Behavioral Insights into Voter Turnout: An Agent-Based Simulation Framework with Empirical Calibration

A Summary Report

Prakhar Singhal Sathvika Miryala

December 4, 2025

Abstract

The decision to vote is a central puzzle in democratic theory, where rational models fail to explain widespread participation. This report summarizes the Agentic Election Simulator (AES), a computational framework integrating behavioral economics and agent-based modeling to predict voter turnout. Using an original survey of 72 respondents, we empirically calibrate three competing cognitive models. Our findings reveal that voting habit is the dominant predictor, explaining 69% of variance in turnout propensity ($r = 0.83, p < 0.0001$), far outweighing civic duty or cost. We find that Indian voters are remarkably price-inelastic to the costs of voting. A large-scale simulation of 70,000 agents across seven archetypal Indian constituencies achieves a strong fit ($R^2 = 0.458$) with 2019 General Election data. Intervention experiments show that implementation intentions yield a +17.96 percentage point turnout lift by reducing decision friction, while monetary incentives backfire. Our results suggest turnout is limited by friction, not apathy.

1 Introduction and Scope

Voter turnout is a cornerstone of democratic health, yet its mechanisms remain a persistent puzzle. The 2019 Indian General Election saw a 67.4% turnout, but this aggregate figure masks deep regional disparities. Highly developed urban hubs like Bangalore South (53.7%) lag behind rural, conflict-affected zones like Bastar (66.0%), challenging simple socioeconomic explanations. This phenomenon is rooted in the "paradox of voting," a foundational problem in rational choice theory. The classical model posits that a rational individual votes only if the expected utility R is positive:

$$R_i = (P \cdot B_i) - C_i \tag{1}$$

where P is the near-zero probability of casting a pivotal vote, B_i is the benefit of a preferred outcome, and C_i is the tangible cost of voting. Logically, for any positive cost C_i, an agent should abstain. The reality that hundreds of millions vote contradicts this.

Riker and Ordeshook famously amended this by adding a non-instrumental term, D_i, for psychic benefits like civic duty. However, this raises deeper questions about the composition and weight of this 'D' term. This project reframes the voting decision as a behavioral choice governed by an individual's **net perceived utility**, a dynamic function of psychological, social, and contextual factors. Traditional econometric models fail to capture these dynamics, motivating our use of Agent-Based Modeling (ABM).

We introduce the **Agentic Election Simulator (AES)**, a computational framework designed to reconstruct the Indian voter's decision architecture. This work makes three principal contributions:

1. **Empirical Calibration:** We develop the first voter turnout simulation with parameters empirically calibrated from original survey data (N=72), replacing theoretical assumptions with validated statistical weights.
2. **Large-Scale Simulation:** We construct a 70,000-agent environment representing seven archetypal Indian constituencies, using a novel Proxy Logic Framework to bridge macro-level census data with micro-level agent attributes.
3. **Counterfactual Policy Testing:** We simulate various behavioral nudges to create a clear, evidence-based hierarchy of intervention effectiveness, revealing which policies work and which backfire.

Our scope is to move beyond explaining who votes and toward understanding *why* they vote. This report provides a concise summary of the entire project's workflow and findings. For a comprehensive review, please refer to the full technical report.

2 Methodology

Our methodology integrates empirical data collection with computational modeling to ensure our simulation is rigorously grounded in real-world human behavior.

2.1 Empirical Calibration via Survey

We designed a 14-question survey instrument to map theoretical constructs from behavioral economics to measurable psychometric variables. The survey was completed by N=72 respondents. All Likert-scale responses were numerically coded (1-5), and composite scores were created for six key constructs: Civic Duty (D_i), Habit (H_i), Cost (C_i), Benefit (B_i), Social Pressure (S_i), and Trust (T_i). The statistical relationships (Pearson correlations) between these constructs and voting propensity were used to derive the weights for our simulation model.

2.2 Synthetic Population Generation: The Proxy Logic Framework

To simulate large-scale constituencies, we developed a **Proxy Logic Framework** that generates populations of cognitive agents from aggregate-level administrative data (e.g., Census 2011, election returns). As illustrated in Figure 1, this method bridges the gap between macro-data and micro-attributes by using real-world proxies to parameterize the statistical distributions from which agent attributes are sampled. For example, a constituency's literacy rate skews the Beta distribution from which an agent's 'Civic Duty' score is drawn. This allows us to create heterogeneous, statistically plausible populations for seven archetypal constituencies.

