Political drift

Luis Luarte

Universidad del Desarrollo, DCCS

November 16, 2025

1 Conceptual framework

Black [1] proposed a single-dimensional policy space, that is, a space going from left to right political ideologies. Its primary justification, in conjunction with Downs [4], was that politics is indeed highly complex; however, there's a significant information cost for voters, leading to the use of a 'good-enough' heuristic, which reduces dimensionality to this 'left-right' dimension, having also found empirical support for this as a 'natural' order. Further, Black [1] proposes a single-peaked utility function, which makes intuitive sense as any politician should have an ideal point to aim towards, thereby situating all political agents in the same rational space.

My work is basically in line with the main assumptions of Black [1] and Downs [4], starting from the point that rational agents take decisions from an ideal or preferred point in a policy space, which seems like a good-enough abstraction of the immense complexity of politics. The exception is that I take these assumptions and set them for a specific time point, relaxing the belief that, from time t to time $t+1$, the system must remain the same. Thus, I do not assume a fixed preference for the agents; instead, a mechanism for modifying preferences from t to $t+1$ is defined, while respecting the classical assumption within any given t. Ideal points are thus dynamic ideal points, following a more modern tradition of treating these points as time-varying parameters [5], rather than fixed attributes. To justify this from the perspective of the agent's composition, I assume a 'bounded rationality' framework, where internal beliefs are updated based on external feedback [6], in conjunction with classical intuitions on how intelligent organisms learn based on their interaction with the environment [7].

As the logical next step I needed some assumptions of the environment in which the agents moves, again departing from Black's where the environment can be reduced to the proposal that the agents vote, here proposal potential value is in interaction with agent position in the political space and the macro-political

1

space, to capture the intuition that no given proposal is the same depending on the ruling party. To capture this, I use techniques drawn from agent-based modeling and computational social science [2], where, instead of a top-down statistical model, I propose a bottom-up Newtonian system with agents as particles, each party being a center of mass exerting a gravitational pull. This approach permits capturing emergent phenomena like 'party lock-in' as the result of micro-level deterministic rules.

The final part of my conceptual framework is the strategic vote, which I don't see as a complete departure from Black's sincere voting, as within any given t while the political space is 1-dimensional the agents utility is multi-dimensional and dependent on the history of decision, an according to that its vote is always sincere within the complex political dynamics. The idea stems from thinking of parties as organizations that aim to solve the collective action problem by using incentives for their members [3], rewarding loyalty and punishing defection. Here, the agent is in a tension between its own ideology and the party's organizational power; thus, the agent is not just a reflection of its ideology but also how it manages it in the personal-versus-party dimension.

2   Preamble

I start by assuming that politicians 'drift' from their original ideological positions. I propose that there's (1) an **environment** in which they drift, and (2) a **mechanism** that determines how and why they'll 'drift'.

This model defines the **environment** as a one-dimensional 'left-right' policy space. This space is not empty; political parties dominate it and act as 'centers of mass' or 'gravitational bodies' that exert a constant, attractive pull on every legislator.

The **mechanism** of this drift is two-fold: it addresses both the 'why' (the strategic vote) and the 'how' (the positional update).

First, the 'why'. I reject the idea that a vote is a simple, sincere expression of ideology. Instead, I propose a strategic utility function where an agent's vote is a calculated trade-off between belief and loyalty.

This utility is a weighted average of:

- Ideological payoff (w): the pure, personal satisfaction of voting for a policy that is close to one's own ideal beliefs.

- Portfolio Payoff ($1 - w$): This refers to the tangible (though possibly not widely recognized) K-bonus (or penalty) associated with voting in line with

one's assigned political role, whether as an 'Officialist' or an 'Opposition' member. I use the term "portfolio" to convey the concept of a rational, macro-oriented approach to ideological decision-making.

