Week 1: Preliminaries and electoral accountability

A second introduction to formal political economy, Trinity Term 2025

Jacob Edenhofer

29/04/2025

* jacob.edenhofer@nuffield.ox.ac.uk

Overview

1. Briefly discuss course logistics
2. Outline the core themes and explain the logic underlying the course structure
3. Introduce four ideas about electoral accountability – two formally, and the other two informally
4. Conclude and provide short outlook

Outline

About me

I am a first-year DPhil Student in Politics, supervised by Federica Genovese and David Rueda. My main research interests are the comparative political economy of climate policy and right-wing populism in Europe.

Climate policy

- Edenhofer and Genovese, 2024 (Link)
- Zwar, Edenhofer, and Flachsland, 2024 (WP link, Ariadne report, Verfassungsblog post)
- Konc et al., 2024 (Short summary)
- Edenhofer, 2024
- Edenhofer and Flachsland, 2025 (WP link)

Populism

- Alabrese et al., 2024 (Short summary, FT coverage, Interactive website)
- Fetzer, Shaw, and Edenhofer, 2024 (VoxEU summary)
- Fetzer, Edenhofer, and Garg, 2025 (FAZ coverage)
- Gratton and Edenhofer, 2025 (Short summary)

@edenhofer_jacob Often wrong, but sometimes useful (Substack)

Recapitulating the objectives of the course

Managing expectations: This course is intended as a complement to Ben Ansell's formal theory class in Hilary.

Constraints: We only have four sessions, and this course assumes very little prior exposure to formal theory or mathematical background knowledge

Best response: Four, formally speaking, fairly light-touch sessions, focused on understanding the intuitions underlying the formal arguments

My hope for this course is threefold: that it will help you develop a better understanding of the workings of modern democracies, equip you with a richer theoretical toolkit, and give you a sense as to where to look if you wish to delve deeper into some of these aspects

Resources, assignments, and teaching style

Central resources:

- Syllabus – quite extensive to reduce search costs for you
- Github course page (Link)

Assignments: You will be able to choose between three types of assignments: (i) formal exercises, (ii) short essays, (iii) simulations. I'll keep them short – well aware that most of you have a lot of other work to do → use of generative AI actively encouraged with some caveats

Teaching style: Motivation + one or two formal deep dives + real-world applications → happy for you to actively shape style of classes, just let me know (e.g. setting up models in class)

How to get most out of the course?

In class: You are free to follow the lectures in whatever way works best for you, though my view is that students get most of out lectures when (i) they minimise sources of distraction and (ii) actively participate (e.g. ask questions, challenge me, spot mistakes)

At home: Please peruse the required readings; if pressed for time, read whatever you find most interesting, but do so carefully (be able to summarise, apply, and extend logic). Generative AI is a fantastic resource, but not easy to use responsibly.

Feedback: This is my first time teaching a course – so, I will likely make mistakes or get things wrong. Let me know asap, rather than taking it on the chin. Feel free to arrange office hours via email – happy to discuss assignments or your thesis.

Outline

The starting point: Democracy and the ideal of self-government

Democracy's original ideal (Rousseau):

→ Set people free through self-government (Dunn, 2018)

→ See Grofman and Feld, 1988 for an excellent discussion

Self-government:

→ Citizens participate in law-making, being bound only by laws they themselves affirm (Przeworski, 2010). Then the laws reflect the will of the people, making them "free".

Key assumptions:

- Population interests are sufficiently homogeneous (consensus is possible).
- Costs of reaching consensus are low.

▸ Summary of Grofman and Feld

The problem of scale and the ideal of self-government

Small-scale self-government:

→ Feasible only in small, homogeneous societies.

→ Historically, other conditions also mattered: exit options and private information (Bates and Lien, 1985; Stasavage, 2020)

→ Question: How does scale change with technology, and why? Examples?

Large-scale democracies:

→ Transaction costs (time, coordination) rise with population size, even when interests are fairly homogeneous.

→ With heterogeneous interests, it is extremely complicated – if not impossible – to find a consensus that satisfies even "minimal" conditions, as Schumpeter, 1942 argued and Arrow, 1951 – generalising Condorcet's paradox – showed.

▸ Example for Condorcet cycle ▸ Arrow's theorem (formal) ▸ Arrow's theorem (intuitive)

How scale affects the nature of politics

Criterion (explanation)	Small-scale polities	Large-scale polities
Efficiency (Matching outputs to preferences/policies)	Preference efficiency; Policies closely match citizen preferences	Policy efficiency; Better internalisation of externalities; Large-scale public goods
Political Relationships (Formality of interaction)	Informal, personal ties; Clientelism, grassroots negotiation	Formal, institutionalised roles; Professionalised bureaucracy
Systems of Rule (Concentration vs. dispersion)	Intensive rule; Concentration of authority over few	Extensive rule; Dispersion of authority across wider populations
Popular Rule (Citizen participation)	Participatory democracy; Direct citizen decision-making	Representative democracy; Rule through elected representatives

Abridged version of the table in Gerring and Veenendaal, 2020

The second-best ideal of self-government

Fundamental implications of scale and preference aggregation:

- Division of labour is necessary to manage the high costs of collective decision-making.
- Laws likely cannot encode the "will of the people". Instead, representatives must craft laws that peacefully resolve societal conflicts.
- To prevent abuses of power:
  - → Representatives must contest **free and competitive elections**.
  - → Elections confer power **temporarily**.
- For losers to accept outcomes:
  - → All citizens must enjoy **equal, enforceable rights** in some domains.

Second-best ideal: Self-government aims to maximise satisfaction (or minimise dissatisfaction) across citizens by combining **representation** with **equal rights** (Przeworski, 2010).

How do modern democracies realise the second-best ideal?

Key implication:

- In large, heterogeneous societies, democracy is less about accurately aggregating preferences, and more about regulating competition for power.

Strategic logic of modern democracy:

- Institutionalise peaceful competition for office among elites and parties.
- Ensure sufficient responsiveness to citizens' interests and protect inviolable rights to guarantee that losers continue to abide by collective decisions.

Our focus:

- Study the institutions that structure and regulate this competition: **Elections, parties, electoral systems, interest groups, legislatures, judiciary, etc.**.

Formal theory helps us understand the “miracle of democracy”

> *“The miracle of democracy is that conflicting political forces obey the results of voting. People who have guns obey those without them. Incumbents risk their control of governmental offices by holding elections. Losers wait for their chances to win office.* ***Conflicts are regulated, processed according to rules****, and thus limited. This is not consensus, yet not mayhem either. Just regulated conflict; conflict without killing.* ***Ballots are ‘paper stones’.”*** *(Przeworski, 2018, p. 118)*

Key points:

- Institutions (elections, parties, legislatures, judiciary) enable large, heterogeneous societies to live in peace while ensuring a minimum level of liberty.
- This outcome may seem underwhelming — societal peace is not the same as justice.
- Yet, preserving this peace is no mean feat either (let’s talk after the mid-terms).

Outline

Setting the stage for today's sessions: What's the deal with elections?

Fundamental role: Competitive, free, and regular elections regulate societal conflicts peacefully because they perform two functions.

1. **Alternation in office**:
   - → Provides losers with incentive to abide by laws crafted by winners.
   - → Elections render conflicts intertemporal — losers wait for another chance (Y. Li, 2019; Przeworski, 1991). Why not just toss a fair coin?
2. **Voting and accountability**:
   - → Voting ensures population preferences are considered (to some extent).
   - → Without voting → dissatisfaction → instability.

"[The losers] will wait, as long as the policies imposed by the winners are not too extreme or as long as their chance to win at the next opportunity is sufficiently high. [...] Winners have to moderate their policies." (Przeworski, 2018, p. 116)

Elections and the intertemporal dimension of political conflicts

Dynamic Payoffs:

- Elections occur infinitely often; discount factor $\delta \in (0, 1)$. Time is discrete: $t = 0, 1, 2, \ldots$
- In each period:
  - $\rightarrow$ Holding power yields payoff $R > 0$,
  - $\rightarrow$ Being in opposition yields payoff L (where $L < R$),
  - $\rightarrow$ Political engagement costs $k > 0$,
  - $\rightarrow$ Probability of winning future elections: π depends non-monotonically on electoral stakes $(S \equiv R - L)$, i.e. $\exists S^* > 0$ s.t. $\pi' > 0$ for $S < S^*$ and $\pi' < 0$ for $S > S^*$

Expected lifetime payoff from complying:

$$V_{\text{comply}} = \frac{\pi(R - L) + L - k}{1 - \delta}$$

Based on: Fearon, 2011; Przeworski, 2018; Przeworski, Rivero, and Xi, 2015

Deriving the compliance payoff: Step-by-step derivation

1. Per-period expected payoff:

$$\text{Payoff} = \pi R + (1-\pi)L - k = \pi(R-L) + L - k$$

2. Lifetime discounted payoff (starting in period $t = 0$):

$$V_{\text{comply}} = \sum_{t=0}^{\infty} \delta^t \left[\pi(R-L) + L - k\right]$$

3. Factor out the constant:

$$V_{\text{comply}} = \left[\pi(R-L) + L - k\right] \cdot \sum_{t=0}^{\infty} \delta^t$$

4. Use the geometric series formula:

$$\sum_{t=0}^{\infty} \delta^t = \frac{1}{1-\delta} \quad \text{(for } 0 < \delta < 1\text{)}$$

5. Final expression:

$$V_{\text{comply}} = \frac{\pi(R-L) + L - k}{1-\delta}$$

The compliance condition

Expected lifetime payoff from fighting:

$V_{\text{fight}} = p\frac{R}{1-\delta} - c$ where p = probability of success, and $c > 0$ is the cost of fighting

Compliance condition: $V_{\text{comply}} \geq V_{\text{fight}}$

$$\frac{\pi(R-L) + L - k}{1-\delta} \geq p \cdot \frac{R}{1-\delta} - c$$

Multiply both sides by $1-\delta$:

$$\pi(R-L) + L - k \geq pR - c(1-\delta)$$

Understanding the cost term in the fighting payoff

Why is c not divided by $1-\delta$?

- The formula:

$$V_{\text{fight}} = p \cdot \frac{R}{1-\delta} - c$$

  assumes that:

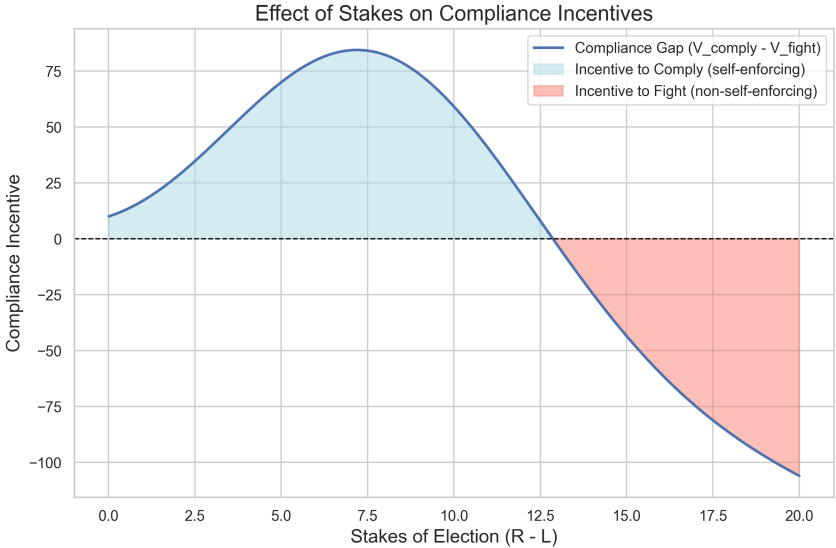
  → Fighting incurs a **one-time cost** $c > 0$ in period $t = 0$,

  → Not a recurring cost over time.

