DO ELECTION SYSTEMS DRIVE POLARIZATION? AN AGENT-BASED MODEL APPROACH

Jamie Hackney

Advisor: Professor Phil Chodrow

A Thesis

Presented to the Faculty of the Computer Science Department

of Middlebury College

May 2024

ABSTRACT

Ideological polarization affects the ability of democratic governments to function. Levels of ideological polarization have been rising globally in recent years. Many theories have been put forward about potential causes, but the connection between the election system used and the degree of ideological polarization has been less studied. We present an agent based model to simulate plurality and instant-runoff ranked choice election systems based on Bounded Confidence and Attraction-Repulsion opinion updates. We analyze the long term behavior of these models and quantify the degree of ideological polarization as the variance among agent opinions. We provide a set of parameters that cause plurality election systems with Attraction-Repulsion opinion updates to end with high variance while ranked choice election systems end with near-zero variance. We conduct sweeps between four pairs of model parameters and discuss the phase transitions in variance they produce.

ACKNOWLEDGEMENTS

I would like to thank Professor Chodrow for advising me on this thesis. Your guidance was invaluable, and much of this work would not have been possible without it. I would also like to thank Professor Robbett and the class of CS702 for providing additional guidance and feedback throughout this process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Polarization, as an abstract concept, describes the process in which members of a group become divided into opposing groups. As it relates to politics, polarization generally exists in two forms: ideological polarization and affective polarization [22]. Whereas affective polarization studies the change in individuals feelings and emotions towards opposing parties, ideological polarization refers to how divergent individuals of opposing parties become in their opinions and political preferences. While both forms of polarization are consequential to the outcome of political processes, this paper will focus on ideological polarization (henceforth referred to simply as polarization).

Although a lack of data has made cross-national measurements and comparisons of polarization difficult, a study conducted by McCoy et al. [26] using a 2020 dataset and a polarization metric from the Varieties of Democracy Institute shows that polarization has generally been increasing globally since 1900. While variation exists throughout the dataset, an important finding was that every region of the world except for Oceania has seen a rise in the level of polarization since 2005. When studying party system polarization using the Comparative Study of Electoral Systems (CSES) dataset with data from 21 established democracies, Dalton [6] found that 8 democracies showed a statistically significant increase in polarization from the mid 1990s to 2020 when measured with the CSES Polarization Index, a measurement of party vote shares and ideological position. More countries demonstrated an increase in their polarization index, but the small number of elections made statistical significance too high a bar.

Polarization has also been increasing in the context of US politics. A study conducted by Abramowitz and Saunders [1] looked at public opinion polls from 1982 to 2004 and found that polarization, when measured as the absolute value of the difference between the number of liberal positions and the number of conservative positions, has been increasing since the 1980s. Furthermore, they showed that this polarization exists in both the general public as well as in political elites. The survey conducted by the Pew Center for their 2014 report *Political Polarization in the American Public* [31] found evidence in line with the findings of Abramowitz and Saunders [1], that is, polarization has been increasing in the American public. Their survey found that from 2004 to 2014, the share of Americans who were uniformly liberal or conservative across most values doubled, from about one-in-ten to one-in-five. A result of this increase of individuals with ideologically consistent views is that the partisan gap in opinions nearly doubled from 1994 to 2014. The study shows that in 1994, 23% of Republicans were more liberal than the median Democrat, and 17% of Democrats were more conservative than the median Republican. In 2014, those shares shrank to 4% and 5%, respectively.

Some degree of polarization has been shown to have positive effects on democratic political processes as it can increase political participation, which generally has a positive impact on democratic political processes [2]. Abramowitz and Saunders [1] showed how intense polarization of the electorate about George W. Bush in the 2004 presidential election was reflected in high levels of voter turnout and engagement in the run-up to the election. The findings of Pew Research's

Political Polarization in the American Public [31] 2014 study are in line with those of Abramowitz and Saunders [1]. The percentage of individuals who always vote and who contributed to a political candidate or group in the past two years are highest among consistently liberal and consistently conservative groups. That pattern holds for other types of political engagement as well, including contacting an elected official, attending a campaign event, and working or volunteering for a candidate or campaign.