Second, the 'how'. A legislator's ideal point $p_i(t)$ is in constant motion, influenced by a 'political force field'. The drift is then the sum of two distinct forces:

- A learning force (ΔP_{learn}) occurs when political reality delivers a 'jolt.' When a legislator loses a vote, they must 'learn' from the winning position, which shifts their ideal point.

- A gravitational force (ΔP_{force}): This represents the Newtonian attraction exerted by all parties involved. I conceptualize this as a weighted gravitational field, where an agent's own party (the one to which they feel closest) exerts an amplified pull (θ_{self}). Creating a lock-in effect that draws agents toward the nearest cluster. The underlying idea is that while politicians may belong to a particular political party, this affiliation does not prevent them from acting in favor of another party.

The drift, therefore, is the dynamic path a rational agent follows as it navigates a field of competing incentives.

The significance of this model lies in its push-and-pull framework, in which political agents (legislators) and political parties operate within the ideological spectrum in parallel. It emphasizes clarity by establishing party positions as the average views of their constituents. The model formalizes typical micro-dynamics, specifically the interactions between legislators and political parties, while also introducing a macro perspective. This macro layer illustrates how the governing party can shape the legislature's overall ideological landscape.

3 Formalization

3.1 Actors, environment and information

Agents (L, J): a set of n legislators, $L = \{1, 2, \ldots, n\}$. A set of m political parties, $J = \{1, 2, \ldots, m\}$. Each legislator $i \in L$ is a member of one party $j \in J$.

Policy space (X): the policy space X is a single dimension representing the 'left-right' ideological spectrum. All positions are scalars: $X \subseteq \mathbb{R}$

Information structure: agents know their own *ideology* $p_i(t)$ (which drifts) and their *most likely political affiliation C_{target}*, which is the nearest political

party in the policy space. There exists a global state variable $G(t) \in J$, which is the party of the ruler at time t. Then the relational agent state $S_i(t)$ based on $G(t)$ is:

$$S_i(t) = \begin{cases} \text{Officialist} & \text{if } C_j(t) \in G(t) \\ \text{Opposition} & \text{if } C_j(t) \notin G(t) \end{cases} \tag{1}$$

Agents know the policy space of all proposals, $x_P(t)$, and the status quo, $x_{SQ}(t)$ (the policy space resulting from the project being rejected).

3.2 Rules of the game

Algorithm 1 Policy drift game loop

1: **for** $t = 1$ **to** T **do**
2:     {1. Observe Global States}
3:     Observe Ruler's party $G(t)$
4:     Observe Agenda: proposal $x_P(t)$, status quo $x_{SQ}(t)$, voting rule $\tau(t)$ {*voting rule refers strictly to the quorum required for approval*}
5:     {2. Compute policy space}
6:     **for** each party $j \in J$ **do**
7:         $C_j(t) \leftarrow \frac{1}{|J|} \sum_{i \in j} p_i(t)$ {Calculate party centers}
8:     **end for**
9:     **for** each agent $i \in L$ **do**
10:        $S_i(t) \leftarrow \text{determineRole}(J_i, G(t))$ {Officialist or Opposition}
11:        $j_{\text{self}}(t) \leftarrow \underset{j \in J}{\arg\min}|C_j(t) - p_i(t)|$ {Find closest party}
12:     **end for**
13:     {3. Voting}
14:     **for** each agent $i \in L$ **do**
15:        Calculate $U_i(\text{Yea})$ and $U_i(\text{Nay})$ {Using w, K, and $S_i(t)$}
16:        $a_i(t) \leftarrow \underset{a \in \{\text{Yea, Nay}\}}{\arg\max} U_i(a)$ {Cast vote to max utility}
17:     **end for**
18:     {4. Outcome}
19:     Tally all votes $a_i(t)$
20:     $O(t) \leftarrow \text{determineOutcome}(a_i(t), \tau(t))$ {Pass or Fail}
21:     {5. Drift}
22:     **for** each agent $i \in L$ **do**
23:        Compute $\Delta p_{\text{learn}}(t)$ (based on $O(t)$ vs $a_i(t)$)
24:        Compute $\Delta p_{\text{force}}(t)$ (based on all $C_j(t)$, k_i, θ_j, η)
25:        $p_i(t+1) \leftarrow p_i(t) + \Delta p_{\text{learn}}(t) + \Delta p_{\text{force}}(t) + \epsilon_i(t)$ {Set ideology for next iteration}
26:     **end for**
27: **end for**

3.3 Utility function

The agent's utility is a weighted sum of ideological satisfaction and a relational loyalty bonus (K).