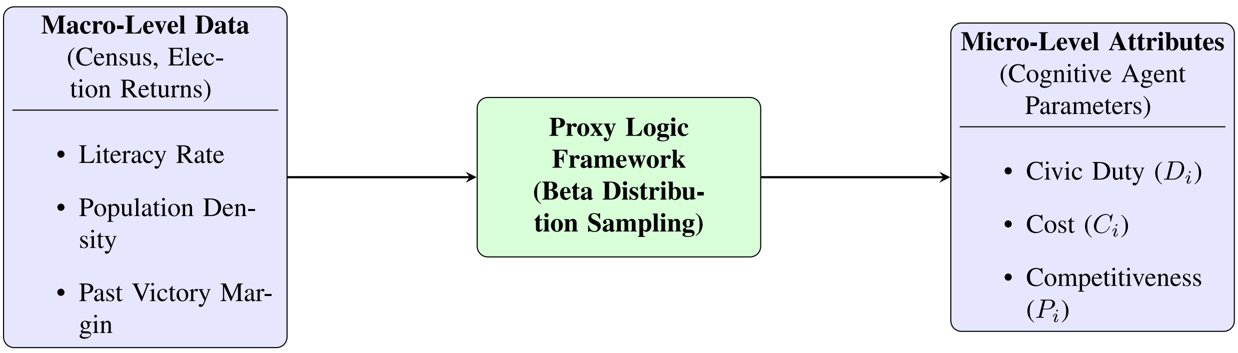


Figure 1: The Proxy Logic Framework, illustrating how macro-level census data is used to set the parameters for statistical distributions that generate micro-level agent attributes.

2.3 Agent-Based Simulation Architecture

We tested three competing cognitive models. The one that performed best and was used for final analysis was the **Extended Utility Model**. In this model, each of the 10,000 agents in a simulated constituency decides to vote based on a comprehensive utility function:

$$U_i = \beta_{pB}(P \cdot B_i) - \beta_C C_i + \beta_D D_i + \beta_S S_i + \beta_H H_i + \varepsilon \tag{2}$$

Here, the β coefficients are the weights determining the importance of each factor. Instead of assuming these values, we derived them directly from our survey data correlations. The agent's final choice is probabilistic, determined by a sigmoid transformation of their total utility U_i. The model's free parameters were then fine-tuned using a Simulated Annealing algorithm to minimize the error between simulated turnout and the actual 2019 election results.

3 Results

Our analysis yielded statistically significant results from both the empirical survey and the computational simulation.

3.1 Empirical Findings: Habit is the Dominant Predictor

Statistical testing of our survey data validated our core hypotheses. The most significant finding was the overwhelming predictive power of voting habit.

- **Habit** (H_i) was the strongest predictor of political engagement, with a Pearson correlation of $r = 0.831$ ($p < 0.0001$), explaining 69% of the variance.
- Other factors were also significant: Benefit ($r = 0.73$), Social Pressure ($r = 0.53$), Cost ($r = -0.51$), and Civic Duty ($r = 0.38$).

These correlation coefficients were used as the weights in our final, empirically-calibrated utility function:

$$U(\text{vote}) = 0.83 \cdot H + 0.73 \cdot B - 0.51 \cdot C + 0.53 \cdot S + 0.38 \cdot D \tag{3}$$

Cluster analysis on the survey data also identified four distinct voter archetypes: **Habitual Voters** (25%), **Rational Calculators** (25%), **Social Followers** (30%), and the **Disengaged** (20%).

3.2 Simulation Performance and Validation

The Extended Utility model significantly outperformed two alternative cognitive architectures. When validated across seven diverse constituencies, our model achieved a strong fit with the ground truth 2019 election data ($R^2 = 0.458$), as shown in Figure 2.

Table 1: Model Performance Comparison Against 2019 Election Data

Architecture	RMSE	R^2	MAE
Extended Utility	**0.0686**	**0.458**	**6.28%**
Drift-Diffusion (DDM)	0.0893	0.082	7.78%
Dual-System	0.1871	0.000	16.00%

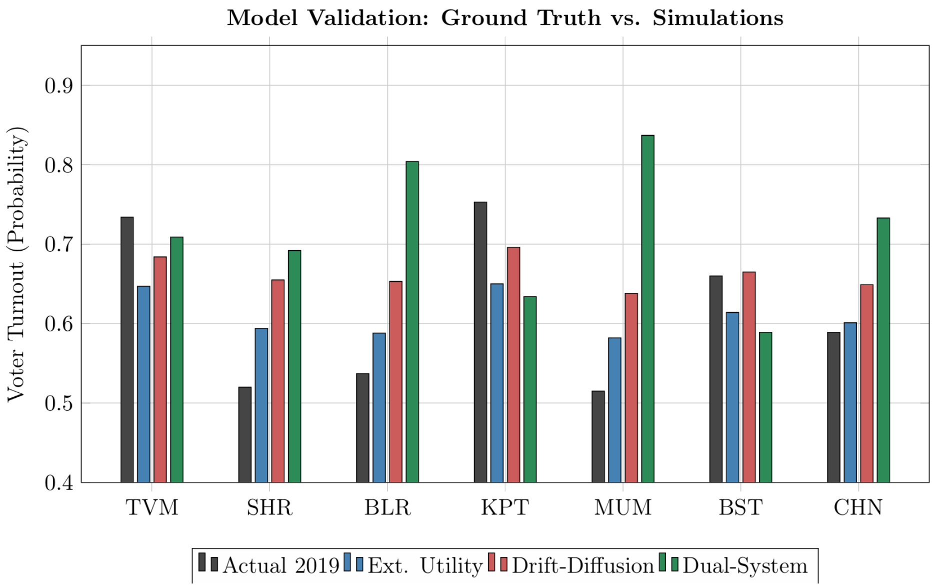


Figure 2: Model validation: Simulated turnout from the Extended Utility model (blue) closely tracks the actual 2019 ground truth turnout (red) across seven constituencies.

