Forecasting the 2019 Spanish Elections: A Comparative Analysis of the “Surprisingly Popular” Algorithm

Leonardo Fabbri

Universitat Pompeu Fabra, Barcelona

Topic: Collective Intelligence, Political Forecasting

Abstract

Collective intelligence mechanisms, based on the ”wisdom of the crowd,” postulate that aggregated group judgments often surpass individual accuracy. While traditional democratic methods assign equal weight to all opinions, recent advancements such as the ”Surprisingly Popular” (SP) algorithm proposed by Prelec et al. (2017) attempt to identify expert knowledge by selecting answers that are more common than the crowd predicts. This study investigates the applicability of the SP algorithm in a predictive geopolitical setting: the 2019 Spanish Elections. Using a survey of 461 participants regarding 39 political outcomes, this research compares the SP algorithm against the Standard Wisdom of the Crowd (Simple Majority) and Confidence-Weighted models. Contrary to the standard theory, the Simple Majority method showed the highest overall performance (F1-score 86%) due to a high baseline accuracy of the crowd. However, the SP algorithm demonstrated superior performance in **Specificity (95%)** and **Precision (92%)**, while maintaining a lower Recall (67%). This finding suggests that SP is a highly conservative metric, which, by minimizing false positives, proves uniquely useful for political forecasting scenarios where the cost of predicting a non-event is high.

1 Introduction

The concept that ”many heads are better than one,” often referred to as the ”wisdom of the crowd,” traces back to Aristotle, yet remains a cornerstone of modern decision-making structures across economics, finance, and political science. The fundamental premise, famously demonstrated by Francis Galton (1907) at a livestock fair, suggests that aggregating independent estimates can filter out idiosyncratic noise, resulting in a collective judgment that is often more accurate than most, if not all, individual guesses.

However, the efficacy of the Standard Wisdom of the Crowd (Simple Majority) relies heavily on the independence of judgments and the condition that the average guess is close to the truth. Traditional methods often fail when the majority holds a shared misconception—a scenario prevalent in domains like general knowledge or, critically, political forecasting where misinformation or common biases can proliferate.

To address this limitation, Prelec, Seung, and McCoy (2017) introduced the **”Surprisingly Popular” (SP) algorithm**. This novel approach moves beyond simple aggregation by incorporating a meta-cognitive component. Instead of asking only ”What is the answer?”, it asks two questions:

1. What do you believe the right answer is? (Belief)

2. What do you predict the popular opinion will be? (Prediction)

The SP algorithm then selects the answer whose actual frequency exceeds the crowd's predicted frequency, thereby exploiting a fundamental principle: experts not only know the correct answer but also recognize that the majority is likely to be wrong. This mechanism allows a sophisticated minority to potentially override an ill-informed majority.

While Prelec et al. (2017) validated the SP model on general knowledge questions with clear, unambiguous answers, its utility in geopolitical forecasting—where the objective "truth" is dynamic and often contested before the event—remains critically under-explored. This paper aims to fill that gap by testing the SP algorithm against traditional methods using survey data collected immediately prior to the 2019 Spanish general, regional, local, and European elections.

2 Literature Review

2.1 Standard Models of Collective Intelligence

The theoretical underpinnings of the wisdom of the crowd are formally captured by the *Condorcet Jury Theorem* (1785), which states that if each individual voter is more likely than not to choose the correct outcome, the probability of the majority choosing the correct outcome increases with the number of voters.

In practice, two primary aggregation models prevail:

- **Simple Majority (SWoC):** The simplest democratic aggregation, where the answer selected by $> 50\%$ of the group is chosen.
- **Confidence-Weighted Aggregation:** An improvement over SWoC where each vote is weighted by the individual's self-reported confidence in their answer. This attempts to give greater weight to knowledgeable individuals, although studies often show a weak correlation between confidence and accuracy.

2.2 The "Surprisingly Popular" Algorithm

The SP algorithm was specifically designed to solve the "single-question crowd wisdom problem," which occurs when the ground truth is needed from a single set of responses without the benefit of repeat polling or external verification.

[Surprisingly Popular (SP) Criterion] Let a be an answer option (e.g., "Yes") and b be the alternative answer (e.g., "No"). Let A be the proportion of the crowd that chooses a (actual popularity), and E_A be the average predicted proportion of the crowd that will choose a (predicted popularity). The SP algorithm selects the answer a if and only if:

$$A > E_A$$

If the actual frequency of an answer exceeds the crowd's prediction of its frequency, that answer is deemed "Surprisingly Popular" and is selected as the collective judgment.

The intuition is that an informed individual knows the correct answer a but also anticipates that uninformed individuals will choose the common, incorrect answer b, thereby underestimating the frequency of a. Conversely, an uninformed individual who chooses b is likely to overestimate the popularity of b and underestimate the popularity of the correct answer a. The difference $(A - E_A)$ thus acts as a filter for separating knowledge from common bias.

