民 3 真诚投 C 会让最差的 A 胜出；改投第二偏好的 B ，结果从 A 变成 B 。这不是选民“不真诚”，而是 FPTP 的收益断点把表达问题变成协调问题。

7.3 两轮制：第一轮进决赛，第二轮扩联盟

两轮制候选人的收益可以写为

$$\Pi_j = R \cdot \Pr(j \in T_2) \cdot \Pr(j \text{ wins runoff} \mid j \in T_2) - F_j,$$

其中 T_2 是第一轮前两名集合。这个式子把目标拆成两步：先进前二，再赢第二轮。

共同例子中，第一轮前两名是 A, B 。第二轮 C, D 阵营都在 A, B 之间选择 B ，于是

$$B = 29 + 22 + 15 = 66, \quad A = 34.$$

B 胜出。对 B 来说，第一轮不必成为第一名，只要进前二并保持对 C, D 选民的可接受性即可。对 A 来说，策略重点是阻止 B 成为所有非 A 阵营的共同第二选择。

表 25: 两轮制下的策略目标

候选人	第一轮目标	第二轮目标
A	保持第一名	分裂 C, D 转移票，避免反 A 多数
B	稳进前二	维持 C, D 的第二偏好
C, D	争取进入前二或换取背书价值	决定向谁转移支持

两轮制中，攻击侧翼候选人可能有短期收益，也可能损害第二轮转移。若 B 为了抢第一轮票猛烈攻击 C ，导致 C 的 22 票只有一半转给 B ，则第二轮变为

$$B = 29 + 11 + 15 = 55, \quad A = 34,$$

B 仍胜；但如果 D 也不转移， $B = 40$ ，优势就很薄。候选人要算的是转移票弹性。

可以把这个问题写成一个阈值。设 C, D 阵营总票为 37，其中比例 t 会在第二轮转给 B 。 B 的第二轮票为

$$v_B^{(2)} = 29 + 37t.$$

要击败 $A = 34$ ，需要

$$29 + 37t > 34,$$

即

$$t > \frac{5}{37} \approx 0.135.$$

在共同例子中， B 只需要很小比例的转移票就能超过 A ，所以 B 可以承受一定攻击成本。但如果第一轮变成 $A = 45, B = 30, C + D = 25$ ，阈值就是

$$30 + 25t > 45 \Rightarrow t > 0.6.$$

这时 B 必须非常重视第二轮可接受性。两轮制的策略强度取决于第一轮差距和被淘汰阵营规模。

7.4 IRV：不要只看第一名，还要避免被排斥

IRV 下，候选人的目标不是简单的第一轮票数，而是在淘汰路径中存活：

$$\Pi_j = R \cdot \Pr(j \text{ survives transfers and reaches majority}) - F_j.$$

共同例子中， D 先被淘汰，15 票转给 B ；随后 C 被淘汰，22 票也转给 B 。 B 的好位置来自两点：它有 29 的第一偏好，足够不被早早淘汰；它又是 C, D 的后续偏好，可以接收转移票。

如果把 B 改成一个强烈冒犯 C, D 阵营的候选人，使 C, D 的后续偏好不再流向 B ，IRV 结果就可能改变。IRV 奖励“可转移性”：你可以不是最多人的第一选择，但不能成为太多人的最后选择。

例 7.3（IRV 下的早淘汰风险）

设四名候选人的第一偏好变为

$$A = 32, \quad B = 24, \quad C = 23, \quad D = 21.$$

B 仍然是很多人的第二偏好，但第一轮只比 C, D 略高。如果 B 因为某次丑闻或策略失误失去 2 票，变成

$$A = 32, \quad B = 22, \quad C = 23, \quad D = 23,$$

B 第一轮就被淘汰。它的“广泛可接受性”还没来得及发挥作用。IRV 因此给中间候选人一个双重要求：既要有足够第一偏好避免早淘汰，又要有足够后续偏好接收转移票。这个机制和两轮制不同；两轮制只要求进入前二，IRV 则要求在每一轮都不成为最低者。

7.5 PR：小党不必赢第一名也能有收益

PR 下政党的收益函数更接近

$$\Pi_j = R s_j + \theta_j P_j - F_j,$$

其中 s_j 是席位数, P_j 是进入联盟或影响政策的概率, θ_j 是政策影响价值。

把共同例子的第一偏好当作政党票, 并分配 10 个席位。D'Hondt 结果大致为

$$A = 4, \quad B = 3, \quad C = 2, \quad D = 1.$$

这改变了 C, D 的策略。FPTP 下它们可能被迫退选; PR 下它们可以保留独立品牌, 因为 22% 和 15% 都能换成席位。

表 26: FPTP 与 PR 下小党进入收益的差别

候选人/政党	第一偏好	FPTP 席位	10 席 PR 近似席位
C	22	0	2
D	15	0	1

所以 PR 的策略不是“争第一名”, 而是“跨门槛、稳住细分群体、进入联盟谈判”。小党甚至可能不想过度中间化, 因为中间化会模糊品牌, 降低核心群体投票动机。

例 7.4 (PR 下小党为什么可能保持差异化)