That said, polarization, especially at extreme levels, tends to also have negative consequences on democratic political processes [26]. For one, polarization can make it difficult for the government to effectively solve issues. Pew Research's *Political Polarization in the American Public* [31] 2014 study shows that polarized groups tend to think that it is the responsibility of the other group to compromise, leading to a lack of progress. It can also contribute to government dysfunction, political conflict, democratic erosion, and incremental autocratization [26]. Specifically, McCoy et al. [26] found that there tends to be an inverse relationship between polarization and the quality of democracy. Finally, polarization may also make the nation harder to govern and alienate voters of the losing party [6].

The existence of polarization may seem surprising given that theoretical work has shown that under the following assumptions, convergence to the median voter's position is the ideal strategy for politicians in two-party systems [10, 34]:

1. Elections are periodic.
2. A single winner is chosen at each election.
3. There are only two Candidates in each election.

4. The Candidate with a majority of votes is the winner.
5. Candidates are fully informed of Voter opinions.
6. Voters are fully informed of Candidate opinions.
7. Opinions exist in one dimension.
8. There is a sufficiently large number of Voters.
9. The Voters opinion distribution is unimodal.
10. Voters vote based on minimizing the Euclidean distance to the possible Candidates.
11. All Voters vote.

Given that levels of polarization have instead been increasing in the U.S., which is effectively a two party system, it would seem important to understand the cause(s). Much work has gone into studying the validity of the assumptions [15, 17, 16, 18]. Several theories beyond the assumptions have also been put forward, including electoral and social change, increasingly restrictive congressional rules, increasingly ideological party leaders, and the greater control majority party leaders have over congressional work [23]. The ideological identity of parties and the political history of a nation may also be determinants in the level of polarization [6].

The relationship between election systems and the degree of polarization has also been explored [24, 25], but these studies have largely investigated this relationship through the lens of electoral rule disproportionality, that is, the level of disproportional representation of the electorate. We propose a method to understand the relationship between the election system *itself* and the degree of polarization through the use of an agent-based model.

1.1 Agent Based Models

Agent based models, or ABMs, describe a general approach to modeling dynamic systems in which key actors of the system are abstracted into *agents*. These agents then interact with each other and the environment according to specified rules of interaction [33]. The long term behavior of the model is then studied to gain an understanding of the system it is meant to model. This process of studying the system level behavior of the model that arises as a result of agent interaction is central to ABMs, as it shows the idea of *emergent behavior*. More generally, emergent behavior is the idea that interactions between low level parts of the system can give rise to complex, system level behavior [33, 32]. Our emergent behavior is polarization, which arises as a result of interactions between the agents in our model. ABMs have been used to model processes in a wide variety of fields, including political decision making processes [32]. Chapter 3 will go into detail about the specific functioning of our model, but from a high-level view, our ABM models voters, candidates, and elections all in the same system to show how polarization develops.

To do so, we view voters and candidates as agents and abstract the idea of opinions and policy preferences to *opinion space*, a d dimensional space where agents can take any position $x \in [0,1]^d$ on some arbitrary amount of d issues. An agent's political opinion is therefore defined entirely by a vector $x \in \mathbb{R}^d$. To make this idea more concrete, consider an example where we care about two issues, healthcare and military spending. Let us define an agent's opinion on healthcare as $x_1 \in [0,1]$ and their opinion on military spending as $x_2 \in [0,1]$. We can define a value of $x_1 = 0$ as meaning an agent believes healthcare should be entirely privately funded, whereas a value of $x_0 = 1$ implies an agent believes healthcare should be entirely funded by the government. Since x_1 exists in continuous space, it can take any value between $[0,1]$ to reflect nuance in the agent's opinion. We can define values for x_2 in a similar way. We can therefore now entirely describe an agents opinion with a vector $x = [x_1, x_2]$. This abstraction to opinion space can be seen as an extension of opinion polling, which often asks respondents to rank their opinions on a numerical scale.