3 Methodology

3.1 Study Design and Participants

The research was conducted using a survey distributed one week prior to the April 2019 Spanish elections via the Behavioural Science Laboratory (BESLab) at Universitat Pompeu Fabra. The total sample size for the analysis was $n = 461$.

3.2 Procedure

Participants were presented with a total of 46 forecasting statements, which were filtered down to 39 binary outcomes (Yes/No) covering various levels of the Spanish political system (general, regional, local, and European elections). For each of the 39 statements, participants provided the three core measures necessary for testing the algorithms:

1. **Binary Prediction (Belief):** The participant's own prediction (Yes/No).
2. **Confidence Rating:** A self-assessed probability (50% to 100%) of their prediction being correct.
3. **Meta-cognitive Prediction:** The estimated percentage (0% to 100%) of other participants who would choose "Yes." This is the critical measure for calculating the E_A in the SP algorithm.

3.3 Evaluation Metrics

The three aggregation methods—Simple Majority (SWoC), Confidence-Weighted, and Surprisingly Popular (SP)—were evaluated against the ground truth (the actual election results) using standard classification metrics:

- **F1-score:** The harmonic mean of Precision and Recall, providing a balanced measure of overall accuracy.
- **Precision (Positive Predictive Value):** The proportion of positive predictions that were actually correct (True Positives/(True Positives + False Positives)).
- **Specificity (True Negative Rate):** The proportion of negative outcomes that were correctly identified as such (True Negatives/(True Negatives + False Positives)).
- **Recall (Sensitivity):** The proportion of actual positive outcomes that were correctly predicted (True Positives/(True Positives + False Negatives)).

4 Results

4.1 Overall Performance Comparison

The Standard Wisdom of the Crowd (Simple Majority) method achieved the highest overall performance, as measured by the F1-score.

The Simple Majority's high F1-score (86%) and Recall (89%) indicate that the crowd was highly accurate on the majority of the 39 political outcomes. This high baseline accuracy (the majority was correct 87% of the time in the final paper analysis) provided limited opportunity for the SP algorithm to demonstrate its corrective power.

Table 1: Summary of Model Performance (F1-score, Precision, and Recall)

Method	**F1-score**	**Precision**	**Recall (Sensitivity)**
Simple Majority	86%	84%	89%
Confidence-Weighted	81%	79%	83%
Surprisingly Popular	77%	92%	67%

4.2 Specificity and Precision Dominance

Despite having the lowest F1-score, the SP algorithm exhibited clear superiority in the metrics related to minimizing errors of commission:

- **Specificity (True Negative Rate):** The SP algorithm scored 95%, significantly higher than Simple Majority (86%) and Confidence-Weighted (81%).
- **Precision (Positive Predictive Value):** The SP algorithm also led with 92%, compared to Simple Majority's 84% and Confidence-Weighted's 79%.

These results demonstrate that when the SP algorithm did issue a "Yes" prediction (a positive outcome), it was correct 92% of the time. Furthermore, it excelled at correctly predicting "No" outcomes (a negative non-event), with a 95% success rate.

5 Discussion and Conclusion

5.1 The Conservative Nature of the SP Algorithm

The primary finding—that Simple Majority outperformed SP in overall F1-score—is not a definitive rejection of SP, but rather a contextual finding. In the 2019 Spanish election predictions, the crowd was exceptionally well-informed, minimizing the incidence of collective illusions that SP is designed to correct.

However, the SP algorithm's striking dominance in **Specificity (95%)** and **Precision (92%)** reveals its true value in practical forecasting. Its high Precision means that its positive predictions are highly reliable, while its high Specificity means it rarely raises a false alarm. The trade-off is its low Recall (67%), meaning it misses many true positive events (errors of omission).

5.2 Implications for Political Forecasting

In geopolitical and political forecasting, not all errors are equally costly. In a scenario where the cost of a *false positive* (e.g., falsely predicting a major political party will lose its majority, or a market crash) is extremely high, the conservative nature of the SP algorithm becomes an advantage. It operates like a high-confidence filter, only predicting an outcome when the actual belief significantly outstrips the expected belief, providing a strong signal of a genuine "surprise."

This study provides compelling evidence that the "Surprisingly Popular" algorithm is a viable and valuable tool for political forecasting, particularly for practitioners who prioritize **signal reliability** over comprehensiveness. It should be considered not as a replacement for Simple Majority, but as a complementary tool in an ensemble forecasting model, used specifically for identifying high-confidence predictions and minimizing the risk of false alarms.

6 References

1. Galton F. (1907). Vox Populi. *Nature*, 75, 450–451.

2. Lee, M. D., Danileiko, I., & Vi, J. (2018). Testing the ability of the surprisingly popular method to predict NFL games. *Judgment and Decision Making*, 13(4), 322.

3. Prelec, D., Seung, H. S., & McCoy, J. (2017). A solution to the single-question crowd wisdom problem. *Nature*, 541(7638), 532.

4. Surowiecki, J. (2004). *The Wisdom of Crowds.* Anchor Books.

5. Sunstein, C. R. (2006). *Infotopia: How many minds produce knowledge.* Oxford University Press.