设 C 有两种策略。保持清晰议题品牌时, 得票为 12%, 联盟谈判中能获得价值 $h = 3$ 的政策影响; 温和化后, 得票升到 15%, 但政策品牌变弱, 联盟影响降为 $h = 1$ 。若每 1% 得票大约带来 0.1 的席位价值, 则

$$\Pi_C^{\text{brand}} = 0.1 \times 12 + 3 = 4.2,$$

$$\Pi_C^{\text{moderate}} = 0.1 \times 15 + 1 = 2.5.$$

即使温和化提高了得票, 保持差异化仍可能更优。PR 的政治产品逻辑更接近产品差异化: 小党靠清晰定位占据一个细分市场, 再在联盟谈判中把席位转化为政策影响。

7.6 联盟议价: 席位小不等于影响小

在 PR 中, 席位不是最终目标, 政策和内阁还要通过 联盟议价³⁹ 形成。设议会席位为

$$A = 46, \quad B = 42, \quad C = 12.$$

任何多数政府都需要超过 50 席。 $A + B = 88$ 可以组成大联合, 但两大党政策差距和职位竞争都很高; $A + C = 58$, $B + C = 54$, 小党 C 可以分别帮助两边形成多数。若 C 与 A 组阁得到政策收益 $u_C(A)$ 和职位收益 $o_C(A)$, 与 B 组阁得到 $u_C(B)$ 和 $o_C(B)$, 它选择

$$\max\{u_C(A) + o_C(A), u_C(B) + o_C(B)\}.$$

³⁹ 联盟议价 (coalition bargaining; 联盟议价): 没有单一政党掌握多数时, 政党围绕组阁、职位和政策让步进行谈判的过程。

假设 $u_C(A) = 4, o_C(A) = 2$, 而 $u_C(B) = 3, o_C(B) = 4$, 则

$$A + C : 6, \quad B + C : 7.$$

C 会选择 B 。如果 A 想争取 C , 需要把职位或政策让步提高至少 1。这个计算说明, 小党影响力来自“枢纽位置”: 没有它就没有多数。FPTP 下 12% 的第三党可能没有席位; PR 下 12 席可能成为组阁关键。

7.7 MMP: 地方候选人与全国品牌分开优化

MMP 下, 政党同时面对两个市场: 地方选区和全国政党票。地方候选人要服务本地中位选民, 全国名单要维护政党整体品牌。一个党可以在地方推出温和候选人, 同时在全国名单上保持清晰政策标签。

这与 MMM 不同。MMM 中, 选区多赢不会减少名单席收益; MMP 中, 选区席会先占用总席位, 名单席用于补偿。因此大党在 MMP 下需要同时考虑两个变量:

$$\Pi_j = R_D d_j + R_L (\bar{s}_j - d_j)^+ - F_j,$$

其中 d_j 是选区席, $\bar{s}_j$ 是政党票决定的总应得席位, $(\bar{s}_j - d_j)^+$ 是可补偿名单席。若 d_j 很高, 名单补偿就会下降。

例 7.5 (MMP 下地方温和化和全国品牌并存)

设政党 A 的全国理想政策为 0.30, 某地方选区中位选民为 0.48。若地方候选人坚持 0.30, 选区胜率为 40%; 若地方候选人移动到 0.45, 选区胜率升到 60%, 但会让全国核心支持者的不满成本为 0.05。设一个选区席价值为 1, 则地方温和化的净收益为

$$0.60 - 0.40 - 0.05 = 0.15 > 0.$$

所以地方候选人会温和化。与此同时, 全国名单仍可维持 0.30 的政党标签, 因为政党票面对的是全国支持者和比例席位。MMP 把“地方候选人”和“全国政党品牌”分成两个可优化变量, 这一点和纯 FPTP 或纯 PR 都不同。

7.8 策略比较的结论

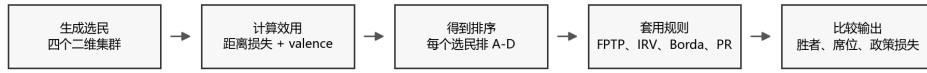
同一组偏好, 在不同制度下给候选人的优化问题不同。FPTP 让候选人争第一名并制造弃保压力; 两轮制让候选人同时争基本盘和第二轮转移; IRV 奖励可转移性; PR 鼓励细分品牌和联盟谈判; MMP 把地方服务和全国品牌放到同一个收益函数里。