- This reflects a single event like a coup, war, or rebellion — the cost is paid once.
- In contrast, if the cost were **recurring** (e.g., cost of maintaining conflict each period), it would be discounted like payoffs:

$$\text{Recurring cost: } \sum_{t=0}^{\infty} \delta^t c = \frac{c}{1-\delta}$$

Illustrating the compliance condition



Comparative statics by regime type and electoral incentives

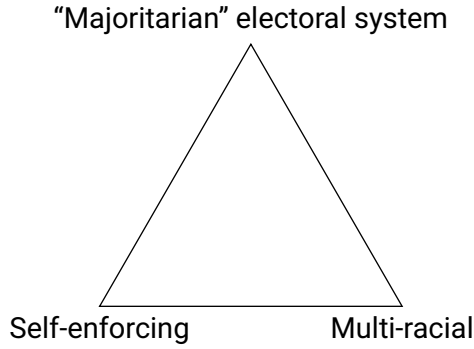
Parameter effect	$\pi(S^*) > p$		$\pi(S^*) < p$	
	$\pi'(S^*) > 0$	$\pi'(S^*) < 0$	$\pi'(S^*) > 0$	$\pi'(S^*) < 0$
$\frac{dS^*}{dc}$	> 0	< 0	> 0	< 0
$\frac{dS^*}{dk}$	< 0	> 0	< 0	> 0
$\frac{dS^*}{d\delta}$	< 0	< 0	< 0	< 0
$\frac{dS^*}{dp}$	< 0	< 0	< 0	< 0

Interpretation: The effects of c and k flip sign depending on whether electoral stakes increase or reduce competitiveness (i.e., the sign of $\pi'(S^*)$), provided the slope of F w.r.t. S reverses. By contrast, the effects of δ and p are consistently signed across institutional contexts.

▸ Formal derivations

A useless model for advanced democracies? An application to the US

Source: Brookings Institution



Question: Is this the key trilemma of US democracy?

Relates to topical debates: desirability and feasibility of electoral reform in the US (Carey and Pocasangre, 2024; Drutman, 2021; Latner, Santucci, and Shugart, 2021; Mainwaring and Drutman, 2023; Santucci, 2022) and whether the GOP is becoming a multi-racial party (Burn-Murdoch, 2024; Ruffini, 2023)

Why do we vote, rather than just tossing a fair coin?

Because voting allows us to communicate to parties and candidates which policies or policy platforms we find relatively more appealing, thus helping to minimise our dissatisfaction / making compliance likelier

Prospective voting (Selection)

- Informational demands: Voters need to know parties' platforms, candidates' competence, and their own ideal point → spatial logic: select candidate closest to own ideal point, conditional on competence
- Mandate conception of representation: voters enter into implicit contract with representative, whose performance is evaluated by the extent to which (s)he implements promised platform (Manin, Przeworski, and Stokes, 1999)

Retrospective voting (Sanctioning)

- Informational demands: "In order to ascertain whether the incumbents have performed poorly or well, citizens need only calculate the changes in their own welfare." (Fiorina, 1978, p. 5) → decision rule + attribution of responsibility
- Accountability conception of representation: "voters vote to retain the incumbent only when the incumbent acts in their best interest, and … [she] chooses policies necessary to get re-elected." (Manin, Przeworski, and Stokes, 1999, p. 40)

Ferejohn, 1986 on electoral control

Goal: Understand how elections can discipline incumbents.

- Voters cannot directly observe politicians' type (competent or incompetent).
- Voters only observe outcomes (e.g., economic performance, public goods).
- Elections function as incentive mechanisms: bad performance is punished by removal from office.
- Politicians maximise their probability of reelection by choosing an effort level $e \geq 0$. Effort represents the incumbent's exertion, competence, or quality of governance, but it is personally costly to the politician. (S)he incurs $c(e)$.

Ferejohn, 1986: Model primitives — Outcome and noise

Observable outcome:

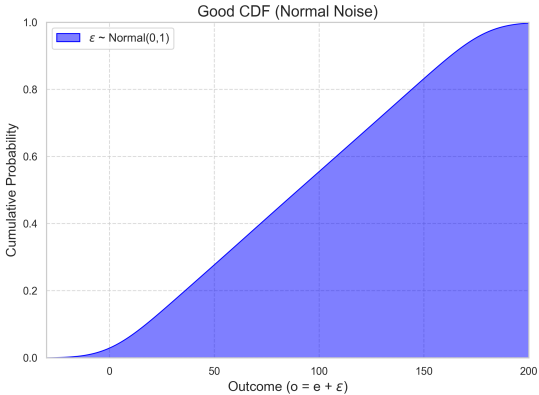
- The observed performance outcome is:

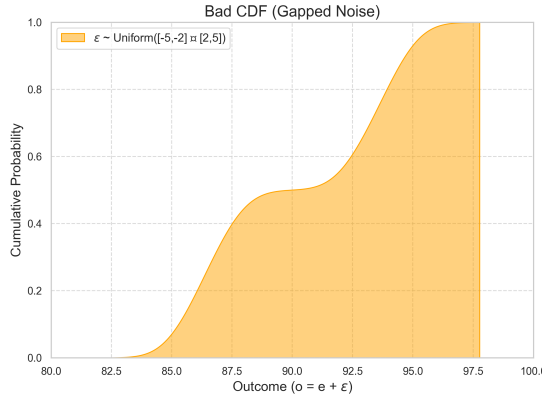
$$o = e + \varepsilon \Rightarrow \epsilon = o - e$$

  where ε is a random variable.

- The noise term, ε, has a continuous, strictly increasing cumulative distribution function $F(\cdot)$, typically mean zero.
- Interpretation: Outcomes are noisy signals of effort; random shocks beyond the incumbent's control affect performance.
- Symmetry of ε is not required: only continuity and strict monotonicity.

Illustrating the assumptions about $F(\cdot)$





Ferejohn, 1986: Model primitives — Voters' information and strategy

Voters' Observations:

- Voters observe only the realized outcome o, not the true effort e.
- They apply a simple cutoff rule:

$$\text{Reelect if } o \geq o^*, \quad \text{replace otherwise.}$$

- Interpretation: Voters do not infer effort (e.g. via Bayes' rule); they judge based solely on observed performance against a threshold (more below).

Ferejohn, 1986: Model Primitives — Incumbent's payoff

Incumbent's Utility:

- If reelected, the incumbent receives:

$$R - c(e)$$

  where:

  - $R > 0$ is the benefit of holding office.
  - $c(e)$ is the cost of effort, increasing and convex, i.e. $c', c'' > 0$.

- If not reelected, the payoff is 0.
- Interpretation: Incumbents weigh the office-holding benefit against the cost of the effort required for staying in power.

Ferejohn, 1986: Timing and equilibrium structure

Timing:

1. **Voters** choose cutoff o^* anticipating incumbent's response.
2. **Incumbent** chooses effort e to maximise the probability of reelection minus cost.
3. **Nature** realises outcome $o = e + \varepsilon$.
4. **Voters** observe o and decide whether to reelect.

Equilibrium Reasoning:

- Voters: Choose o^* to discipline the incumbent into exerting high effort e^*.
- Incumbent: Chooses e given o^*, solving

$$c'(e) = R \cdot f(o^* - e)$$

Ferejohn, 1986: How is the cutoff o^* determined?

Voter strategy:

- Voters choose o^* to maximise their expected utility.
- Specifically, voters set o^* high enough to discipline the incumbent into choosing a desired (high) effort level e^*.
- Interpretation: The cutoff acts as an incentive device — only good enough performance is rewarded.

Important:

- o^* must balance two goals:
  - $\rightarrow$ Induce high effort (e^*) by threatening removal.
  - $\rightarrow$ Ensure reelection is still realistically achievable, otherwise no effort is exerted.
- If the cutoff is set too high, even good incumbents would give up.

Ferejohn, 1986: The incumbent's optimisation problem

Incumbent's problem:

- Given the cutoff o^*, the incumbent chooses effort e to maximise:

$$\max_{e} \quad R \cdot (1 - F(o^* - e)) - c(e)$$

where:

- $\rightarrow$ R is the benefit from reelection,
- $\rightarrow$ $F(o^* - e)$ is the probability of an outcome below o^* (i.e., being voted out of office),
- $\rightarrow$ $c(e)$ is the cost of effort.

Ferejohn, 1986: Deriving the FOC

Taking the first-order condition:

$$R \cdot f(o^* - e) - c'(e) = 0$$

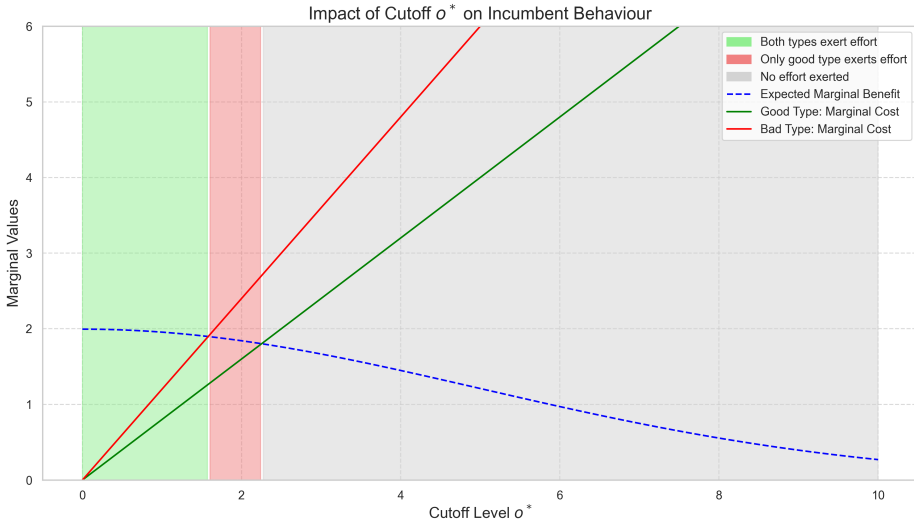
which rearranges to:

$$\underbrace{c'(e)}_{\text{Marginal Cost (certain)}} = \underbrace{R \cdot f(o^* - e)}_{\text{Expected Marginal Benefit (uncertain via } \varepsilon)}$$

Interpretation:

- The incumbent exerts effort until the certain marginal cost equals the expected marginal benefit of improving reelection chances.
- Higher sensitivity of outcomes ($f(o^* - e)$) makes effort more attractive.

Illustrating the FOC



Ferejohn, 1986: Equilibrium conditions

Equilibrium definition:

- A pair (o^*, e^*) constitutes an equilibrium if:
  1. **Incumbent Best Response**: Given o^*, the incumbent chooses effort e^* solving:

$$c'(e^*) = R \cdot f(o^* - e^*)$$

  2. **Voters' Optimal Cutoff**: Voters choose o^* to induce e^* that maximizes their expected utility, ensuring effort is exerted but reelection remains feasible.

Interpretation: Voters set standards that discipline the incumbent without discouraging all effort; the incumbent balances marginal costs and expected marginal benefits accordingly.

Mathematical tool: Implicit Function Theorem (IFT)

Problem setup:

- Suppose we have a condition $G(e, R) = 0$ that implicitly determines e as a function of R.
- Example: The incumbent's FOC links effort e and reward R.

IFT Intuition:

- Even if we cannot solve explicitly for e,
- We can still understand how small changes in R affect e^* by differentiating $G(e, R) = 0$ implicitly.