Importantly, this vector x also describes a *position* in opinion space, which in this example is on a two-dimensional plane. An agent moving around in opinion space therefore corresponds to them changing their opinion. How agents move was a key question, and our approach is explained in detail in Chapter 3, but is largely based on rules of interaction defined for the agents. The abstraction to opinion space, along with the idea of movement in opinion space based on rules of interaction, form the fundamental basis of our model. Fig. 1.1 provides an example visual representation of our model. Agents are shown in two-dimensional opinion space, with Candidate agents being represented by circles and Voter agents being represented by X's. Voters are color coded to the candidate they voted for. The figure shows the model's state at time step 500.

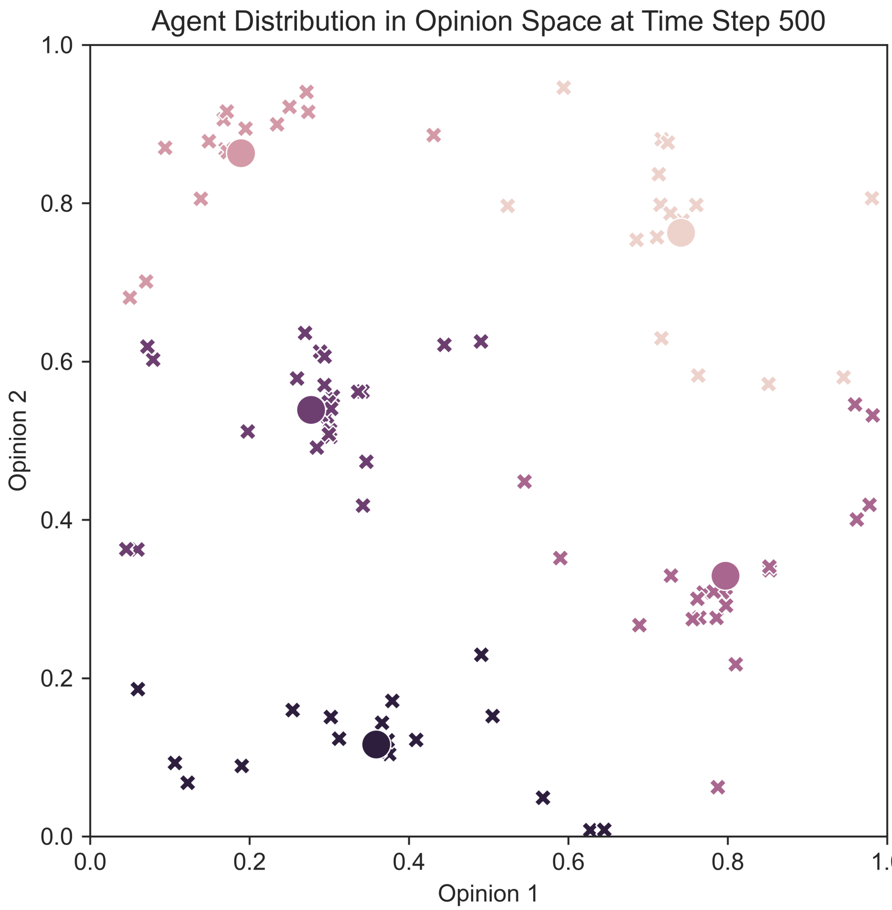


Figure 1.1: The distribution of agents in opinion space [0, 1] at time step 500. Voters are represented by X's and Candidates are represented by circles. Voters are color coded based on the candidate they support.

A chief advantage of this approach is that it allows us to model both the electorate and the candidates in the same system to see how each becomes polarized. Additionally, we are able to control model parameters to see how changes in agent actions affect the long term behavior of the system. The major disadvantage, however, is that ABMs in general offer low *ecological validity*, which describes how well the model matches reality [32].

To discuss the long term behavior of our model in relation to polarization, we first need to define how we will measure polarization. In accordance to Axelrod, Daymude, and Forrest [3], we will define polarization as the variance in opinions of the agents:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \tag{1.1}$$

As variance is only defined in one dimension, the analysis in this paper that involves variance will be in one dimensional opinion space.

Chapter 2 will provide a review of literature related to our work. Specifically, literature related to ABMs in political decision making, ABMs in opinion dynamics, and ABMs and other opinion dynamic models related to polarization will be discussed. Following that, Chapter 3 will describe how the model functions, both informally and formally. The purpose of the informal definition is to allow readers with limited technical knowledge to develop an intuitive understanding of the model and how it functions. The purpose of the formal definition will be to provide exact mathematical descriptions of the model and it's rules so that it may be replicated. Finally, Chapter 4 will discuss the results obtained from studying the long term behavior of the model, and Chapter 5 will provide possible interpretations of those results.