所以制度影响的不只是“最后谁赢”。它会提前改变谁参选、谁退选、谁结盟、谁中间化、谁保持差异化。

8 小型仿真：从选民坐标到制度结果

仿真把前面的纸面例子放进一个二维政策空间。选民有坐标，候选人也有坐标；效用由政策距离和候选人质量决定；排序由效用推出；制度规则再读取这些排序。

仿真计算链条：从二维偏好到制度结果



仿真不是经验实证；它把前文的效用函数、排序规则和席位公式放到同一组数据里手算核对。

图 7: 仿真的计算链：坐标生成效用，效用生成排序，排序进入制度算法。

候选人设定为

$$A = (0.05, 0.90, q = 0), \quad B = (0.00, 0.25, q = 0.08),$$

$$C = (0.75, -0.15, q = 0), \quad D = (-0.75, -0.15, q = 0).$$

选民由四个集群生成：340 人靠近 A ，290 人靠近 B ，220 人靠近 C ，150 人靠近 D 。每个选民对候选人 j 的效用为

$$U_i(j) = -[(x_i - E_j)^2 + (y_i - C_j)^2] + q_j.$$

效用越高，排序越靠前。

8.1 先算一个选民：排序从效用来

仿真中的每个选民 i 有一个二维理想点 (x_i, y_i) 。候选人 j 的位置为 (E_j, C_j) ，质量项为 q_j 。效用写成

$$U_{ij} = -[(x_i - E_j)^2 + (y_i - C_j)^2] + q_j.$$

脚本并不直接指定“谁喜欢谁”，而是先计算四个效用，再按效用从高到低生成排序。

例 8.1（一个选民怎样生成排序）

设某位选民的理想点为 $(0.10, 0.20)$ 。候选人位置和质量项为

$$A = (0.05, 0.90, q_A = 0), \quad B = (0.00, 0.25, q_B = 0.08),$$

$$C = (0.75, -0.15, q_C = 0), \quad D = (-0.75, -0.15, q_D = 0).$$

则

$$U_A = -[(0.10 - 0.05)^2 + (0.20 - 0.90)^2] + 0 = -0.4925,$$

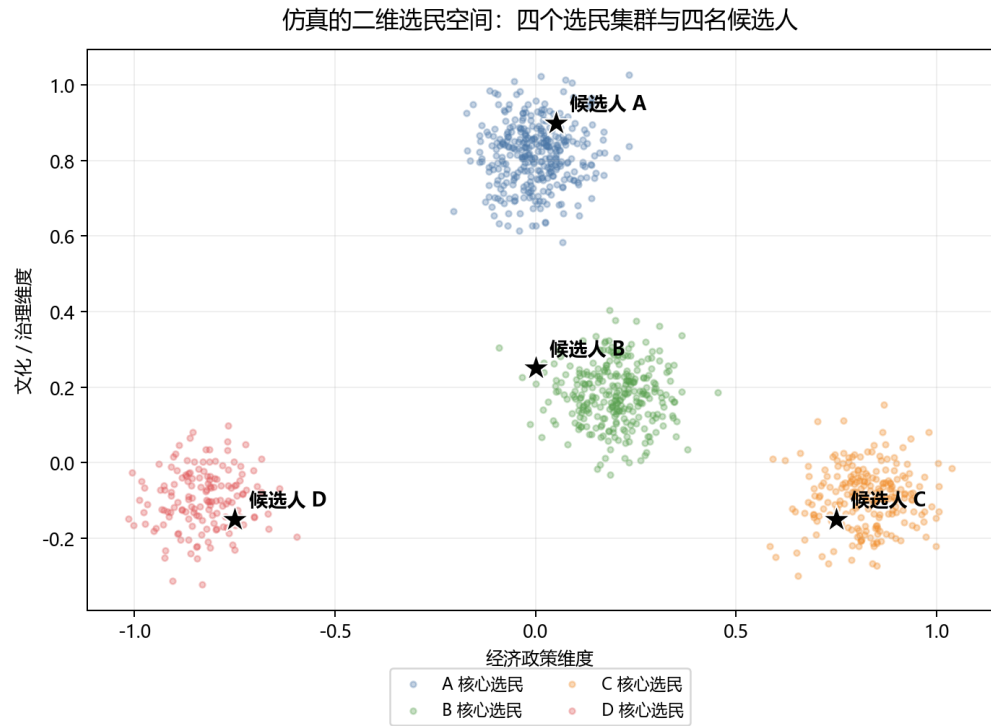


图 8: 仿真的二维选民空间与候选人位置。

$$U_B = -[(0.10 - 0.00)^2 + (0.20 - 0.25)^2] + 0.08 = 0.0675,$$
$$U_C = -[(0.10 - 0.75)^2 + (0.20 + 0.15)^2] = -0.5450,$$
$$U_D = -[(0.10 + 0.75)^2 + (0.20 + 0.15)^2] = -0.8450.$$

排序为

$$B \succ A \succ C \succ D.$$

这个单个选民的计算展示了 B 的 valence 如何进入排序。若没有 $q_B = 0.08$, B 的效用仍为 -0.0125 , 在这里仍会胜过 A , 但优势变小; 对边界选民来说, 质量项可能直接改变第一偏好。

表 27: 仿真算法的输入与输出

步骤	输入	运算	输出
生成选民	四个选民集群、随机扰动	得到 1000 个理想点	选民坐标
计算效用	候选人坐标和 q_j	$U_{ij} = -d_{ij}^2 + q_j$	每个选民对四人的效用
生成排序	每个选民的四个效用	从高到低排序	完整偏好剖面
输入制度	完整剖面或第一偏好	套用 FPTP、IRV、Borda、PR 等算法	胜者或席位
评价结果	胜者、席位、选民坐标	计算成对多数矩阵和政策距离损失	制度比较表

这里先记两个输出量: pairwise matrix⁴⁰ 用来检查谁能在两两比较中击败所有对手;

⁴⁰ **pairwise matrix** (pairwise majority matrix; 成对多数矩阵): 记录每个候选人在一对一比较中能获

policy loss⁴¹ 用来衡量单一胜者离选民总体分布有多远。

8.2 第一步：从效用得到第一偏好

脚本对每个选民计算四个效用值，然后把效用最高的候选人记为第一偏好。运行结果为

$$A = 333, \quad B = 297, \quad C = 220, \quad D = 150.$$

这一步解释了为什么 FPTP 会选出 A 。 A -core 是最大的单一集群，虽然 A 不是全体选民的空间中心，但它拥有最多第一偏好。

表 28: 仿真的第一偏好票

候选人	A	B	C	D
第一偏好票	333	297	220	150

从第一偏好票就能看出制度冲突的来源。 A 的 333 票只比 B 多 36 票，但 FPTP 把这个小差距放大成“唯一赢家”。 C, D 合计 370 票，高于 A ，但它们分散在两个候选人上。若这些选民大多把 B 排在后续偏好，排序制度就会把分散票重新整合；FPTP 不会。