Key Idea: Think of e "adjusting" automatically to keep $G(e, R) = 0$ true when R changes.

Mathematical tool: Implicit Function Theorem (IFT), cont'd

Conditions for applying IFT:

- $G(e, R)$ must be continuously differentiable.
- The partial derivative $\frac{\partial G}{\partial e} \neq 0$ near the solution (i.e., the FOC crosses properly, not flat).

IFT Comparative statics formula:

$$\frac{de^*}{dR} = -\frac{\frac{\partial G}{\partial R}}{\frac{\partial G}{\partial e}}$$

Interpretation: This formula tells us how equilibrium effort e^* changes with changes in the reward R.

Exogenous parameters in Ferejohn's model

Symbol	Meaning	Assumptions / Properties
R	Reward from reelection	$R > 0$
o^*	Voter-set cutoff performance	Chosen to discipline without demotivating
$F(\cdot)$	CDF of noise ε	Continuous, strictly increasing
$f(\cdot)$	Density of noise ε	Positive, smooth (e.g., normal)
$c(e)$	Cost of effort function	Increasing and convex: $c'(e) > 0, c''(e) > 0$

Note: These parameters are taken as given when analysing the incumbent's optimal effort choice.

Comparative statics — Effect of cutoff o^*

Applying IFT:

$$\frac{de^*}{do^*} = \frac{\overbrace{R}^{>0} \cdot \overbrace{f'(o^* - e)}^{<0}}{\underbrace{c''(e)}_{>0} - \underbrace{R}_{>0} \cdot \underbrace{f'(o^* - e)}_{<0}} \quad \Rightarrow \quad \boxed{\frac{de^*}{do^*} < 0}$$

Intuition:

- Higher cutoff o^* increases the required performance.
- If $f'(o^* - e) < 0$ (declining density), marginal benefit from effort falls (single-peaked and symmetric).
- Thus, incumbents respond by exerting less effort.

Conclusion: Tougher cutoffs o^* discourage effort.

Lessons from Ferejohn, 1986 — Intuition

Key Insights:

- Elections act as instruments of performance-based accountability.
- Voters can discipline incumbents without needing full information — simple cutoff rules suffice.
- Incumbents respond by exerting effort to meet voter expectations and retain office.

Determinants of Accountability:

- *Higher reward* for reelection (R) strengthens incentives to work hard.
- *Tougher cutoffs* (o^*) discipline incumbents — but risk deterring effort if too strict.
- *Lower cost convexity* and *less noisy outcomes* (steeper f) improve electoral control.

Bottom line: Even simple electoral mechanisms can discipline politicians — but their success depends crucially on rewards, costs, and information noise.

Limitations of Ferejohn, 1986 — and the road ahead

Key limitations:

- No modelling of types: incumbents' intrinsic competence is unobserved and undiscussed.
- Purely retrospective control: no role for policy promises or ideological proximity.
- No misattribution and extraneous events: events beyond incumbents' control might lead voters to unfairly punish or reward them if attribution is not fully rational.
- No distinction between voters and electorates: aggregation across many voters is not modelled.
- No learning or communication: voters react mechanically to outcomes without receiving messages or cues from political elites.
- Assumes stationary cutoffs: Non-stationary (e.g. increasingly lenient) performance standards can better sustain incumbent effort (Acharya, Lipnowski, and Ramos, 2025; Gieczewski and C. Li, 2024).

Next steps:

- **Selection**: How voters screen for "good types" (Fearon, 1999).
- **Aggregation**: Why electorates can be more rational than individual voters (Ashworth and Fowler, 2020).
- **Information and persuasion**: How elites communicate effectively despite voter ignorance (Lupia and McCubbins, 1998).

Fearon, 1999: Model primitives — Players and structure

Players:

- Electorate (E): collective actor representing voters.
- Incumbent Politician (I).

Timeline:

- Two-period game:
  - → **Period 1**: Incumbent chooses policy $x \in \mathbb{R}$, welfare signal, z, is observed, voters decide whether to reelect.
  - → **Period 2**: Whoever holds office (incumbent or new candidate) chooses policy $y \in \mathbb{R}$.

Goal:

- Understand whether elections work by selecting good types or sanctioning bad behaviour.

The electorate's ideal point and preferences

- Voters have a single-peaked preference centered at $x = 0$.
- **Utility function** for a policy choice $x \in \mathbb{R}$:

$$u(x) = -x^2$$

- Interpretation:
  - $\rightarrow$ Utility declines quadratically as policy deviates from 0 in either direction.
  - $\rightarrow$ Policies farther from the ideal are increasingly worse (losses grow faster the further away).

$\rightarrow$ Analogous for second period

▸ Illustration of (non-)single-peaked preferences

Fearon, 1999: Model primitives — Welfare signal and noise

Outcome signal (first period):

$$z = -x^2 + \varepsilon$$

- Policy choice by the incumbent, denoted by $x \in \mathbb{R}$
- Noise term, ε, drawn from a continuous, symmetric, strictly unimodal (why strict?) distribution with pdf, $f(\cdot)$, and mean zero.
- Interpretation:
  - $\rightarrow$ z is a noisy signal of policy quality. Maximum welfare attained at $x = 0$, the electorate's ideal point. This makes electoral control harder.

Outcome signal (second period):

$$z' = -y^2 + \varepsilon' \quad \text{(if re-elected or replaced)}$$

- Same structure; ε' is independent draw from $f(\cdot)$

Types of politicians

Politicians may differ in their intrinsic type. Types are private information — voters cannot observe a politician's type directly.

- **Pure selection:**
  - → Some politicians are good types: they choose policies close to the electorate's ideal ($x = 0$).
  - → Others are bad types: they systematically choose policies far from the ideal (e.g., $x = \hat{x} > 0$).
- **Pure sanctioning:**
  - → All politicians are identical and opportunistic.
  - → Without electoral incentives, incumbents would selfishly choose $x = 1$ (far from ideal).
  - → Elections are needed to discipline behaviour.
- **Mixed case:**
  - → Let $\theta \in \{Good, Bad\}$. With probability $\alpha \in (0, 1)$, a politician is a good type, and with $1 - \alpha$ a bad one. If replaced, a new candidate is drawn randomly, so the expected quality is unchanged.

Bottom line: Elections either **select good types** or **discipline bad behaviour**.

Voter strategy — cutoff rule

Voter decision rule:

- After observing the first-period welfare signal z, voters apply a cutoff rule.
- **Rule:** Reelect incumbent if and only if $z \geq k$, where $k \in \mathbb{R}$ is the threshold. That is: $-x^2 + \varepsilon \geq k \Rightarrow \varepsilon \geq k + x^2$

Interpretation:

- If the welfare signal is sufficiently high ($z \geq k$), voters reelect.
- If the signal is too low ($z < k$), voters replace the incumbent with a new random candidate.

Why a cutoff?

- Imperfect monitoring: Voters can't observe policy choice x or type directly.
- A simple cutoff balances the risks of mistakenly retaining bad types versus unfairly removing good ones.

The selection–sanctioning trade-off

- Selection:
  - → Voters set a strict cutoff k to weed out bad types.
  - → Tight cutoff = fewer bad types survive → better future policy.
- Sanctioning:
  - → Reelection incentives discipline all politicians (even bad types) into behaving better now.
  - → Requires that incumbents think reelection is *achievable* — not hopeless.

Problem:

- If k is too strict → Bad types give up → No incentive to behave well (blunts efficacy of sanctioning).
- If k is too lenient → Too many bad types get re-elected (impinges on selection).

Bottom Line: Voters face a fundamental tension between *screening* and *disciplining* politicians.

Politician's problem (first period)

Incumbent's decision: Choose policy $x \in \mathbb{R}$ to maximise expected utility, given the voter's cutoff rule k.

Elements of payoff:

- Immediate welfare from first-period policy:

$$u(x) = -x^2$$

- Probability of reelection:

$$\Pr(\text{reelect}) = 1 - F(k + x^2) \quad \text{where } F \text{ is the CDF of } \varepsilon$$

- Second-period payoff if reelected: depends on type and policy in that period

Politician's objective function — General setup

Incumbent's objective: Choose $x \in \mathbb{R}$ to maximise expected utility:

$$EU(x) = -x^2 + \delta\left(R(x)W + (1 - R(x))\mathbb{E}_\theta[W + u(\hat{y})]\right)$$

where:

- First-period welfare: $-x^2$
- Discount factor: $\delta \in (0, 1)$
- Value of office: $W > 0$
- $R(x) = 1 - F(k + x^2)$ is the probability of re-election (cutoff rule based on noisy $z = -x^2 + \varepsilon$).
- $\mathbb{E}_\theta[W + u(\hat{y})]$ is expected payoff if replaced, where expectation is over type distribution:
  $\rightarrow$ $u(\hat{y}) = -\hat{y}^2$ is welfare under replacement.

Note: $R(x)$ and $\mathbb{E}[W + u(\hat{y})]$ encode the election mechanics and second-period incentives.

Specifying second-period payoffs: Pure selection model

Politicians differ by unobservable types (good vs. bad). Their second-period behaviour depends entirely on type, not incentives (no further reelection).

Good types:

- Choose second-period policy $y = 0$ (voters' ideal point).
- Maximises voter welfare: $u(0) = 0$.

Bad types:

- Choose $y = \hat{y} > 0$, deviating from the ideal.
- Deliver welfare $u(\hat{y}) = -\hat{y}^2$

Replacement candidate:

- If voters replace the incumbent, the new candidate is drawn randomly from the type distribution.
- Thus, the expected second-period payoff if replacing the incumbent is:

$$\mathbb{E}_\theta[W + u(\hat{y})] = W - \mathbb{E}_\theta[\hat{y}^2], \text{where } W > 0 \text{ is the value of holding office}$$

Specifying second-period payoffs: Pure sanctioning model

In the pure sanctioning model, all politicians are identical and opportunistic. There are no hidden types: second-period behaviour is uniform across all incumbents and replacements.

Second-period behaviour:

- Once re-elected (no further reelection), incumbents choose their selfishly preferred policy, e.g. $y = 1$.
- Welfare outcome for voters: $u(1) = -(1)^2 = -1$.

Replacement candidate:

- Replacement candidates are identical to incumbents.
- Thus, expected second-period payoff if replaced is:

  $\mathbb{E}[W + u(1)] = W - 1$, fixed and known

Key implication: Elections serve to discipline behaviour during the first term, not to select better types.