CHAPTER 2

RELATED WORK

2.1 ABMs in Political Decision Making

ABMs have been used to study political processes because they offer a means to understand these systems as dynamic and interactive processes which are not so easily described with analytical methods [32]. For example, Fowler and Smirnov [12] designed an ABM to study the Downsian paradox, which posits that because the probability of one's vote changing the outcome of an election is low, and the expected gain from the outcome of an election is likely smaller than the cost to vote, rational voters should not turn out to vote [10, 32]. Their model also studied the effect that mutual influence between voters has on turnout. This was achieved by making a voter's decision on whether to turnout to an election dependent on both their satisfaction from the previous election as well as observations of their neighbors satisfaction. This modeling choice is an example of bounded rationality, where agent's incomplete information and knowledge of the state of the system leads to decisions that may not be considered rational if they did have complete information [32].

Sobkowicz [36] extended the concept of voters incomplete access to information to understand how opinion formation can be used to predict real world polling and election results. Using the cusp-catastrophe model [11, 37] and probabilistic opinion updates based on information access, it was shown that an ABM can be used to predict the outcome of a complex political process [32]. Specifically, the model was calibrated to predict the 2015 polish presidential election using polling data, and its predictions were generally aligned with the final election results [32].

Clough [4] and Clough [5] developed an ABM to study Duverger's law, which states that two-party systems are likely to emerge from single-member plurality systems and multi-party systems are likely to emerge from proportional representation systems [32]. These models focused on the idea of strategic voting, where voters vote for parties based on prospective ratings. They calculate each parties prospective rating as a function of the expected utility they would receive from that party if it won and the likelihood of each party winning, which is derived from learning the voting preferences of their neighbors. Each voter and party is given an ideological position from 0 to 100, and the expected utility is calculated according to the distance between their position and the party's position [32].

Kim, Taber, and Lodge [20] showed that people tend to use motivated reasoning when forming political opinions [32]. To do so, they developed an ABM where agents form opinions with motivated reasoning and Bayesian learning theories. They calibrated the agents with election survey data and by comparing model predictions to new rounds of election data, they showed that motivated reasoning was a more accurate predictor of attitude formation than Bayesian learning.

DeMarzo, Vayanos, and Zwiebel [8] showed that persuasion bias, or susceptibility to repeated information, influences agent's opinion formation. To do so, they developed an ABM based on boundedly rational opinion formation where individuals fail to account for receiving repeated information. They showed that persuasion bias implies the phenomena of social influence and unidimensional opinions. The phenomenon of social influence is where one's influence over a group opinion depends not only on the accuracy of the information, but also their connections in the communication network. The unidimensional opinions phenomenon is where individual's opinions over a multidimensional set of a issues converges to a single dimension, often denoted as a left-right spectrum in political science applications.

2.2 ABMs in Opinion Dynamics

ABMs have also been widely used to study opinion dynamics, that is, what causes individuals to become polarized in their opinions, reach consensus, or fragment such that opinions are clustered in groups.

Schweighofer, Garcia, and Schweitzer [35] developed an ABM to reproduce the unidimensional opinions phenomenon. They derive their interaction rules from cognitive dissonance theory and structural balance theory and develop two methods for how agents perceive similarity in multi-dimensional, continuous opinion space, based on the paradigms of proximity voting and directional voting. The proximity voting paradigm implies that agents perceive similarity as the euclidean distance in opinion space, whereas directional voting implies that opinion similarity depends on which *side* of an opinion each agent is on.