General utility equation:

$$U_i(\text{Action}) = \underbrace{w \cdot (\text{Ideological payoff})}_{\text{Personal beliefs}} + \underbrace{(1-w) \cdot (\text{Portfolio payoff})}_{\text{Party loyalty}} \qquad (2)$$

Where $w \in [0,1]$ is the weight the agent places on pure ideology.

With ideological payoff defined as:

$$U_{\text{Ideological}}(\text{Yea}) = -(p_i(t) - x_P(t))^2 \qquad (3)$$

$$U_{\text{Ideological}}(\text{Nay}) = -(p_i(t) - x_{SQ}(t))^2 \qquad (4)$$

The portfolio payoff is a discrete bonus or penalty K based on the agent's relational state $(S_t(t))$. We assume that the government's intent is always in favor of its proposal.

If an agent is **Officialist**, the utility of voting "favor" is:

$$U_i(\text{Yea}) = w \cdot (-(p_i(t) - x_P)^2) + (1-w) \cdot (K) \qquad (5)$$

and the utility of voting "against":

$$U_i(\text{Nay}) = w \cdot (-(p_i(t) - x_{SQ})^2) + (1-w) \cdot (-K) \qquad (6)$$

If the agent is **Opposition**, the utility of voting "favor" is:

$$U_i(\text{Yea}) = w \cdot (-(p_i(t) - x_P)^2) + (1-w) \cdot (-K) \qquad (7)$$

and the utility of voting "against":

$$U_i(\text{Nay}) = w \cdot (-(p_i(t) - x_{SQ})^2) + (1-w) \cdot (K) \qquad (8)$$

Then the complete formulation, including all the components of the utility function, is:

$$U_i(a, S_i(t)) = \begin{cases} w \cdot (-(p_i(t) - x_P)^2) + (1-w) \cdot K & \text{if } S_i(t) = \text{Officialist and } a = \text{Yea} \\\\ w \cdot (-(p_i(t) - x_{SQ})^2) + (1-w) \cdot (-K) & \text{if } S_i(t) = \text{Officialist and } a = \text{Nay} \\\\ w \cdot (-(p_i(t) - x_P)^2) + (1-w) \cdot (-K) & \text{if } S_i(t) = \text{Opposition and } a = \text{Yea} \\\\ w \cdot (-(p_i(t) - x_{SQ})^2) + (1-w) \cdot K & \text{if } S_i(t) = \text{Opposition and } a = \text{Nay} \end{cases}$$
$$(9)$$

3.4 Dynamics

Ideology $p_i(t)$ drifts according to:

$$p_i(t+1) = p_i(t) + \Delta p_{learn}(t) + \Delta p_{force}(t) \tag{10}$$

Where Δp_{learn} is the teaching signal that guides the drift according to personal satisfaction:

$$\Delta p_{learn}(t) = \alpha_i(t) \cdot (x_{winner}(t) - p_i(t)) \tag{11}$$

Where $x_{winner}(t)$ is the scalar value of the winning policy, and $\alpha_i(t) = \alpha_{loss}$ if the agent loses (Yea for rejected proposal, Nay for approved proposal) or $\alpha_i(t) = \alpha_{win}$ if the agent wins (Yea for approved proposal or Nay for rejected proposal). Thus α constitutes two free parameters.