Deriving the FOC for the politician's first-period choice

Recall the politician's first-period maximisation problem:

$$\max_x \quad -x^2 + \delta \left(R(x) \cdot W + (1 - R(x)) \cdot \mathbb{E}_\theta[W + u(\hat{y})] \right)$$

Taking the derivative w.r.t. x**:**

$$\frac{d}{dx} \left[-x^2 + \delta \left(R(x)W + (1 - R(x))\mathbb{E}_\theta[W + u(\hat{y})] \right) \right]$$

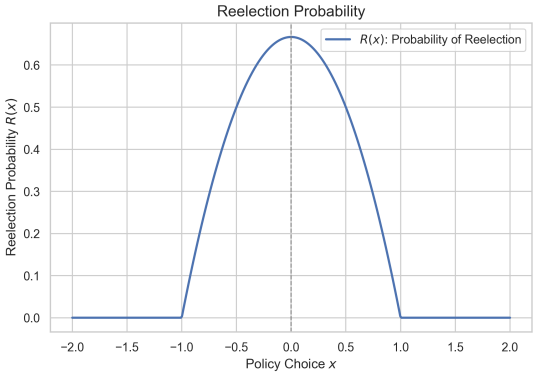
yields:

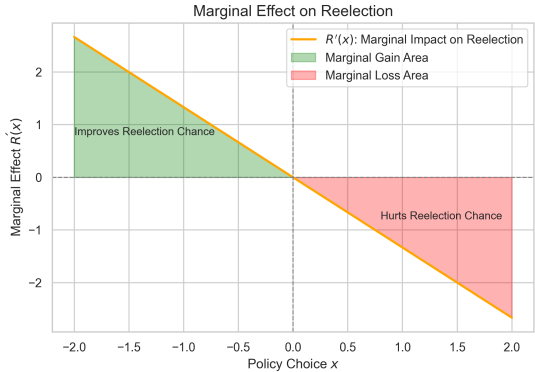
$$-2x + \delta R'(x) \left(W - \mathbb{E}_\theta[W + u(\hat{y})] \right) = 0$$

Thus, the FOC is:

$$\boxed{2x = \delta R'(x) \left(W - \mathbb{E}_\theta[W + u(\hat{y})] \right) = \delta R'(x) \mathbb{E}_\theta(\hat{y}^2)}$$

Visualising the trade-off implied by the FOC





Equilibrium structure under different models

Pure sanctioning model:

- No type heterogeneity: all politicians solve the FOC.
- Equilibrium behaviour: All incumbents choose $x^*(k)$ to maximise the probability of reelection.

Pure selection model:

- Good types choose $x = 0$ (ideal policy) without incentives.
- Bad types solve the FOC strategically.

Mixed model:

- Good types: $x = 0$.
- Bad types: Choose $x^*(k)$ solving FOC.
- Voters' cutoff k trades off selection and sanctioning.

Comparative statics - Setup

General First-Order Condition (FOC):

$$2x = \delta R'(x)\mathbb{E}_\theta[\hat{y}^2]$$

Where:

- $R(x) = 1 - F(k + x^2)$: probability of reelection,
- $R'(x) = -2xf(k + x^2)$,
- δ: discount factor,
- $\mathbb{E}_\theta[\hat{y}^2]$: expected second-period welfare loss.

Effect of cutoff k — Part 1

Goal: Find the sign of $\frac{\partial x^*}{\partial k}$.

- $R'(x)$ depends directly on k, through the composite function $F(k + x^2)$.
- $x = x^*(k)$ also changes with k.

Differentiate both sides of the FOC:

$$\frac{d}{dk}(2x) = \delta \left(\frac{d}{dk} R'(x) \right) \mathbb{E}_\theta[\hat{y}^2]$$

$$2\frac{dx}{dk} = \delta \left(\frac{\partial R'(x)}{\partial x} \cdot \frac{dx}{dk} + \frac{\partial R'(x)}{\partial k} \right) \mathbb{E}_\theta[\hat{y}^2]$$

Why the partial derivative with respect to k?

- $R'(x) = -2xf(k + x^2)$ depends on k *directly* via $f(k + x^2)$.
- Even if x were held fixed, changing k shifts the argument of the pdf.
- So we must account for this when differentiating $R'(x)$.

Effect of cutoff k — Part 2

Recall: $R'(x) = -2xf(k + x^2)$

$$\frac{\partial R'(x)}{\partial x} = -2f(k + x^2) - 4x^2 f'(k + x^2)$$
$$\frac{\partial R'(x)}{\partial k} = -f'(k + x^2)$$

Plug back into the total derivative:

$$2\frac{dx}{dk} = \delta \left(R''(x)\frac{dx}{dk} - f'(k + x^2) \right) \mathbb{E}_\theta[\hat{y}^2]$$
$$\left(2 - \delta R''(x)\mathbb{E}_\theta[\hat{y}^2]\right) \frac{dx}{dk} = -\delta f'(k + x^2)\mathbb{E}_\theta[\hat{y}^2]$$
$$\frac{\partial x^*}{\partial k} = \frac{-\delta f'(k + x^2)\mathbb{E}_\theta[\hat{y}^2]}{2 - \delta R''(x)\mathbb{E}_\theta[\hat{y}^2]}$$

Sign: $f'(\cdot) < 0$ (for symmetric, unimodal f at $k > 0$) $\quad \Rightarrow \quad \frac{\partial x^*}{\partial k} > 0.$

Effect of discount factor δ

Differentiate FOC with respect to δ:

$$2\frac{dx}{d\delta} = \underbrace{R'(x)}_{<0}\underbrace{\mathbb{E}_\theta[\hat{y}^2]}_{>0} + \underbrace{\delta}_{>0}\underbrace{R''(x)}_{\text{typically} <0}\frac{dx}{d\delta}\underbrace{\mathbb{E}_\theta[\hat{y}^2]}_{>0}$$

$$\left(\underbrace{2}_{>0} - \underbrace{\delta R''(x)\mathbb{E}_\theta[\hat{y}^2]}_{\text{typically} <0}\right)\frac{dx}{d\delta} = -\underbrace{R'(x)}_{<0}\underbrace{\mathbb{E}_\theta[\hat{y}^2]}_{>0}$$

$$\frac{\partial x^*}{\partial \delta} = \frac{-R'(x)\mathbb{E}_\theta[\hat{y}^2]}{2 - \delta R''(x)\mathbb{E}_\theta[\hat{y}^2]}$$

Sign:

- Numerator is > 0 (since $-R'(x) > 0$).
- Denominator is > 0 (small enough δ) $\Rightarrow \quad \frac{\partial x^*}{\partial \delta} < 0$

Effect of replacement quality $\mathbb{E}_\theta[\hat{y}^2]$

Differentiate FOC with respect to $\mathbb{E}_\theta[\hat{y}^2]$:

$$
\begin{aligned}
2\frac{dx}{d\mathbb{E}_\theta[\hat{y}^2]} &= \delta\left(R'(x) + R''(x)\frac{dx}{d\mathbb{E}_\theta[\hat{y}^2]}\right) \\
\left(2 - \delta R''(x)\mathbb{E}_\theta[\hat{y}^2]\right)\frac{dx}{d\mathbb{E}_\theta[\hat{y}^2]} &= -\delta R'(x) \\
\frac{\partial x^*}{\partial \mathbb{E}_\theta[\hat{y}^2]} &= \frac{-\delta R'(x)}{2 - \delta R''(x)\mathbb{E}_\theta[\hat{y}^2]}
\end{aligned}
$$

Sign:

- $R'(x) < 0 \quad \Rightarrow \quad \frac{\partial x^*}{\partial \mathbb{E}_\theta[\hat{y}^2]} > 0.$

Summary: Effects on bad types' policy choice

Parameter	**Effect on x^***	**Interpretation**
Voter cutoff k	$\frac{\partial x^*}{\partial k} > 0$	Stricter cutoff improves behaviour
Discount factor δ	$\frac{\partial x^*}{\partial \delta} < 0$	More patience improves behaviour
Replacement quality $\mathbb{E}_\theta[\hat{y}^2]$	$\frac{\partial x^*}{\partial \mathbb{E}_\theta[\hat{y}^2]} > 0$	Worse replacements worsen behaviour

Bottom line: Bad types' effort depends on both electoral standards and future office value.

Lessons from Fearon, 1999 — Intuition

- Elections serve a dual role: **selection** of good types and **sanctioning** of bad behaviour.
- There is a fundamental trade-off: better selection weakens sanctioning incentives, and vice versa.
- Even with noisy outcomes, simple cutoff rules allow voters to control politicians.
- Fearon deepens Ferejohn, 1986's insights by incorporating unobservable types — voters care about *who* politicians are, not just what they do.
- Elections embody a tension between the **mandate** (selection) and **accountability** (sanctioning) conceptions of democratic representation (Manin, Przeworski, and Stokes, 1999).

Bottom line: Elections are powerful but imperfect tools for both disciplining politicians and selecting good leaders.

Limitations of Fearon, 1999 — and the road ahead

Key limitations:

- Static cutoff rules: Voters use fixed thresholds, no dynamic or adaptive standards.
- Simple reaction behaviour: Politicians react mechanically to reelection incentives, no richer signalling or promises.
- No endogenous candidate entry: Who runs for office is taken as given.
- Single-dimensional policy space: Only one policy dimension is modelled.

Extensions:

- **Strategic control rules** (Anesi and Buisseret, 2022): Voters can design more sophisticated reelection thresholds to better induce effort even under uncertainty — relaxing Fearon's trade-off between selection and sanctioning.
- **Aggregation and information**: Further work (e.g., Ashworth and Fowler, 2020, Lupia and McCubbins, 1998) explores how voter aggregation and elite communication shape electoral accountability.

Zooming out briefly: Why is accountability so difficult?

Theoretical assumptions

- **Prospective voting**:
  - → Voters know platforms and understand policy consequences
  - → Politicians seek reelection by keeping promises
  - → Effective selection of good politicians
- **Retrospective voting**:
  - → Voters know about outcomes
  - → Voters attribute responsibility accurately (disentangle exogenous factors from endogenous ones)
  - → Effective sanctioning of bad incumbents

Empirical challenges

- **Prospective voting**:
  - → Blurry platforms or partisan intoxication
  - → Lack of political knowledge or interest
  - → Not always optimal to punish deviations
  - → Inability to assess consequences of platforms
- **Retrospective voting**:
  - → Attribution difficult (e.g., exogenous shocks or partisan intoxication)
  - → Multi-dimensionality: one vote, many issues: "One cannot control many targets with one instrument" (Przeworski, 2018, p. 96)

Illustration #1: Prospective voting and the mandate conception of accountability

> *When the government announces that the conditions are not what they were anticipated to be, voters cannot be sure if implementing the mandate is still in their best interest. And if implementing the mandate is not the best the government can do, the threat of punishing incumbents who deviate from it is not credible (Przeworski, 2018, p. 94)*

Illustration #2: Do people actually have policy preferences, or are they just blind partisans?

Notable challenge: Achen and Bartels, 2017 argue that voters do not really have independent policy preferences; their preferences are entirely endogenous to their social groups → more extreme version: partisan intoxication or cheerleading

FYI: My reading of the empirical evidence is different, but this is a theoretical class

More importantly: Fowler, 2020 argues convincingly that much of the empirical evidence adduced in favour of that hypothesis is observationally equivalent with policy voting (see table on right)

Evidence	Intoxication interpretation	Policy voting interpretation
Party ID predicts vote choice.	Party identification influences vote choices.	Policy preferences influence party identification and vote choices.
Party ID is correlated across generations	People blindly adopt the party of their parents.	Values, economic circumstances, and policy preferences are correlated across generations.
Party ID is stable over time — even more so than issue positions.	People blindly stick with whatever party they initially adopted, even when their policy positions change.	Policy preferences are correlated over time, surveys measure policy positions imperfectly, and party identification reflects a complex weighted average of positions across a range of issues.
Elite cues sometimes influence issue positions.	People don't care about policy and they blindly adopt whatever their party leaders say.	Rational, policy voters should sometimes take cues from elites that share their values and interests.
Most Republicans support President Trump despite his personal shortcomings.	People blindly support leaders from their party, even when it cuts against their personal interests.	Despite Trump's personal shortcomings, many Republicans likely support him because of his policy positions. Furthermore, Republican loyalty to Trump is overstated since people change their party identification in response to their approval of the president.
When the parties changed positions on racial issues, white southerners were slow to change their party identification.	Party identification is sticky and largely unrelated to policy interests.	But they did change their voting behavior. And to the extent that they continued to support Democrats in non-presidential elections, it was only the conservative Democrats who agreed with them on policy.

Illustration #3: Partisan intoxication and retrospective accountability: A topical example?