Axelrod, Daymude, and Forrest [3] developed an ABM with rules of interaction termed *Attraction-Repulsion*, based on the idea that when individuals interact with others with similar opinions, they tend to get closer in opinion, but when interacting with individuals with dissimilar opinions, they tend to grow apart in opinion. This rule of interaction can be seen as an extension of the bounded confidence model developed by Deffuant et al. [7], Krause [21], and Weisbuch et al. [38]. Specifically, the attraction repulsion model defined three parameters, E, T, R, representing the exposure, tolerance, and responsiveness of agents, respectively. At each time step, a random agent interacts with another with probability $(\frac{1}{2})^{\frac{d}{E}}$, where d is the distance between the two agents in opinion space. If they interact, and if their distance is less than the tolerance T, then the agent moves $\frac{d}{R}$ towards the other agent. On the other hand, if they are more than T apart, then the agent moves $\frac{d}{R}$ away.

2.3 Models Focused on Polarization

ABMs, as well as other types of opinion dynamics models, have been studied to specifically understand the development of polarization in a population. In the context of opinion dynamics, polarization is where population opinions converge to groups occupying opposite ends of opinion space. Axelrod, Daymude, and Forrest [3] showed that the long term behavior of the attraction repulsion model depends largely on the parameters. To study the characteristics of opinions in the model, they defined the degree of polarization as the variance in opinions. For low values of tolerance ($T \leq 0.15$), they found that the model produced extreme polarization, i.e. high variance, while for high values of tolerance ($T \geq 0.55$), the population converged around a single position. They found that the long term behavior of the model is largely invariant with respect to R, that is, the result of whether the population becomes polarized or converges to a single opinion is mostly independent of how much they move based on interactions, except for small values of R. Finally, they also found a similar result in the effect of E on the long term behavior. Except for small values of E, whether or not the population polarizes is largely dependent on T, implying that polarization depends largely on agent's willingness to consider dissimilar opinions positively.

Kashima et al. [19] showed that ideological commitment and opinion interpretation influence the polarization of political discourse. They use a connectionist approach called the tensor product model to show ideological thinking. In their model, agents move in opinion space as a result of encountering various opinions. These opinions are first passed through an interpreter mechanism before an opinion is generated. Agents also have a memory that represents the associations between the interpreted opinions and the represented source. Agents express opinions by accessing memory using their self representation as a cue, which is then passed through the interpreter in reverse to generate an output expression.

Combining the interpretation-based and memory-based ideological mechanisms gives four types of ideological thinking: fully ideological, ideological interpreter, ego-involved ideologist, and non-ideological. Most similar to the agents developed in our model are the ego-involved ideologist and non-ideological types, as they interpret received opinions without bias. Ego-involved ideologists tended to retain a learned ideology, but would change opinions depending on whether opinions they interact with confirm or contradict that ideology. Whether these agents polarize, therefore, is dependent on the opinions they interact with. Similarly, non-ideological agents change their opinions based on the opinion they interact with, but whether or not they polarize depends on communication and social network dynamics. Specifically, on the Parsegov et al. [29] and Friedkin et al. [13]'s model of multidimensional social influence on social networks, these agents always formed consensus when there was no mobility, but could form divergent opinion clusters with mobility.

Mueller and Tan [27] used a simulation approach to show that standard approaches to modeling opinion dynamics involving an influence bound are not unique in their ability to produce fragmentation and polarization. An influence bound, present in our model as well, is based on the idea that people only interact with opinions that are similar enough to their own, often represented by agents ignoring or being repulsed by opinions that are further away from some threshold. Mueller and Tan [27]'s work shows that a cognitively grounded model, alternative to influence bounds, and based on assumptions about knowledge or belief spaces, can also produce the group opinion behaviors seen in influence bounded models.

Hegselmann [14] looked at four models of opinion dynamics: the classic model, the Friedkin-Johnsen model, the time-variant model, and the bounded confidence model to understand when opinion formation within an interacting group leads to consensus, polarization, or fragmentation. The bounded confidence model they implement, as developed by Deffuant et al. [7], Krause [21], and Weisbuch et al. [38] gives opinion updates according to:

$$x_i(t+1) = |I(i, x(t))|^{-1} \sum_{j \in I(i,x(t))} x_j(t) \text{ for } t \in T \tag{2.1}$$

where $x_i(t)$ is agent i's opinion at time step t, $I(i, x) = \{1 \leq j \leq n, |x_i - x_j| \leq \epsilon_i\}$, and $\epsilon_i > 0$ is the confidence bound for agent i. When studying the bounded confidence model with analytic methods, they show that the model approaches consensus if any two agents can be connected by the bound from another agent.