And also drifts within the policy space according to Newton's force:

$$\Delta p_{\text{force}}(t) = k_i \cdot \sum_{j \in J} \left[\underbrace{(C_j(t) - p_i(t))}_{\text{Direction of force}} \cdot \underbrace{\frac{(\theta_j(t))^\eta}{|C_j(t) - p_i(t)|^2}}_{\text{Inverse-square law}} \right] \tag{12}$$

Where k_i is the agent's susceptibility to the policy space. $C_j(t)$ is the parties centers of mass at time t computed as:

$$C_j(t) \leftarrow \frac{1}{|J|} \sum_{i \in j} p_i(t) \tag{13}$$

θ_j represents the reach of the most likely political affiliation pull, while η is a coefficient to represent how much the agent cares about 'political definitions'.

Finally, to represent all idiosyncratic, or any other 'noise' that makes them deviate from a perfectly rational path is included as:

$$p_i(t+1) \leftarrow p_i(t) + \Delta p_{\text{learn}}(t) + \Delta p_{\text{force}}(t) + \epsilon_i(t) \tag{14}$$

Where $\epsilon_i(t) \sim \mathcal{N}(0, \sigma_i^2)$.

4 Predictions

Due to the dynamic and stochastic nature of my proposed model, I can't predict a static equilibrium point (s) without running the system on actual data. Instead, I propose deriving predictions as a set of comparisons from the model key parameters, which will drive the most salient dynamics.

The estimation process will find the set of parameters $(\alpha_{loss,i}, \alpha_{win,i}, w_i, K_i, ki, \eta_i, \theta_j, \sigma_i)$ for each legislator i that maximizes the log-likelihood of their observed voting history:

$$\hat{\Theta} = \underset{\Theta}{\arg\max} \sum_{i=1}^{n} \sum_{t=1}^{T} [V_{i,t} \cdot \log(P_{i,t}(\text{Yea})) + (1 - V_{i,t}) \cdot \log(1 - P_{i,t}(\text{Yea}))] \quad (15)$$

Where, $V_{i,t}$ is the observed vote of agent i on proposal t (1 for Yea, 0 for Nay). And $P_{i,t}(\text{Yea})$ is the model's predicted probability of a 'Yea' vote, defined by a logit function of the utility difference $D_i(t)$:

$$P_{i,t}(\text{Yea}) = \frac{1}{1 + e^{-D_{i(t)}}} \quad (16)$$

Where, $D_{i(t)} = U_{i(\text{Yea})} - U_{i(\text{Nay})}$

4.1 Hypothesis 1: asymmetric learning from legislative outcomes

I hypothesize that the legislator's ideological drift is not random, but a 'learning' process in which they are more sensitive to 'losses' (voting against the winning side) than to 'wins'. More concretely, $LL(\Theta)$, which will yield $\hat{\alpha}_{loss,i}$ and $\hat{\alpha}_{win,i}$ for each legislator i, then I'll simply test if $H_1 : \mu(\hat{\alpha}_{loss}) > \mu(\hat{\alpha}_{win})$ with $H_0 : \mu(\hat{\alpha}_{loss}) = \mu(\hat{\alpha}_{win})$ with a paired t-test

4.2 Hypothesis 2: systemic drift

As a consequence of hypothesis 1, I predict that Opposition members will exhibit a statistically significant ideological drift towards the Officialist party over the course of the ruling period. With $\hat{p}_i(t)$ for every legislator i, and including a normalization of the Officialist party's mean position $C_{\text{Officialist}}(t)$ to 0, I will, for each Opposition legislator i, fit a simple linear regression to their $\hat{p}_i(t)$ over a whole ruling period ($t = 1 \ldots T$):

$$\hat{p}_i(t) = \beta_{0,i} + \beta_{1,i} \cdot t + \epsilon_i(t) \quad (17)$$

This will produce a vector of slopes representing each Opposition member's ideological drift. Then I'll perform a simple one-sample t-test with $H_0 : \mu(\beta_1^{\text{Opposition}}) = 0$, and $H_1 : \mu(\beta_1^{\text{Opposition}}) > 0$.