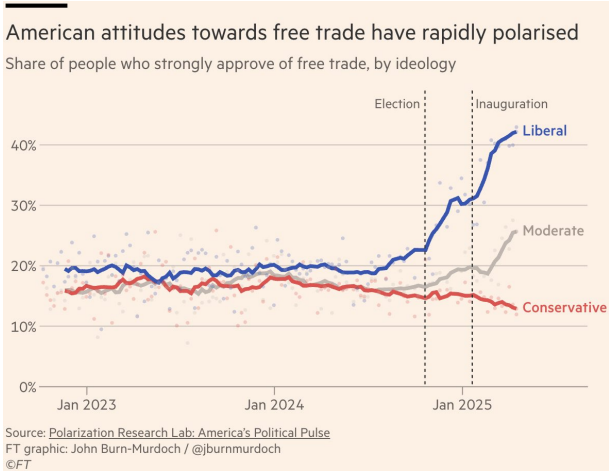


Illustration #4: Exogenous events, attribution of responsibility, and retrospective accountability

Much-discussed empirical challenge to retrospective voting: Exogenous voters – such as shark attacks (Achen and Bartels, 2017) or wins of college football teams – affect vote choice → clear evidence that voters are bad at attribution (Caplan, 2008)?

- Empirically, these findings are very much contested (Fowler and Hall, 2018; Fowler and Montagnes, 2015)

More important point for our purposes: Should we theoretically expect (rational) voters to *not* respond to exogenous events? What events should they respond to, and how? Should we expect voter competence to lead to better accountability?

Answering this question requires theory. Ashworth and Bueno De Mesquita, 2014; Ashworth, Bueno de Mesquita, and Friedenberg, 2018 address this challenge.

Illustration (cont'd)

We won't delve into their argument – this is too complicated for an introductory course like this.

Central intuitions:

1. Some exogenous shocks (e.g., hurricanes) – those that affect voter welfare *and* are related to government performance – provide voters with valuable information about incumbents' types (e.g., their competence). Can foster better selection; rational voters should respond to some shocks.
2. Whether accountability is undermined by voters responding to exogenous events depends on how more or less voter competence affects politicians' incentives. Effect can be non-monotonic: more competence can be bad when it for instance induces pandering

$\rightarrow$ Great examples of how an empirical debate inspires theoretical innovation, which helps us understand better what evidence is necessary to adjudicate between different perspectives

Lupia and McCubbins, 1998: “The Democratic Dilemma“

> *It is widely believed that there is a mismatch between the requirements of democracy and most people’s ability to meet these requirements. If this mismatch is too prevalent, then effective self-governance is impossible. The democratic dilemma is that the people who are called upon to make reasoned choices may not be capable of doing so. (Lupia and McCubbins, 1998, p. 1)*

Remember our starting point: Schumpeter, 1942, Arrow, 1951, and the problem of scale showed that we’d be remiss to mainly ground democratic legitimacy in the consistent aggregation of *all* individual preferences (next week more on that)

A way out: Voters just need to be able to retrospectively sanction politicians; this might be enough to provide decent incentives and keep democracy working.

But what if voters are too stupid or lazy? How can they know enough to navigate the trade-off between sanctioning and selection?

The role of information and trust

- Citizens often lack the time or ability to become fully informed.
- Instead of acquiring full knowledge, they rely on **information shortcuts** (e.g., endorsements, party cues).
- These shortcuts are only effective if the **source is trustworthy**.
- **Trust** becomes central: citizens must assess *who to believe*.
- Lupia and McCubbins, 1998 stress the importance of **credible communication** for informed decision-making.

Formal Model – Setup and intuition

- Lupia and McCubbins adapt a version of the **cheap talk** model from Crawford and Sobel, 1982
- Two actors:
  - → **Advisor (Sender)**: Has private information and potentially biased preferences.
  - → **Citizen (Receiver)**: Must make a decision based on advisor's message.
- Communication is **costless** and **non-binding** — the advisor can lie freely.
- Problem: If preferences diverge, advisor has incentive to mislead.
- Central question: *When is communication credible?*

When can citizens trust “advisors“?

- Trust is possible when:
  - → The advisor’s preferences are **not too divergent** from the citizen’s.
  - → The citizen can **detect deception** or punish dishonesty (directly or indirectly).
  - → The advisor has **incentives to be honest** — e.g., reputation or future interaction.
- Credible communication can emerge even with conflicting interests if these conditions hold.
- **Real-world democratic institutions** support trust:
  - → **Elections** and **public scrutiny** create incentives for truth-telling.
  - → **Media competition** increases the cost of misinformation.
  - → **Oversight bodies** and **transparency laws** can reduce asymmetries and make verification possible.
- Lupia and McCubbins argue these institutions allow citizens to use heuristics **effectively and safely**.

Methodological interlude: Cheap Talk vs Bayesian Persuasion

- **Cheap Talk Models** (Crawford and Sobel, 1982):
  - → Communication is **costless and non-binding**.
  - → Credibility hinges on **preference alignment** between sender and receiver.
  - → No commitment: sender chooses messages freely each time.
  - → Typically leads to **partial information revelation**.
- **Bayesian Persuasion Models** (Kamenica and Gentzkow, 2011; Little, 2023):
  - → Sender **commits ex ante** to an information structure or signalling strategy.
  - → Credibility arises from **commitment**, not necessarily from aligned preferences.
  - → Sender shapes the receiver's beliefs through designed signals, even with preference divergence.
- **Bottom line:** Both models are rational and Bayesian, but differ in assumptions about sender behaviour, commitment, and what enables credible persuasion.

Methodological interlude (cont'd)

- **Little, 2023** identifies five mechanisms enabling persuasion under Bayesian rationality:
  1. **Costly signalling**: Messages are more credible when they are costlier for uninformed senders to mimic.
  2. **Common interests**: Partial alignment between sender and receiver facilitates credible communication.
  3. **Verifiable information**: Messages that can be externally verified enhance credibility.
  4. **Reputation concerns**: Senders may convey truthful information to maintain a reputation for honesty or competence.
  5. **Commitment to messaging strategy**: Known as *Bayesian persuasion*, where senders commit to a strategy that influences receivers' beliefs.
- **Comparison with Crawford and Sobel, 1982**:
  - → Focuses on *cheap talk*—costless, non-binding communication between sender and receiver.
  - → Credibility arises only when sender and receiver have sufficiently aligned preferences.
  - → No mechanisms for commitment or external verification are considered.

Empirical challenge: Heuristic projection

- **Main finding:** Voters often *misinterpret interest group cues* due to **heuristic projection**—they assume endorsed politicians share their own views, even when they do not (Broockman, Kaufman, and Lenz, 2024).
- **Implication:** Cues from interest groups may **fail to improve accountability** and can even increase approval for misaligned representatives.
- **Empirical basis:** Three studies using real IG ratings and participants' actual representatives.
  - → Voters struggle to accurately place IGs ideologically.
  - → IG cues increase misperceptions and overconfidence.
- **Relevance to Lupia and McCubbins, 1998:**
  - → Challenges the assumption that voters can identify *trustworthy* or *aligned* sources.
  - → Suggests that without institutional or informational support, heuristics may **mislead** rather than guide.

When do elite cues matter empirically?

Elite cues do not have uniform effects. Their effects depend on several scope conditions:

1. **Political interest:** Cues are more influential for politically interested individuals (Cavaillé and Neundorf, 2023).
2. **Material interest alignment:** Individuals are less receptive when elite cues contradict their material interests (Cavaillé and Neundorf, 2023).
3. **Issue importance:** Greater receptivity to elite cues on issues people care less about (Barber and Pope, 2024).
4. **Issue salience and the media:** Cues matter more for issues extensively covered by the media.

Is democracy broken? Revisiting the role of voters

Many political scientists claim **voters are irrational**, uninformed, and swayed by irrelevant factors (Achen and Bartels, 2017).

Heuristics and elite cues might help, but the argument is both theoretically and empirically not as strong as we would like to confidently believe in the efficacy of electoral accountability.

This raises a crucial question: Do flawed voters necessarily lead to flawed democratic outcomes?

Voters vs. electorates

- A key distinction: the **individual voter** is not the same as the **electorate**.
- Many voters may be:
  - → Partisan
  - → Poorly informed
  - → Susceptible to bias or heuristics
- Yet elections often yield **reasonable aggregate outcomes**:
  - → Sensitivity to candidate quality and ideology
  - → Responsiveness to government performance
- Ashworth and Fowler, 2020: What matters is not the “typical voter,” but the **pivotal voters** whose behaviour decides outcomes.

Theoretical frameworks: Why electorates can be rational, even if most voters are not

- **Statistical argument (The Macro Polity, Condorcet Jury Theorem):**
  - → If individual voter errors are uncorrelated, they "cancel out" in the aggregate (Erikson, Mackuen, and Stimson, 2008).
  - → Assumes independence — a strong and often unrealistic condition.
- **Strategic pivotal voter argument (Ashworth and Fowler, 2020):**
  - → Electoral outcomes are determined by swing voters or those sensitive to candidate quality.
  - → Politicians have incentives to exert effort even if most voters are unresponsive.
- **Endogenous information acquisition:**
  - → Pivotal voters may **seek out more information**, improving aggregate outcomes.
  - → Supported by Mikosch et al., 2024 and other studies on motivated learning and turnout.

▸ Refresher on Condorcet's Jury Theorem

Implications for democratic theory and empirical research

- **Flawed voters do not imply flawed democracy.**
  - → Aggregate behaviour can exhibit rationality even if individuals do not.
- **Pivotal voters and turnout dynamics** help preserve electoral accountability.
- **Scholarly focus should shift:**
  - → From voter irrationality to electorate-level responsiveness.
  - → From idealised norms to empirical mechanisms of accountability.
- **Caution against overgeneralising**: Poor individual-level knowledge or behaviour $\neq$ democratic failure.

Outline

Course logistics and preliminaries

Introducing the core themes

Four big ideas about electoral accountability

- Ferejohn's model of electoral control
- Fearon's model of sanctioning vs. selection
- Informational shortcuts to the rescue?
- Electorates vs. voters

Takeaways and outlook

Taking stock: What have we learned?

- From ideals to institutions: Democracy's original vision of self-government is infeasible at scale; modern democracy relies on representation and rules-based competition to manage conflicts peacefully.
- The second-best ideal: Self-rule becomes regulated conflict — elections distribute power peacefully, making outcomes tolerable for winners and losers alike.
- Elections and incentives: Formal models show how elections can discipline politicians — but only under the right conditions (stakes, noise, expectations).
- Limits of accountability: Voters face trade-offs between sanctioning and selection. Misattribution, low information, or high stakes can erode incentives.
- Trust and shortcuts: Heuristics and elite cues help voters navigate complexity — if sources are credible and institutions foster trust.
- Why democracy can work anyway: Even if individual voters are flawed, electorates can still behave rationally. What matters most is pivotal voters and the structure of competition.

References I

Acharya, Avidit, Elliot Lipnowski, and João Ramos (2025). “Political accountability under moral hazard”. *American Journal of Political Science* 69.2, pp. 641–652. DOI: 10.1111/ajps.12860.

Achen, Christopher H. and Larry M. Bartels (Aug. 2017). *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton, Princeton University Press.

Alabrese, Eleonora et al. (2024). “Levelling up by levelling down: The economic and political cost of Brexit”. *Working Paper*.

Anesi, Vincent and Peter Buisseret (Nov. 2022). “Making Elections Work: Accountability with Selection and Control”. *American Economic Journal: Microeconomics* 14.4, pp. 616–644. DOI: 10.1257/mic.20200311.