8.3 第二步：两轮制和 IRV 怎样把票转给 B

两轮制先取第一偏好前两名： A 和 B 。第二轮只比较 A, B 。所有在 C 或 D 阵营中更偏好 B 而非 A 的选民都会转向 B 。脚本结果为

$$A = 333, \quad B = 667.$$

所以两轮制选出 B 。

IRV 的过程更细。第 1 轮， $D = 150$ 最少，被淘汰；这些票转给下一存活偏好，主要进入 B 。第 2 轮， $C = 220$ 最少，被淘汰；这些票也转给 B 。结果如表 29。

这里可以看到第七章的策略含义。 B 不只是有 297 张第一偏好票；它还是 C, D 阵营愿意转向的候选人。IRV 读取了这部分后续偏好。

得多少票的矩阵。

⁴¹ **policy loss** (policy-distance loss; 政策距离损失): 用胜者政策位置到所有选民理想点的平均距离衡量代表偏离程度的指标。

表 29: 仿真的 IRV 淘汰过程

轮次	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
第 1 轮	333	297	220	150, 淘汰
第 2 轮	333	447	220, 淘汰	–
第 3 轮	333	667, 胜出	–	–

8.4 第三步：Borda 和 Condorcet 为什么也选 B

Borda 把排序位置转成分数。仿真输出为

$$S_A = 1752, \quad S_B = 2297, \quad S_C = 1320, \quad S_D = 631.$$

B 的分数最高。原因仍然是广泛可接受性：*B* 在很多选民排序中不是第一名，但经常排在第二名。

可以用平均排名理解这个分数。四名候选人时，Borda 分数越高，平均排名越靠前。若某候选人被 300 人排第一、700 人排最后，它的第一偏好很强，但 Borda 分数会很低。*B* 的优势来自“很少被排到最后”。这和第七章的策略含义一致：在读取完整排序的制度下，候选人要避免成为大量选民的不可接受对象。

成对多数矩阵更直接。行候选人与列候选人相比，单元格表示行候选人得到的票数。

表 30: 仿真的成对多数矩阵

行对列	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
<i>A</i>	–	333	572	847
<i>B</i>	667	–	780	850
<i>C</i>	428	220	–	672
<i>D</i>	153	150	328	–

从表 30 看，*B* 对 *A, C, D* 都超过 500 票。因此 *B* 是 Condorcet winner。FPTP 选 *A*，不是因为 *A* 在多数比较中更强，而是因为 FPTP 没有读取成对比较。

8.5 第四步：PR 把集群保留为席位

如果把第一偏好视为政党票，并用 10 席 PR 分配，D’Hondt 得到

$$A = 4, \quad B = 3, \quad C = 2, \quad D = 1.$$

Sainte-Lague 得到

$$A = 3, \quad B = 3, \quad C = 2, \quad D = 2.$$

两者都没有把选民压缩成一个赢家。差别在于，小党 D 在 D'Hondt 下得到 1 席，在 Sainte-Lague 下得到 2 席。这里又回到第六章：席位公式会改变中小党收益。

若把 PR 输出解释为“社会偏好的离散近似”，D'Hondt 的 4, 3, 2, 1 更有利于较大集团，Sainte-Lague 的 3, 3, 2, 2 更保护较小集团。二者都比 FPTP 保留更多群体信息，但内部仍有分配差异。仿真因此不是只比较“单赢家 vs PR”，还比较了 PR 公式内部的边际席位规则。

8.6 第五步：政策距离损失

最后计算每个制度胜者与所有选民的平均二次距离，即 policy loss：

$$\text{Loss}(j) = \frac{1}{N} \sum_{i=1}^N [(x_i - E_j)^2 + (y_i - C_j)^2].$$

脚本输出为

$$\text{Loss}(A) = 0.7854, \quad \text{Loss}(B) = 0.4300.$$

所以 FPTP 胜者 A 的平均政策距离更高；两轮制、IRV、Borda 和 Condorcet 选出的 B 平均距离更低。

这个结果不需要夸大。它只说明，在这个二维分布下， A 是最大局部集群的代表， B 更接近整体折中。若评价目标是最大第一偏好， A 有制度优势；若评价目标是多数可接受性或平均政策距离， B 更合适；若评价目标是保留多个集群，PR 的席位向量更合适。

表 31: 仿真结果如何对应前文理论

仿真现象	对应章节	解释
FPTP 选 A	第 5 章 FPTP	第一偏好最大集团胜出
Runoff/IRV 选 B	第 5 章转移票、第 7 章策略	后续偏好把 C, D 支持者转向 B
Borda 选 B	第 5 章 Borda	平均排序较高
Condorcet 选 B	第 4 章多数比较	B 击败每个对手
PR 给 C, D 席位	第 6 章席位公式	小群体没有被压成零
B policy loss 更低	第 3 章空间效用	B 更接近整体二维分布中心

仿真把前面分散的手算例子放到同一个二维场景里。机制链条很清楚：选民位置生成效用，效用生成排序，制度读取排序的不同部分，最后得到不同胜者、席位和政策损失。

9 制度设计权衡

前面的模型和仿真已经给出同一组输入：

$$A = 333, \quad B = 297, \quad C = 220, \quad D = 150.$$

不同制度给出的结果不同。FPTP 选 A ；两轮制、IRV、Borda 和 Condorcet 选 B ；PR 输出一个席位向量。制度设计要比较的，是这些结果分别牺牲了什么、保留了什么。