4.3 Hypothesis 3: agent typologies

Finally, in a more loose sense, I propose that my model can classify different types of politicians by analyzing the combination of their other fitted parameters. Of the many possible combinations, I find that combining Low/High levels of K_i and w_i links how the 'micro' and the 'macro' of the political space can define a particular agent. I propose the four types of politicians:

- $K \uparrow, w \uparrow$ (the conflicted): this describes the model's core tension; both personal ideology and party loyalty pull these politicians. They will likely be on the party's side for some votes but will defect on ideologically salient ones.

- $K \downarrow, w \uparrow$ (the ideologue): this describes a politician that's entirely driven by personal ideology, insensitive to party pressure, and whose votes reflect their sincere ideology.

- $K \downarrow, w \downarrow$ (the apathetic): this politician has no drive. Their voting behavior is weak and will likely be dominated by other forces, such as the noise term.

- $K \uparrow, w \downarrow$ (the loyalist): this politician's behavior is driven by $S_i(t)$, with a weak personal ideology.

Finally, I propose that a bimodal distribution of loyalists and non-loyalists will permit the system-wide drift towards the ruling party center. The intuition here is that if there's no loyalty to the party's center, it will be very weak, so its induced gradient will be weak as well. On the other hand, if all agents are loyalists, there's no chance for drift. Having a balance between those types permits the system-wide drift with defined political centers and drifting agents that are more peripheral to parties.

References

[1] Duncan Black. "On the Rationale of Group Decision-making". en. In: *Journal of Political Economy* 56.1 (Feb. 1948), pp. 23–34. ISSN: 0022-3808, 1537-534X. DOI: 10.1086/256633. URL: https://www.journals.uchicago.edu/doi/10.1086/256633 (visited on 11/16/2025).

[2] Rosaria Conte and Mario Paolucci. "On agent-based modeling and computational social science". In: *Frontiers in Psychology* 5 (July 2014). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2014.00668. URL: http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00668/abstract (visited on 11/16/2025).

[3] Gary W. Cox and Mathew D. McCubbins. *Setting the Agenda: Responsible Party Government in the U.S. House of Representatives*. 1st ed. Cambridge University Press, Sept. 2005. ISBN: 978-0-521-85379-8 978-0-521-61996-7 978-0-511-79112-3. DOI: 10.1017/CBO9780511791123. URL: https://www.cambridge.org/core/product/identifier/9780511791123/type/book (visited on 11/16/2025).

[4] Anthony Downs. "An Economic Theory of Political Action in a Democracy". en. In: *Journal of Political Economy* 65.2 (Apr. 1957), pp. 135–150. ISSN: 0022-3808, 1537-534X. DOI: 10.1086/257897. URL: https://www.journals.uchicago.edu/doi/10.1086/257897 (visited on 11/16/2025).

[5] Andrew D. Martin and Kevin M. Quinn. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999". en. In: *Political Analysis* 10.2 (2002), pp. 134–153. ISSN: 1047-1987, 1476-4989. DOI: 10.1093/pan/10.2.134. URL: https://www.cambridge.org/core/product/identifier/S1047198700009931/type/journal_article (visited on 11/16/2025).

[6] Thomas J Sargent. *Bounded Rationality in Macroeconomics: The Arne Ryde Memorial Lectures*. en. Oxford University PressOxford, Mar. 1993. ISBN: 978-0-19-828864-0 978-1-383-01764-9. DOI: 10.1093/oso/9780198288640.001.0001. URL: https://academic.oup.com/book/52262 (visited on 11/16/2025).

[7] Richard S. Sutton and Andrew Barto. *Reinforcement learning: an introduction*. eng. Second edition. Adaptive computation and machine learning. Cambridge, Massachusetts London, England: The MIT Press, 2020. ISBN: 978-0-262-03924-6.