Arrow, Kenneth J. (1951). *Social Choice and Individual Values*. 3rd ed. New Haven: Yale University Press.

References II

Ashworth, Scott and Ethan Bueno De Mesquita (Aug. 2014). "Is Voter Competence Good for Voters?: Information, Rationality, and Democratic Performance". *American Political Science Review* 108.3, pp. 565–587. DOI: 10.1017/S0003055414000264.

Ashworth, Scott, Ethan Bueno de Mesquita, and Amanda Friedenberg (2018). "Learning about Voter Rationality". *American Journal of Political Science* 62.1, pp. 37–54. DOI: 10.1111/ajps.12334.

Ashworth, Scott and Anthony Fowler (Aug. 2020). "Electorates versus Voters". *Journal of Political Institutions and Political Economy* 1.3, pp. 477–505. DOI: 10.1561/113.00000016.

Barber, Michael and Jeremy C. Pope (Apr. 2024). "Does issue importance attenuate partisan cue-taking?" en. *Political Science Research and Methods* 12.2, pp. 435–443. DOI: 10.1017/psrm.2023.28.

Bates, Robert H. and Da-Hsiang Lien (Mar. 1985). "A Note on Taxation, Development, and Representative Government". en. *Politics & Society* 14.1, pp. 53–70. DOI: 10.1177/003232928501400102.

References III

Broockman, David E., Aaron R. Kaufman, and Gabriel S. Lenz (Jan. 2024). "Heuristic Projection: Why Interest Group Cues May Fail to Help Citizens Hold Politicians Accountable". *British Journal of Political Science* 54.1, pp. 69–87. DOI: 10.1017/S0007123423000078.

Burn-Murdoch, John (Mar. 2024). "American politics is undergoing a racial realignment". *Financial Times*.

Caplan, Bryan (Aug. 2008). *The Myth of the Rational Voter: Why Democracies Choose Bad Policies (New Edition): Why Democracies Choose Bad Policies*. English. Princeton, NJ: Princeton University Press.

Carey, John M. and Oscar Pocasangre (2024). *Can Proportional Representation Lead to Better Governance?* Tech. rep. Protect Democracy | New America.

Cavaillé, Charlotte and Anja Neundorf (Dec. 2023). "Elite Cues and Economic Policy Attitudes: The Mediating Role of Economic Hardship". *Political Behavior* 45.4, pp. 1355–1376. DOI: 10.1007/s11109-021-09768-w.

References IV

Chung, Hun and John Duggan (Feb. 2020). "A Formal Theory of Democratic Deliberation". en. *American Political Science Review* 114.1, pp. 14–35. DOI: 10.1017/S0003055419000674.

Crawford, Vincent P. and Joel Sobel (1982). "Strategic Information Transmission". *Econometrica* 50.6, pp. 1431–1451. ISSN: 0012-9682. DOI: 10.2307/1913390.

Dowding, Keith (Jan. 2006). "Can Populism Be Defended? William Riker, Gerry Mackie and the Interpretation of Democracy". en. *Government and Opposition* 41.3, pp. 327–346. DOI: 10.1111/j.1477-7053.2006.00182.x.

Druckman, James N. (Nov. 2004). "Political Preference Formation: Competition, Deliberation, and the (Ir)relevance of Framing Effects". en. *American Political Science Review* 98.4, pp. 671–686. DOI: 10.1017/S0003055404041413.

Drutman, Lee (2021). "Elections, Political Parties, and Multiracial, Multiethnic Democracy: how the United States Gets It Wrong". *New York University Law Review* 96, p. 985.

References V

Dryzek, John S. and Christian List (2003). "Social Choice Theory and Deliberative Democracy: A Reconciliation". *British Journal of Political Science* 33.1, pp. 1–28.

Dunn, John (Nov. 2018). *Setting the People Free: The Story of Democracy, Second Edition*. English. 2nd edition. Princeton: Princeton University Press.

Edenhofer, Jacob (2024). *Is corporatism dirty or clean? Examining the effects of corporatism on climate policy*.

Edenhofer, Jacob and Christian Flachsland (2025). "Introducing the Climate Politics Framework (CPF): An application to German climate policy". *Working Paper (mimeo)*.

Edenhofer, Jacob and Federica Genovese (2024). *When and why compensation can unlock the green energy transition*. Tech. rep. Oxford: Progressive Politics Research Network.

Erikson, Robert S., Michael B. Mackuen, and James A. Stimson (Jan. 2008). *The Macro Polity*. Cambridge: Cambridge University Press.

References VI

Fearon, James D. (1999). “Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance”. *Democracy, Accountability, and Representation*. Ed. by Adam Przeworski, Bernard Manin, and Susan C. Stokes. Cambridge: Cambridge University Press, pp. 55–97.

— **(Nov. 2011).** “Self-Enforcing Democracy”. *The Quarterly Journal of Economics* 126.4, pp. 1661–1708. ISSN: 0033-5533. DOI: 10.1093/qje/qjr038.

Ferejohn, John (Jan. 1986). “Incumbent performance and electoral control”. en. *Public Choice* 50.1, pp. 5–25. ISSN: 1573-7101. DOI: 10.1007/BF00124924.

Fetzer, Thiemo, Jacob Edenhofer, and Prashant Garg (2025). “Local Decline and Populism”. en. *Economic Letters* Forthcoming. ISSN: 1556-5068. DOI: 10.2139/ssrn.4991823.

Fetzer, Thiemo, Callum Shaw, and Jacob Edenhofer (2024). *Informational Boundaries of the State*.

References VII

Fiorina, Morris P. (1978). “Economic Retrospective Voting in American National Elections: A Micro-Analysis”. *American Journal of Political Science* 22.2, pp. 426–443. DOI: 10.2307/2110623.

Fishkin, James (July 2018). *Democracy When the People Are Thinking: Revitalizing Our Politics Through Public Deliberation*. English. Oxford: Oxford University Press.

Fishkin, James, Valentin Bolotnyy, et al. (Nov. 2024). “Can Deliberation Have Lasting Effects?” en. *American Political Science Review* 118.4, pp. 2000–2020. DOI: 10.1017/S0003055423001363.

— **(Jan. 2025).** “Scaling Dialogue for Democracy: Can Automated Deliberation Create More Deliberative Voters?” en. *Perspectives on Politics*, pp. 1–18. DOI: 10.1017/S1537592724001749.

Fishkin, James, Alice Siu, et al. (Nov. 2021). “Is Deliberation an Antidote to Extreme Partisan Polarization? Reflections on “America in One Room””. en. *American Political Science Review* 115.4, pp. 1464–1481. DOI: 10.1017/S0003055421000642.

References VIII

Fowler, Anthony (Apr. 2020). "Partisan Intoxication or Policy Voting?" English. *Quarterly Journal of Political Science* 15.2, pp. 141–179. DOI: 10.1561/100.00018027a.

Fowler, Anthony and Andrew B. Hall (Oct. 2018). "Do Shark Attacks Influence Presidential Elections? Reassessing a Prominent Finding on Voter Competence". *The Journal of Politics* 80.4, pp. 1423–1437. ISSN: 0022-3816. DOI: 10.1086/699244.

Fowler, Anthony and B. Pablo Montagnes (Nov. 2015). "College football, elections, and false-positive results in observational research". *Proceedings of the National Academy of Sciences* 112.45, pp. 13800–13804. DOI: 10.1073/pnas.1502615112.

Gaertner, Wulf (June 2009). *A Primer in Social Choice Theory: Revised Edition*. English. Revised edition. Oxford ; New York: Oxford University Press, USA.

Gerring, John and Wouter Veenendaal (July 2020). *Population and Politics: The Impact of Scale*. English. Cambridge: Cambridge University Press.

References IX

Gieczewski, Germán and Christopher Li (June 2024). “Incumbent Performance and Electoral Control: A Comment”. *Quarterly Journal of Political Science* 19.3, pp. 331–353. DOI: 10.1561/100.00023038.

Goodin, Robert E. and Kai Spiekermann (May 2018). *An Epistemic Theory of Democracy*. English. Oxford: Oxford University Press.

Gratton, Gabriele and Jacob Edenhofer (2025). “The Rise and Fall of Technocratic Democracies: Unstable Majorities and Delegation to Technocrats”. en.

Grofman, Bernard and Scott L. Feld (1988). “Rousseau’s General Will: A Condorcetian Perspective”. *The American Political Science Review* 82.2, pp. 567–576. DOI: 10.2307/1957401.

Hibbing, John R. and Elizabeth Theiss-Morse (2002). *Stealth Democracy: Americans’ Beliefs About How Government Should Work*. Cambridge: Cambridge University Press.

Kamenica, Emir and Matthew Gentzkow (Oct. 2011). “Bayesian Persuasion”. en. *American Economic Review* 101.6, pp. 2590–2615. ISSN: 0002-8282. DOI: 10.1257/aer.101.6.2590.

References X

Konc, Théo et al. (2024). *The German Energy Price Shock and Populism*.

Landa, Dimitri and Adam Meirowitz (2009). "Game Theory, Information, and Deliberative Democracy". *American Journal of Political Science* 53.2, pp. 427–444. DOI: 10.1111/j.1540-5907.2009.00379.x.

Latner, Michael, Jack Santucci, and Matthew Shugart (Aug. 2021). *Multi-seat Districts and Larger Assemblies Produce More Diverse Racial Representation*. DOI: 10.2139/ssrn.3911532.

Li, Yuhui (Sept. 2019). *Dividing the Rulers: How Majority Cycling Saves Democracy*. English. Michigan: University of Michigan Press.

Little, Andrew T (July 2023). "Bayesian explanations for persuasion". en. *Journal of Theoretical Politics* 35.3, pp. 147–181. ISSN: 0951-6298. DOI: 10.1177/09516298231185060.

Lupia, Arthur and Mathew D. McCubbins (1998). *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge: Cambridge University Press.

References XI

Mainwaring, Scott and Lee Drutman (2023). *The Case for Multiparty Presidentialism in the US*. Tech. rep. Protect Democracy | New America.

Manin, Bernard, Adam Przeworski, and Susan C. Stokes (1999). "Elections and Representation". *Democracy, Accountability, and Representation*. Ed. by Adam Przeworski, Susan C. Stokes, and Bernard Manin. Cambridge: Cambridge University Press, pp. 29–54.

Maskin, Eric (Mar. 2020). "A modified version of Arrow's IIA condition". en. *Social Choice and Welfare* 54.2-3, pp. 203–209. DOI: 10.1007/s00355-020-01241-7.

— **(Feb. 2025).** "Borda's Rule and Arrow's Independence Condition". *Journal of Political Economy* 133.2, pp. 385–420. DOI: 10.1086/732892.

Mikosch, Heiner et al. (Apr. 2024). "Uncertainty and Information Acquisition: Evidence from Firms and Households". *American Economic Journal: Macroeconomics* 16.2, pp. 375–405. DOI: 10.1257/mac.20220047.

References XII

Niemeyer, Simon et al. (Feb. 2024). “How Deliberation Happens: Enabling Deliberative Reason”. en. *American Political Science Review* 118.1, pp. 345–362. DOI: 10.1017/S0003055423000023.

Przeworski, Adam (1991). *Democracy and the Market: Political and Economic Reforms in Eastern Europe and Latin America*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139172493.

— **(2010).** *Democracy and the Limits of Self-Government*. Englisch. Cambridge: Cambridge University Press.

— **(Jan. 2018).** *Why Bother With Elections?* Englisch. Cambridge, UK: Polity.