三个指标从前文算例自然得到。

第一，政策距离损失：

$$\text{Loss}(j) = \frac{1}{N} \sum_i d(x_i, y_j)^2.$$

它衡量单一胜者离选民理想点有多远。

第二，票席失真。这里使用 Gallagher index⁴² 的思想 [4]：

$$G = \sqrt{\frac{1}{2} \sum_j (s_j - v_j)^2},$$

其中 v_j 是得票率， s_j 是席位率。它衡量选票比例和席位比例的偏离。

第三，有效政党数⁴³ [8]：

$$N = \frac{1}{\sum_j s_j^2}.$$

它衡量席位结构是集中在一个党，还是分散在多个党。

9.1 把指标代入同一个例子

先看 FPTP。它把 1000 张选票压缩成一个赢家 A ，等价于席位率

$$s_A = 100\%, \quad s_B = s_C = s_D = 0.$$

得票率为

$$v_A = 33.3\%, \quad v_B = 29.7\%, \quad v_C = 22.0\%, \quad v_D = 15.0\%.$$

票席失真约为

$$G_{\text{FPTP}} = \sqrt{\frac{1}{2} [(100 - 33.3)^2 + (0 - 29.7)^2 + (0 - 22)^2 + (0 - 15)^2]} \approx 54.95.$$

⁴² **Gallagher index** (Gallagher index / least squares index; 盖拉格指数 / 最小二乘失真指数)：衡量政党得票率与席位率偏离程度的比例性指标，数值越大表示票席失真越强。

⁴³ **有效政党数** (effective number of parties; 有效政党数)：按席位或得票份额加权计算的政党数量指标，用来刻画政党系统碎片化程度。

有效政党数为

$$N_{\text{FPTP}} = \frac{1}{1^2} = 1.$$

政策距离损失为

$$\text{Loss}(A) = 0.7854.$$

再看 D'Hondt PR。席位为

$$A = 4, \quad B = 3, \quad C = 2, \quad D = 1,$$

席位率为 40%, 30%, 20%, 10%。票席失真为

$$G_{\text{PR}} = \sqrt{\frac{1}{2}[(40 - 33.3)^2 + (30 - 29.7)^2 + (20 - 22)^2 + (10 - 15)^2]} \approx 6.08.$$

有效政党数为

$$N_{\text{PR}} = \frac{1}{0.4^2 + 0.3^2 + 0.2^2 + 0.1^2} \approx 3.33.$$

PR 没有单一政策赢家，因此不能直接给一个 $\text{Loss}(j)$ 。政策会在议会和联盟中继续形成。

Sainte-Lague PR 的席位为

$$A = 3, \quad B = 3, \quad C = 2, \quad D = 2,$$

席位率为 30%, 30%, 20%, 20%。票席失真为

$$G_{\text{SL}} = \sqrt{\frac{1}{2}[(-3.3)^2 + (0.3)^2 + (-2)^2 + (5)^2]} \approx 4.47.$$

有效政党数为

$$N_{\text{SL}} = \frac{1}{0.3^2 + 0.3^2 + 0.2^2 + 0.2^2} \approx 3.85.$$

和 D'Hondt 相比，Sainte-Lague 给 D 多一席，降低了票席失真，也提高了有效政党数。这个变化来自第六章讲过的奇数除数：中小党更容易拿到下一席。

表 32: 同一组偏好下的制度指标比较

制度	输出	政策距离损失	票席失真 G	有效政党数 N
FPTP	A 胜	0.7854	54.95	1.00
Runoff / IRV / Borda / Condorcet	B 胜	0.4300	—	1.00
D'Hondt PR	4, 3, 2, 1	—	6.08	3.33
Sainte-Lague PR	3, 3, 2, 2	—	4.47	3.85

表 32 显示三个目标不能同时最优。FPTP 简单，输出清楚，但票席失真大，且在仿真中选出政策距离损失较高的 A 。排序和两轮制度仍然输出单一胜者，但能利用后续偏好，在这个例子中选出平均距离更低的 B 。PR 保留多个群体的代表性，票席失真小，有效政党数高，但政策结果要经过议会谈判。

9.2 把制度评价写成目标函数

制度选择可以写成一个加权目标函数。设评价者关心四件事：政策距离损失 L 、票席失真 G 、政党碎片化 N 带来的协调成本、以及问责清晰度 A_c 。一个简单形式是

$$W = -\alpha L - \beta G - \eta C(N) + \rho A_c,$$

其中 $C(N)$ 是随有效政党数上升而增加的组阁协调成本， $\alpha, \beta, \eta, \rho$ 是评价权重。

这个式子不必真的估计每个参数。它把争论分解成几种权重：若问责清晰度权重 ρ 很大，FPTP 的单一赢家会有优势；若比例性权重 β 很大，PR 会有优势；若平均政策距离权重 α 很大，仿真中选出 B 的制度会更好；若联盟谈判成本权重 η 很大，过高的有效政党数会被扣分。

例 9.1（不同权重下的制度排序）

把 FPTP 的问责清晰度记为 $A_c = 1$ ，PR 的问责清晰度记为 $A_c = 0.5$ 。暂时只比较 FPTP 与 D'Hondt PR，并把 PR 的政策损失留给联盟阶段，不直接计入。若评价者取

$$\beta = 1, \quad \rho = 80,$$

则 FPTP 相对 PR 的问责收益约为

$$80(1 - 0.5) = 40,$$

但票席失真劣势为

$$54.95 - 6.08 = 48.87.$$

在这个权重下，比例性损失超过问责收益，PR 更好。若 ρ 提高到 120，问责收益变成 60，FPTP 可能更好。制度排序依赖社会目标函数。

这个结构和效用最大化相似。商品组合要在偏好和预算约束下比较，制度也要在给定的目标下比较。更具体地说，社会愿意用多少比例性换取问责清晰度，用多少代表性换取组阁稳定，用多少简单性换取更丰富的偏好表达。