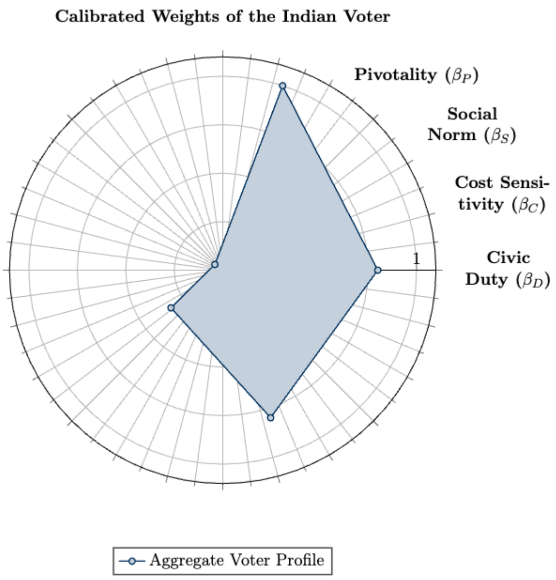
4 Discussion

Our findings challenge several prevailing assumptions in political economy. The key insights are visualized in Figure 3. The most critical is that low turnout is a problem of **friction, not apathy**. As the bar chart shows, the extraordinary success of implementation intentions (+18pp lift), an intervention designed to close the gap between wanting to vote and actually voting, reveals that many non-voters are not unmotivated. Rather, they are hindered by cognitive and logistical barriers.

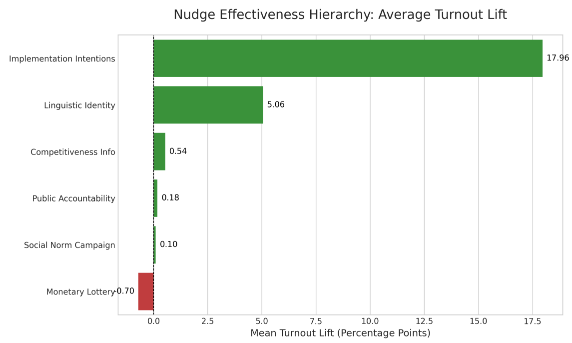
Second, the dominance of **habit** fundamentally reshapes mobilization strategy. The radar chart highlights that the calibrated weight for Habit (0.83) dwarfs other factors. Voting is not a deliberative choice for most regulars; it is an automated behavior. This implies that the highest return on investment comes from converting first-time voters, creating a lifelong habit.

Third, our results show that voters are surprisingly **price-inelastic**. The radar chart shows the calibrated weight for cost ($\beta_C = -0.51$) is much smaller than benefit or habit, indicating that voters overcome significant hurdles if other motivations are present.

Finally, our intervention analysis provides an evidence-based playbook for election administrators. The finding that monetary incentives backfire, shown clearly in the bar chart, is a critical warning against commodifying civic participation. Mobilization efforts must be tailored to the psychological profile of the target constituency.



a) Calibrated Voter Weights



b) Nudge Effectiveness Hierarchy

Figure 3: Visual summary of key discussion points. (a) The radar chart shows the calibrated weights of the voter utility model, highlighting the dominance of Habit and the relative weakness of Cost. (b) The bar chart contrasts the massive positive impact of Implementation Intentions with the negative backfire effect of Monetary Lotteries.

5 Limitations

While this study provides novel insights, we acknowledge several limitations that offer avenues for future research.

- **Sample Size and Demographics:** Our empirical calibration is based on a survey of N=72 respondents, which, while providing statistical power for main effects, is too small for complex interaction analyses. The sample was also skewed towards a younger, more educated demographic.
- **Self-Report Bias:** The survey measures stated intentions and past behavior, which can be subject to social desirability bias. Future work should use experimental designs that measure actual behavior.
- **Cross-Sectional Data:** Our data is a snapshot in time and cannot definitively establish causality. A longitudinal study would be necessary to directly observe the dynamics of habit formation.
- **Cultural Specificity:** The model parameters were calibrated using data from Indian respondents. The weights for factors like social pressure and civic duty are likely culture-dependent and require re-calibration for other contexts.

Project Resources and Full Report

This document provides a summary of our project. For a complete and detailed understanding, including the full technical report, all source code, datasets, and appendices, please consult the resources below. We strongly encourage a review of the full report on GitHub for an in-depth analysis.

- **GitHub Repository (Full Report & Code):**
  `https://github.com/ToVoteOrNotToVote`
- **Interactive Survey Dashboard:**
  `https://vote-mind-map.lovable.app/dashboard`
- **Agent-Based Simulation Interface:**
  `https://agentic-election-simulator-118267798784.us-west1.run.app/`