Przeworski, Adam, Gonzalo Rivero, and Tianyang Xi (Sept. 2015). “Elections as a conflict processing mechanism”. *European Journal of Political Economy* 39, pp. 235–248. DOI: 10.1016/j.ejpoleco.2015.05.006.

Riker, William H. (1982). *Liberalism Against Populism: A Confrontation Between the Theory of Democracy and the Theory of Social Choice*. English. Prospect Heights, Ill: Waveland.

References XIII

Ruffini, Patrick (Nov. 2023). *Party of the People: Inside the Multiracial Populist Coalition Remaking the GOP*. English. New York: Simon & Schuster.

Santucci, Jack (Aug. 2022). *More Parties or No Parties: The Politics of Electoral Reform in America*. English. Oxford: Oxford University Press.

Schumpeter, Joseph A. (1942). *Capitalism, Socialism, and Democracy: Third Edition*. Englisch. 3rd Edition. New York: Harper Perennial Modern Classics.

Sen, Amartya (2017). *Collective Choice and Social Welfare*. Expanded Edition. London: Penguin.

Stasavage, David (June 2020). *The Decline and Rise of Democracy: A Global History from Antiquity to Today*. Princeton (N.J.): Princeton University Press.

Sunstein, Cass R. (2000). "Deliberative Trouble? Why Groups Go to Extremes". *The Yale Law Journal* 110.1, pp. 71–119. DOI: 10.2307/797587.

References XIV

Sunstein, Cass R. (2002). “The Law of Group Polarization”. en. *Journal of Political Philosophy* 10.2, pp. 175–195. DOI: 10.1111/1467-9760.00148.

Veri, Francesco (Feb. 2025). “Fostering Reasoning in the Politically Disengaged: The Role of Deliberative Minipublics”. *Political Studies Review* 23.1, pp. 254–275. DOI: 10.1177/14789299241239909.

Weale, Albert (1984). “Social Choice versus Populism? An Interpretation of Riker’s Political Theory”. *British Journal of Political Science* 14.3, pp. 369–385.

Wilson, Robert (Dec. 1972). “Social choice theory without the Pareto Principle”. *Journal of Economic Theory* 5.3, pp. 478–486. DOI: 10.1016/0022-0531(72)90051-8.

Zwar, Claudia, Jacob Edenhofer, and Christian Flachsland (2024). “Introducing and applying the Climate Institutions Framework (CIAF)”. *Working Paper*.

Outline

Sharpening Rousseau's ideal

Condorcet's paradox and Arrow's (im)possibility theorem

What implications, if any, do these results have for democracy?

Comparative statics of our Przeworski-type model

Grofman and Feld, 1988 on Rousseau's *Volonté Générale*

Three central elements of Rousseau's notion of the common good:

- There exists a common good independent of individual wills.
- Citizens have imperfect individual judgement about the common good.
- Collective decision-making by majority vote is the most reliable method for approximating the common good.

Connection to Condorcet's Jury Theorem:

- If individuals vote sincerely and independently, majority rule converges to correct decisions as group size increases.

A refresher on Condorcet's jury theorem

Setup:

- There is a binary decision: correct (C) or incorrect (I).
- Each voter i votes independently with probability $p > 0.5$ of choosing C correctly.
- Majority rule is used to determine the collective decision.

Theorem (Condorcet):

As the number of voters $n \to \infty$,

$$\mathbb{P}(\text{Majority decision is correct}) \to 1.$$

Interpretation:

- Under independence and individual "competence" $p > 0.5$, majority voting is a highly reliable aggregator of individual judgements.

A refresher on Condorcet's jury theorem (cont'd)

Key intuition:

- Each voter has a better-than-random chance of voting correctly ($p > 0.5$).
- Errors cancel out because voters vote independently.
- The "signal" (correct votes) dominates the "noise" (errors) as the group size grows.

Why majority rule "works":

- "Law of large numbers": With more voters, the proportion voting correctly converges to p.
- Since $p > 0.5$, the correct option increasingly wins the majority.

$\rightarrow$ See Goodin and Spiekermann, 2018 for various extensions and extensive discussion of the theorem's relevance for democratic theory

Conclusions by Grofman and Feld, 1988

Fragility of Rousseau's argument:

- Factions undermine independence by introducing correlated errors.
- Self-interest undermines sincerity, shifting citizens away from voting for the common good.

Implication:

- Rousseau's confidence in collective decision-making depends critically on citizens' sincere, independent pursuit of the common good.

Implication for scale:

- Scale and heterogeneity can undermine the conditions (independence, sincerity) needed for reliable aggregation.

Outline

Sharpening Rousseau's ideal

Condorcet's paradox and Arrow's (im)possibility theorem

What implications, if any, do these results have for democracy?

Comparative statics of our Przeworski-type model

The difficulty of collective choice with heterogenous preferences

Problem: How to aggregate individual preferences into a collective decision?

Warning sign:

→ Condorcet's paradox (1785):

Even with rational individuals, collective preferences can be intransitive.

Key move:

→ Arrow, 1951 generalises Condorcet's insight to all voting rules.

Next: Understand Condorcet's paradox.

Condorcet's paradox

Condorcet's insight (1785):

$\rightarrow$ Majority rule with pairwise comparisons can yield intransitive collective preferences.

Setup:

- Three individuals: A, B, C
- Three alternatives: x_1, x_2, x_3
- Each individual has transitive preferences.

Observation:

$\rightarrow$ Majority preference between alternatives can cycle, even if individuals are rational.

Next: See a concrete example.

Condorcet cycle: An example

Three individuals and their preferences:

Voter	1st choice	2nd choice	3rd choice
A	x_1	x_2	x_3
B	x_2	x_3	x_1
C	x_3	x_1	x_2

Majority preferences in pairwise comparisons:

- $x_1 \succ x_2$ (A and C prefer x_1 over x_2)
- $x_2 \succ x_3$ (A and B prefer x_2 over x_3)
- $x_3 \succ x_1$ (B and C prefer x_3 over x_1)

Conclusion: Cycle arises: $x_1 \succ x_2 \succ x_3 \succ x_1$.

Primitives of collective choice: Individuals and preferences

Individuals:

- $N = \{1, 2, \ldots, n\}$ is the finite, non-empty set of individuals.
- Cardinality: $\#N \geq 2$.

Alternatives:

- X denotes the finite, non-empty set of alternatives.
- Cardinality: $\#X \geq 3$.

Individual preferences:

- Each individual $i \in N$ has a preference relation R_i over X.
- $R_i \in R(X)$, where $R(X)$ is the set of weak orders (complete, reflexive, transitive relations).

Primitives of collective choice: Profiles and Aggregation Rules

Preference profiles:

- $R^N(X)$ is the set of all n-tuples of individual preferences.
- An element $\langle R_i \rangle_{i \in N}$ is called a preference profile.

Preference aggregation rule (PAR):

- A PAR is a mapping: $f : R^N(X) \to B(X)$.
- $B(X)$ = set of complete and reflexive (but not necessarily transitive) binary relations over X.

Goal:

$\to$ Use a PAR to aggregate individual preferences into a collective social relation.

Arrow's *minimal* conditions: Setting the stage

Goal: Formalise minimal, "reasonable" conditions for collective decision-making.

Conditions Arrow imposes on preference aggregation rules (PARs):

- Ensure collective preferences are coherent and respect individual inputs.
- Prevent trivial or dictatorial outcomes.

Warning: Even these seemingly innocuous conditions have profound consequences.

Next: Let's look at the conditions in some more detail.

Arrow's conditions: Overview

Condition	Meaning	Examples ruled out
Unanimity	If everyone in society prefers x to y, then society must prefer x to y.	Any non-Paretian SWF
Universal domain	The SWF must be able to generate a social preference for all logically possible rational profiles.	Any domain-restricted SWF (e.g., weak Pareto)
Transitivity	If society prefers x to y and y to z, it must prefer x to z.	Majority rule
Independence of Irrelevant Alternatives (IIA)	Social preference between x and y should depend only on individuals' preferences between x and y.	Borda count
Weak Pareto	If everyone strictly prefers x to y, society must strictly prefer x to y.	Inverse dictatorship + null aggregation (Wilson, 1972)
Non-dictatorship	No single individual's preferences always dictate social preferences.	Dictatorial aggregation

Arrow's Impossibility Theorem

Theorem (Arrow, 1951)

For a finite set of individuals ($\#N \geq 2$) and three or more alternatives ($\#X \geq 3$), any preference aggregation rule satisfying:

- *Unrestricted Domain,*
- *Weak Pareto,*
- *Independence of Irrelevant Alternatives (IIA),*
- *and Transitivity,*

must be dictatorial.

→ See Sen, 2017 or Gaertner, 2009 for proofs.

Profound implication:

→ Extremely difficult to aggregate individual preferences consistently.

Outline

Sharpening Rousseau's ideal

Condorcet's paradox and Arrow's (im)possibility theorem

What implications, if any, do these results have for democracy?

Comparative statics of our Przeworski-type model

Riker's interpretation of Arrow

Riker, 1982: *Liberalism Against Populism*

- Arrow's theorem shows that populist democracy—democracy as pure preference aggregation—is arbitrary.
- Any non-dictatorial preference aggregation rule (PAR) satisfying weak Pareto and IIA must be intransitive.
- Thus, social preferences can change arbitrarily even when individual preferences stay the same.
- Preference aggregation alone cannot (normatively) ground democratic legitimacy Weale, 1984.

Problems with Riker's argument

- Arrow's theorem shows only the possibility of preference cycles—not that they necessarily occur in practice (Dowding, 2006). We also don't know how likely they are to arise.
- Riker assumes that electoral accountability guarantees less arbitrariness, but this is not demonstrated.
- Electoral mechanisms require demanding conditions to reliably improve outcomes (Ferejohn, 1986) Start of summary

Deliberative democracy: A response to Riker?

Main Idea:

- Arrow's theorem hinges crucially on the unrestricted domain — that is, any pattern of individual preferences must be admissible.
- Deliberation can induce *single-peaked preferences* (Dryzek and List, 2003).
  - → It helps people identify the unique option that is best for them (or establish that they are indifferent between all options).
  - → Rests on the idea that non-single-peaked preferences are due to a lack of information that can be remedied via deliberation
- **Alternative approaches:** Other theorists (e.g., Maskin, 2020, 2025) propose relaxing Arrow's IIA condition instead of restricting domains.

What is single-peakedness?

Definition:

- A voter's preference is single-peaked if alternatives can be ordered along a line (e.g., left–right), such that:
  - → Each voter has one most-preferred alternative p.
  - → Moving away from p in either direction makes outcomes worse.

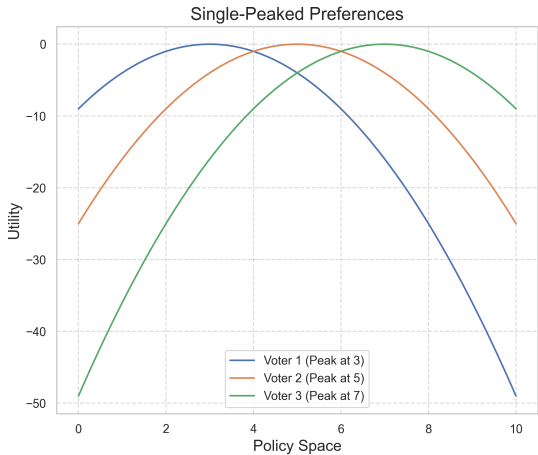
Formally:

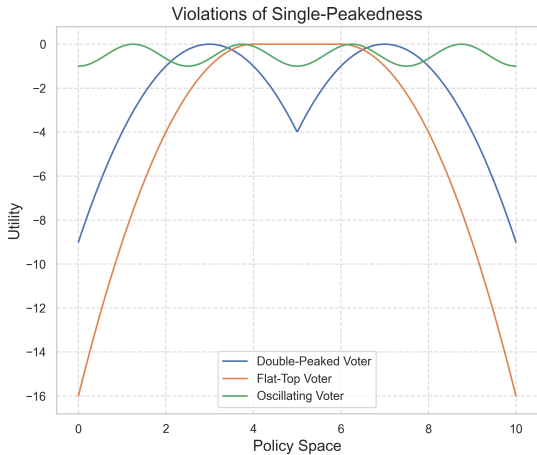
- For all alternatives p'', p', p with $p'' \leq p' \leq p$ or $p \leq p' \leq p''$, we have:

$$U(p'') \leq U(p') \quad \text{and} \quad U(p') \leq U(p)$$

where $U(\cdot)$ denotes the voter's utility.