9.3 指标从哪里来

政策距离损失来自空间投票模型。若选民效用由距离决定，平均距离就是一个直接的福利近似。它不等于完整社会福利，因为它没有包含 valence、偏好强度和联盟政策，但它能回答一个明确问题：胜者在政策空间上离选民远不远。

票席失真来自比例代表的评价问题。选民投了多少票，最后得到多少席位？如果一个制度把 33.3% 的票变成 100% 的席位，它的问责性可能很清楚，但比例性很差。

有效政党数来自政党系统研究。它不是简单数政党个数，而是按席位权重计算。一个 90 席大党加十个 1 席小党，与四个规模接近的政党，政治含义不同。

这三个指标的作用，是把“制度好不好”拆开。一个制度可以在一个指标上好，在另一个指标上差。讨论制度时必须先说目标。

说明 9.1（指标的限度）

这些指标仍然很粗。政策距离损失忽略了偏好强度和 valence；Gallagher index 只看票席比例，不看政府能否形成；有效政党数只看席位分散，不直接衡量意识形态距离。它们的价值在于能和前面的模型相连：距离损失来自空间效用，票席失真来自席位公式，有效政党数来自政党进入与联盟谈判。若继续做研究，可以在这些指标上加入真实选举数据、选民调查或议会联盟数据。

9.4 前沿制度：把 tradeoff 换一种形式

传统制度多半在“一个人一票、选一个候选人”附近变化。近年的计算社会选择和机制设计文献把问题继续往前推：如果选民可以批准多个候选人、如果要选出一个委员会、如果偏好强度也能进入机制，制度会怎样改变？这些制度没有消灭前面的不可能性结论，它们只是换了一组取舍。

9.4.1 Approval-based multiwinner voting

Approval-based multiwinner voting⁴⁴ 让每位选民提交一个批准集合 A_i ，制度从候选人集合中选出 k 人委员会 W 。这类规则适合解释“多赢家代表性”：班级委员会、议会名单、专家小组和公共项目包，都不一定只需要一个赢家。Lackner 和 Skowron 对这一方向有系统综述 [9]。

最简单的 approval voting 会选批准票最多的 k 个候选人，但它可能让最大群体包揽全部席位。PAV⁴⁵ 的想法是给同一选民的第一个代表较高价值，第二个代表价值较低，第三个更低。若选民 i 在委员会中有 $t_i = |W \cap A_i|$ 个被批准候选人，PAV 给他的效用为

$$H(t_i) = 1 + \frac{1}{2} + \cdots + \frac{1}{t_i},$$

其中 $H(0) = 0$ 。委员会得分为

$$\sum_i H(|W \cap A_i|).$$

这个调和数形式把“边际代表收益递减”写进规则。一个群体已经有一名代表后，再给它第二名仍有价值，但价值只有 $1/2$ ；第三名价值只有 $1/3$ 。

⁴⁴ **Approval-based multiwinner voting** (approval-based multiwinner voting；基于批准票的多赢家投票)：选民提交批准集合，规则从候选人中选出固定规模委员会的多赢家投票框架。

⁴⁵ **PAV** (Proportional Approval Voting；比例批准投票)：用调和收益计算委员会得分、鼓励不同群体获得比例代表的批准票多赢家规则。

例 9.2 (PAV 如何阻止最大群体包揽)

设要选 $k = 3$ 人委员会。有 6 名选民：

3 人批准 $\{a, b, c\}$, 2 人批准 $\{d, e\}$, 1 人批准 $\{f\}$.

若委员会为 $W_1 = \{a, b, c\}$, 前三名选民各得到三个代表, 效用为

$$H(3) = 1 + \frac{1}{2} + \frac{1}{3} = \frac{11}{6}.$$

总得分为

$$3 \times \frac{11}{6} = 5.5.$$

另外三名选民没有代表。若委员会为 $W_2 = \{a, b, d\}$, 前三名选民各有两个代表, 后两名选民各有一个代表, 得分为

$$3H(2) + 2H(1) = 3 \left(1 + \frac{1}{2}\right) + 2 = 6.5.$$

W_2 的 PAV 得分更高。规则宁愿让最大群体少拿第三个代表, 也要让第二个群体至少得到一个代表。

这个例子和 D'Hondt、Sainte-Lague 的边际席位逻辑很像：代表性不是简单数第一偏好，而是看每个群体是否已经被代表。差别在于，PAV 不需要预先把候选人分成政党；它直接用选民批准集合衡量代表性。

9.4.2 Method of Equal Shares

MES⁴⁶ 把代表性写成预算分摊问题。每个选民初始有同样数量的虚拟预算，候选人有固定成本；如果某个候选人的批准者能共同支付成本，就可以把该候选人选入委员会，并从这些批准者预算中扣款。Peters 和 Skowron 把 MES 放在 Phragmen、Thiele 和比例性公理的比较中讨论 [13]。

例 9.3 (Equal Shares 的预算直觉)

仍设 6 名选民、委员会大小 $k = 3$, 每个候选人成本为 2, 每位选民初始预算为 1, 总预算正好买 3 名候选人。批准集合仍为

3 人批准 $\{a, b, c\}$, 2 人批准 $\{d, e\}$, 1 人批准 $\{f\}$.