Illustrating (non-)single-peaked preferences





Back to Fearon

Why single-peakedness matters: Black's theorem(s)

Black's Possibility Theorem (1958):

- If all voters have single-peaked preferences over a one-dimensional policy space,
- And the number of voters is odd,
- Then pairwise majority rule produces a transitive social preference ordering.

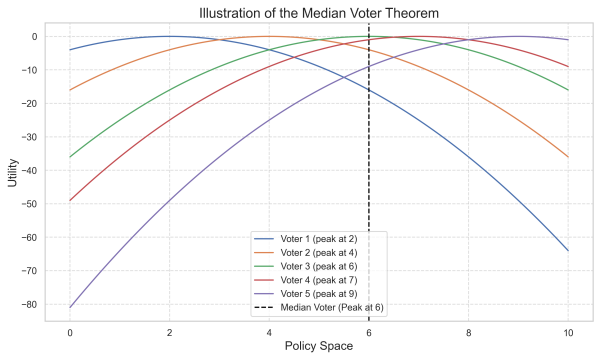
Black's Median Voter Theorem (refinement):

- The median voter's ideal point is a Condorcet winner:
- It defeats every other alternative in pairwise majority votes.

Implications for democracy:

- No cycles (no Condorcet paradox),
- Stable and non-arbitrary majority rule outcomes.

Black's Median Voter Theorem: Visual explanation



- Voters' ideal points are ordered: $2, 4, 6, 7, 9$.
- Median voter = voter at 6.
- Pairwise comparisons:
  - $\rightarrow$ Against 4: Majority (6, 7, 9) prefer 6.
  - $\rightarrow$ Against 7: Majority (2, 4, 6) prefer 6.
  - $\rightarrow$ Against 2 or 9: Majority still prefer 6.
- Thus, the median wins all pairwise majority votes.

The deliberative response: Elegant, but unrealistic?

Main Idea:

- While deliberative democracy offers an elegant response to Arrow and Riker, it faces two major challenges:
  1. **Theoretical challenges**: Strategic behaviour, agenda-setting, and institutional fragility.
  2. **Empirical challenges**: Evidence of group polarisation, biased updating, and failures of deliberation in practice.

Next Steps:

- We first examine theoretical concerns, and then turn to the empirical ones.

Theoretical challenge: Strategic incentives in deliberation

Key insights from Landa and Meirowitz, 2009:

- Deliberation is often strategic, not purely truth-seeking.
- Participants may mislead, withhold information, or selectively disclose to influence outcomes.

Why?

- Non-common values: Participants often pursue different, conflicting goals.
- Cheap talk: Communication is often unverifiable and costless, encouraging manipulation.
- Costly verification: Even when evidence exists, verifying it is expensive and may not happen.

Conclusion: Without careful institutional design, deliberation may **amplify strategic manipulation** rather than promote consensus.

Theoretical challenge: Instability and path dependence

Key insight from Chung and Duggan, 2020:

- Different modes of deliberation (discussion formats) produce very different outcomes.
- Only under strong and idealised conditions does deliberation converge predictably.

Main findings:

- **Myopic discussion**:
  → Unstructured; easily leads to cycling and indeterminate outcomes.
- **Constructive discussion**:
  → Guarantees convergence, but outcomes are path-dependent.
- **Debate (strategic deliberation)**:
  → Can yield a unique compromise, but only under strong strategic rationality assumptions.

Conclusion: Deliberation does not automatically stabilise outcomes.

Does democratic deliberation “work”? Two perspectives

Pro-Deliberation (Optimistic)

- Structured deliberation improves political knowledge, reduces polarisation, and can also durably boost participation (Fishkin, 2018; Fishkin, Bolotnyy, et al., 2024, 2025; Fishkin, Siu, et al., 2021).
- Deliberative forums can improve the “quality“ of reasoning (Niemeyer et al., 2024; Veri, 2025)

Contra-Deliberation (Sceptical)

- Deliberation can increase group polarisation (Sunstein, 2000, 2002).
- Citizens prefer leaders to “get things done” without debate (Hibbing and Theiss-Morse, 2002).
- Deliberation vulnerable to framing and manipulation (Druckman, 2004).

Conclusion: Deliberation can work — but only under **carefully designed, small-scale** conditions. Scaling it up remains difficult.

Deliberation and the challenge of democratic legitimacy

Key takeaways:

- Deliberation offers a possible way to circumvent Arrow's impossibility by restricting the domain of preferences.
- However, theoretical and empirical evidence shows that deliberation is fragile.
- Without ideal conditions, deliberation is often fails to stabilise preferences or mitigate arbitrariness.

Bottom line:

- Deliberation **can** enrich democracy — but it cannot "solve" the problems Arrow and Riker highlight.
- Robust democracy requires **institutions that manage conflict and strategic behaviour** $\rightarrow$ problem of scale remains unresolved

Outline

Sharpening Rousseau's ideal

Condorcet's paradox and Arrow's (im)possibility theorem

What implications, if any, do these results have for democracy?

Comparative statics of our Przeworski-type model

The comparative statics of the compliance condition illustrated

Threshold for compliance:

$$R^* \geq \frac{\pi L - L + k - c(1-\delta)}{\pi - p} \quad \text{(if } \pi > p\text{)}$$

Let $S = R - L$, so $R = S + L$, and rewrite (using the equality case):

$$S^* = \frac{\pi(S^*)L - L + k - c(1-\delta)}{\pi(S^*) - p}$$

Implicit function form: Define

$$F(S, c) := \frac{\pi(S)L - L + k - c(1-\delta)}{\pi(S) - p} - S \quad \Rightarrow \quad F(S^*(c), c) = 0$$

Comparative statics (cont'd)

By the implicit function theorem:

$$\frac{dS^*}{dc} = -\frac{\partial F/\partial c}{\partial F/\partial S} = \frac{(1-\delta)/(\pi - p)}{\partial F/\partial S}$$

Where:

$$\frac{\partial F}{\partial S} = \frac{\pi'(S)L(\pi(S)-p) - \pi'(S)(\pi(S)L - L + k - c(1-\delta))}{(\pi(S)-p)^2} - 1$$

Note: The sign of $\frac{dS^*}{dc}$ depends on both $\pi'(S^*)$ and $\pi(S^*) - p$; it cannot be signed by π' alone.

Comparative statics: Effect of fighting cost c

Threshold condition:

$$S^* = \frac{\pi(S^*)L - L + k - c(1-\delta)}{\pi(S^*) - p}$$

Define: $F(S, c) = \frac{\pi(S)L - L + k - c(1-\delta)}{\pi(S) - p} - S \quad \Rightarrow \quad F(S^*(c), c) = 0$

Implicit function theorem:

$$\frac{dS^*}{dc} = -\frac{\partial F/\partial c}{\partial F/\partial S} = \frac{(1-\delta)/(\pi - p)}{\partial F/\partial S}$$

Denominator term:

$$\frac{\partial F}{\partial S} = \frac{\pi'(S)L(\pi(S) - p) - \pi'(S)\left[\pi(S)L - L + k - c(1-\delta)\right]}{(\pi(S) - p)^2} - 1$$

Comparative statics: Effect of fighting cost c – cont'd

Key implication:

- The numerator of $\frac{dS^*}{dc}$ is positive if $\pi(S^*) > p$,
- But the *sign of the derivative as a whole* depends on $\pi'(S^*)$,
- Since $\pi(S)$ is hump-shaped, $\pi'(S^*)$ may be positive or negative.

Conclusion: The effect of c on compliance stakes S^* is ambiguous and depends on where S^* lies on the $\pi(S)$ curve — increasing c may raise or lower the threshold.

Comparative statics: Effect of political engagement cost k

Threshold condition:

$$S^* = \frac{\pi(S^*)L - L + k - c(1-\delta)}{\pi(S^*) - p}$$

Define:

$$F(S,k) = \frac{\pi(S)L - L + k - c(1-\delta)}{\pi(S) - p} - S \quad \Rightarrow \quad F(S^*(k),k) = 0$$

Implicit function theorem:

$$\frac{dS^*}{dk} = -\frac{\partial F/\partial k}{\partial F/\partial S} = \frac{1/(\pi(S^*) - p)}{\partial F/\partial S}$$

Comparative statics: Effect of political engagement cost k – cont'd

Denominator (as before):

$$\frac{\partial F}{\partial S} = \frac{\pi'(S)L(\pi(S) - p) - \pi'(S)\left[\pi(S)L - L + k - c(1-\delta)\right]}{(\pi(S) - p)^2} - 1$$

Key implication:

- The numerator is strictly positive when $\pi(S^*) > p$,
- The sign of the total derivative again depends on the denominator, which involves $\pi'(S^*)$,
- As with c, the hump-shaped nature of $\pi(S)$ means that the effect of k on S^* is also potentially non-monotonic.

Conclusion: Raising the cost of political engagement generally increases S^*, but the size (and possibly the sign) of the effect depends on where S^* lies along the $\pi(S)$ curve.

Comparative statics: Effect of patience δ

Define:

$$F(S,\delta) = \frac{\pi(S)L - L + k - c(1-\delta)}{\pi(S) - p} - S \quad \Rightarrow \quad F(S^*(\delta),\delta) = 0$$

Implicit function theorem:

$$\frac{dS^*}{d\delta} = -\frac{\partial F/\partial \delta}{\partial F/\partial S} = \frac{c/(\pi(S^*) - p)}{\partial F/\partial S}$$

Denominator: same as in earlier slides — depends on $\pi'(S^*)$

Key implication:

- Numerator is positive if $\pi(S^*) > p$ and $c > 0$; denominator again depends on $\pi'(S^*)$ due to the implicit dependence of the threshold on stakes

Comparative statics: Effect of violent success probability p

Define:

$$F(S,p) = \frac{\pi(S)L - L + k - c(1-\delta)}{\pi(S) - p} - S \quad \Rightarrow \quad F(S^*(p), p) = 0$$

Implicit function theorem:

$$\frac{dS^*}{dp} = -\frac{\partial F/\partial p}{\partial F/\partial S}$$

Compute:

$$\frac{\partial F}{\partial p} = \frac{\pi(S)L - L + k - c(1-\delta)}{(\pi(S) - p)^2} > 0 \quad \text{(under regular conditions)}$$

Comparative statics: Effect of violent success probability p – cont'd

Key implication:

- Numerator is positive when the numerator of the threshold expression is positive (i.e. when compliance is attractive),
- Therefore, $\frac{dS^*}{dp} < 0$ if denominator $\partial F/\partial S$ is positive (standard case),
- That is, greater likelihood of violent success reduces the electoral stakes needed to prefer fighting.

Conclusion: As p increases, compliance becomes harder to sustain. Political actors need increasingly high electoral stakes S^* to prefer democratic participation over rebellion.