候选人 a 有 3 名批准者, 三人共同支付成本 2, 每人支付 $2/3$, 各剩 $1/3$ 。候选人 d 有 2 名批准者, 每人支付 1, 二人预算用完。此时最大群体三人合计还剩 1, 已经不足以再单独买一个成本为 2 的候选人；小群体也没有预算。若规则一开始连续选择 a, b, c , 最大群体三人

⁴⁶ **MES** (Method of Equal Shares; 等额份额法): 把每位选民视为拥有相同预算, 并让批准者共同支付候选人成本的多赢家代表性规则。

需要支付总成本 6，但它们总预算只有 3，做不到。预算约束因此阻止一个群体用重复批准包揽全部委员会。

MES 的直觉是预算公平。每个选民有相同购买力，候选人是公共项目，批准者共同分摊成本。制度设计的目标不再只是“谁票多”，而是让不同群体的预算都能转化为相应代表。它把比例性解释成一种预算约束下的公平。

9.4.3 Quadratic voting

前面所有制度主要处理排序或批准集合，很少处理偏好强度。Quadratic voting⁴⁷ 直接把强度纳入机制：选民可以购买多张赞成票或反对票，但购买 v 张票需要支付 v^2 个 voice credits⁴⁸ [10]。成本表为

v	0	1	2	3
v^2	0	1	4	9

边际成本依次为 1, 3, 5, ...。偏好越强，买更多票越有意义；但成本递增会阻止某个人轻易买下大量影响力。

一个简单的一阶条件可以说明为什么是平方成本。设选民购买 v_i 张票，改变结果的边际概率为 p ，他对结果改变的价值为 b_i 。净收益近似为

$$pb_i v_i - v_i^2.$$

最大化得到

$$pb_i - 2v_i = 0, \quad v_i = \frac{pb_i}{2}.$$

最优票数和偏好强度 b_i 成正比。线性成本下，强偏好者容易把预算一次性砸到最大；平方成本下，表达强度仍然可能，但边际成本不断上升。

例 9.4（偏好强度如何改变投票解释）

设一个公共项目有 5 名轻度支持者，每人价值 +1，还有 1 名强烈反对者，价值 -6。普通多数制只看人数，项目以 5 : 1 通过。若用 quadratic voting，轻度支持者各买 1 张赞成票，总赞成票为 5，成本各为 1；强烈反对者若有足够 credits，买 3 张反对票成本为 9，反对票为 3，项目仍通过。若他买 5 张反对票，成本为 25，项目打平或失败。平方成本使强烈偏好可以表达，但表达强烈反对必须付出显著更高成本。制度把“人数多数”和“强度少数”放到同一个计算里。

Quadratic voting 的困难也很明显。它需要给每个人相同或可控制的 voice credits，否则富人购买力会变成政治权力；它还要求选民理解成本函数，并防止交易、合谋和身份重复。它适合说明机制设计如何处理偏好强度，却不能简单替代普通政治选举。

⁴⁷ **Quadratic voting** (quadratic voting; 平方投票)：选民可购买多张赞成票或反对票，但购买票数的成本按平方增长的偏好强度表达机制。

⁴⁸ **voice credits** (voice credits; 投票信用)：平方投票中用于购买投票影响力的虚拟预算单位。

9.4.4 计算复杂性也是制度约束

前沿制度还带来计算问题。PAV 要在许多可能委员会中寻找最高得分组合，候选人一多，组合数量迅速增长。若有 m 名候选人，要选 k 人委员会，可能委员会数为

$$\binom{m}{k}.$$

当 $m = 30, k = 10$ 时，

$$\binom{30}{10} = 30045015,$$

已经超过三千万种。规则在理论上漂亮，不等于现实中容易计票、解释和审计。计算社会选择的贡献就在这里：它把“规则满足什么公理”和“规则能不能有效算出结果”放在一起研究。

这些制度回到同一条主线：PAV 和 MES 处理多赢家代表性，解决的是 PR 与 approval 的结合问题；quadratic voting 处理偏好强度，回应的是普通投票只读取序数信息的缺陷；计算复杂性则表明，机制的可行性也是约束。它们没有绕开 Arrow 或 Gibbard-Satterthwaite，只是把不可能性定理留下的 tradeoff 重新分配到比例性、强度表达、预算公平、计算可行性和可解释性之间。

10 结论与制度选择建议

Hotelling-Downs 模型给出的中位选民逻辑，是一维、两候选人、单峰偏好下的基准。多维偏好进入以后，选民有议题权重、valence 和底线约束；多数关系可能循环，政党票也可能把逐议题多数压成相反的整体结果。制度在这个位置发挥作用：它规定选民能报告什么信息，也规定这些信息如何变成胜者、席位和谈判权。

回到开头的模型，最重要的收获是看清楚结论从哪些假设中来。一维、单峰、两候选人、真诚投票、单赢家，这些条件合在一起，会把政治竞争推向中位点；一旦加入多维议题、valence、底线约束、策略投票、政党进入和席位公式，政治市场的收益函数就变了。候选人和选民会在不同制度给出的约束下重新做最优反应。这个视角比简单评价“哪种制度最好”更有解释力。

10.1 没有脱离目标的最优制度

FPTP 读取第一偏好，容易形成清楚赢家，也容易制造弃保和票席失真。两轮制和 IRV 读取后续偏好，使广泛可接受的候选人更有机会胜出。Borda 和 Condorcet 进一步利用完整排序或成对多数信息，但也带来复杂性。PR 不再追求一个赢家，而是把多个选民群体转化为席位。第九章的目标函数

$$W = -\alpha L - \beta G - \eta C(N) + \rho A_c$$

把这个判断写得更清楚：制度排序取决于社会更重视哪一种损失和哪一种收益。

表 33: 不同制度目标下的选择方向

主要目标	更容易被支持的制度方向	需要付出的代价
清楚问责和单一赢家	FPTP、两轮制、排序单赢家规则	代表性下降，小党和第二偏好被压缩
多数可接受性	两轮制、IRV、Condorcet、Borda	计票和解释更复杂，策略空间仍存在
比例代表和小群体进入	PR、Sainte-Lague、MMP、STV 类制度	联盟谈判成本上升，问责链条变长
偏好强度表达	quadratic voting 等机制	预算分配、合谋和理解成本更高
多赢家代表性	PAV、MES、多赢家 approval 规则	计算复杂性和制度解释难度上升

表 33 不是给出唯一答案，而是把制度选择变成可以讨论的条件句：如果目标是问责清晰，单赢家规则会更有吸引力；如果目标是降低票席失真，比例代表更自然；如果目标是让小群体不被压成零，多赢家和比例规则更合适；如果目标是表达偏好强度，就要面对预算和合谋问题。

10.2 分析主线的最终表述

选举制度分析要抓住四个环节：偏好、约束、策略、结果。偏好不是单一排序，而可能来自多维效用、议题权重和候选人质量；约束不是外在背景，而包括选票格式、席位公式、门槛和议程；策略不是道德瑕疵，而是行为人对规则的最优反应；结果也不只是“谁赢”，还包括席位结构、政策距离损失、票席失真、联盟谈判和可操纵性。

因此，选举制度不是简单计票方法，而是一套把私人偏好转化为公共选择的机制。没有一种制度能同时最大化代表性、问责性、稳定性、简单性和防操纵性。比较制度时，最有用问题不是“哪种制度最好”，而是：在给定的社会目标函数下，哪一种制度把最重要的信息保留下来，并把最不能接受的策略扭曲控制在较低水平。

参考文献

[1] Kenneth J. Arrow. *Social Choice and Individual Values*. Yale University Press, New Haven, 2 edition, 1963. Book; first edition 1951.

[2] Michel L. Balinski and H. Peyton Young. *Fair Representation: Meeting the Ideal of One Man, One Vote*. Yale University Press, New Haven, 1982. Book; no DOI used.

[3] Anthony Downs. *An Economic Theory of Democracy*. Harper and Row, New York, 1957. Book; no DOI used.

[4] Michael Gallagher. Proportionality, disproportionality and electoral systems. *Electoral Studies*, 10(1):33–51, 1991. doi:10.1016/0261-3794(91)90004-C.

- [5] Allan Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4):587–601, 1973. doi:[10.2307/1914083](https://doi.org/10.2307/1914083).
- [6] Harold Hotelling. Stability in competition. *The Economic Journal*, 39(153):41–57, 1929. doi:[10.2307/2224214](https://doi.org/10.2307/2224214).
- [7] Ralph L. Keeney and Howard Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge University Press, 1993. doi:[10.1017/CB09781139174084](https://doi.org/10.1017/CB09781139174084).
- [8] Markku Laakso and Rein Taagepera. “effective” number of parties: A measure with application to west europe. *Comparative Political Studies*, 12(1):3–27, 1979. doi:[10.1177/001041407901200101](https://doi.org/10.1177/001041407901200101).
- [9] Martin Lackner and Piotr Skowron. *Multi-Winner Voting with Approval Preferences*. Springer, 2023. Open access book. doi:[10.1007/978-3-031-09016-5](https://doi.org/10.1007/978-3-031-09016-5).
- [10] Steven P. Lalley and E. Glen Weyl. Quadratic voting: How mechanism design can radicalize democracy. *AEA Papers and Proceedings*, 108:33–37, 2018. doi:[10.1257/pandp.20181002](https://doi.org/10.1257/pandp.20181002).
- [11] Kenneth O. May. A set of independent necessary and sufficient conditions for simple majority decision. *Econometrica*, 20(4):680–684, 1952. doi:[10.2307/1907651](https://doi.org/10.2307/1907651).
- [12] Richard D. McKelvey. Intransitivities in multidimensional voting models and some implications for agenda control. *Journal of Economic Theory*, 12(3):472–482, 1976. doi:[10.1016/0022-0531\(76\)90040-5](https://doi.org/10.1016/0022-0531(76)90040-5).
- [13] Dominik Peters and Piotr Skowron. Proportionality and the limits of welfarism. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 793–794, 2020. doi:[10.1145/3391403.3399465](https://doi.org/10.1145/3391403.3399465).
- [14] Charles R. Plott. A notion of equilibrium and its possibility under majority rule. *American Economic Review*, 57(4):787–806, 1967. Classic article; stable bibliographic record checked, no DOI used.
- [15] George Rabinowitz and Stuart Elaine Macdonald. A directional theory of issue voting. *American Political Science Review*, 83(1):93–121, 1989. doi:[10.2307/1956436](https://doi.org/10.2307/1956436).
- [16] Douglas W. Rae and Hans Daudt. The ostrogorski paradox: A peculiarity of compound majority decision. *European Journal of Political Research*, 4(4):391–398, 1976. doi:[10.1111/j.1475-6765.1976.tb00542.x](https://doi.org/10.1111/j.1475-6765.1976.tb00542.x).

- [17] Mark Allen Satterthwaite. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, 1975. doi:[10.1016/0022-0531\(75\)90050-2](https://doi.org/10.1016/0022-0531(75)90050-2).
- [18] Donald E. Stokes. Spatial models of party competition. *American Political Science Review*, 57(2):368–377, 1963. doi:[10.2307/1952828](https://doi.org/10.2307/1952828).