These results were derived under the assumption of a uniform confidence bound, i.e., $\epsilon_i = \epsilon_j$ for all agents i, j. Simulations were run to understand the opinion dynamics of varying the bound. Specifically, simulations were run with symmetric but increasing bounds, opinion independent asymmetric bounds, and opinion dependent asymmetric bounds. With increasing symmetric and asymmetric bounds, the opinion dynamics go from polarization to consensus. Additionally, symmetric bounds lead to symmetric polarization, while asymmetric bounds lead to asymmetric polarization. While these simulations explored various asymmetric and changing bounds, they were all homogeneous in the sense that the size and shape of the bounds is the same for all agents. On the contrary, processes in our model lead to heterogeneous confidence intervals among agents.

CHAPTER 3

METHODS

To develop on understanding of the relationship between election systems and polarization, we propose an ABM to simulate elections. Here, our model will first be described informally so that readers without previous exposure to ABMs will be able to intuit a basic understanding of how the model functions. Then, the model will be described formally in Section 3.2 using the "Overview, Design concepts, and Details" (ODD) protocol as described in Railsback and Grimm [33], to allow for reproduction of the model and its results.

3.1 Informal Model Definition

First, we provide an informal definition of our model so that readers with limited technical knowledge can gain an intuitive understanding of the process. Our model defines two agents, Voters and Candidates, and a set of rules that describe how the model state changes over time. For Voters these rules relate to updating agent opinions as a result of interactions between Voters and Candidates. For Candidates, these rules relate to updating opinions to increase their expected performance in elections. The model progresses in discrete time steps, or rounds, which are either *movement rounds* or *election rounds*. Elections happen in cycles, such that there are some constant number of movement rounds before each election. The order of processes in movement rounds is:

1. A random Voter moves in opinion space by changing their opinion based on

interacting with another random Voter and Candidate.

2. Every 20^{th} round, Candidates strategically move in opinion space by changing their opinion to increase their probability of winning the upcoming election.

In election rounds, the process order is:

1. A movement round happens.
2. Candidates who have reached their term limit are removed from the model.
3. A winner is chosen according the election system specified, and candidates are politically advantaged or disadvantaged depending on how they placed in the election.
4. Candidates stochastically enter and exit the model.

Section 3.2 goes into precise detail about mechanisms describing each change, but a general understanding of the model is that every round Voters update their opinions, and every 20^{th} round Candidates update their opinions. The specific ways they change their opinion and move strategically are detailed in the equations in Section 3.2. Then, if the round is an election round, the winner is elected based on one of two election systems: *plurality* or *instant-runoff ranked choice*. The election has some consequences for the model, which can be thought of as changing the political power of candidates based on how they performed in the election. With this informal but intuitive idea of the model in mind, the formal definition follows in Section 3.2.

3.2 Formal Model Definition

Here, we describe the model formally and in detail with the goal of allowing reproducibility of the model and it's results. The model description follows the ODD protocol detailed in Railsback and Grimm [33]. This section contains references to and discussion about model parameters. For convenience, a term in *italics* in this section references a model parameter or agent attribute, who's initial value can be found in Appendix B.

3.2.1 Purpose

The purpose of this model is to simulate different single-winner election systems and the opinion dynamics of voters and candidates in those systems. The goal is to compute the variance among agent opinions to quantify the degree of polarization present in each election system.

3.2.2 Entities, State Variables, and Scales

The model consists of two types of agents, Voters and Candidates. Let V be the set of Voters and C be the set of Candidates. Each agent is defined primarily by an opinion vector, such that every Voter i has an opinion $x_i \in [0,1]^d$, and every Candidate j has an opinion $y_j \in [0,1]^d$, where d is the dimensionality of opinion space (*Number of Opinions*). While each agent's opinion vector is its primary descriptor for the purposes of studying opinion dynamics, there are also other attributes used to define actions agents take at each time step which are listed in Table B.1.

Table B.2 lists the model parameters that are used to both initialize the model and define processes during the run.

Spatially, the model and agents exist in continuous $[0,1]^d$ opinion space, where each dimension represents an opinion an agent may vary along and taken together they form an opinion vector $x_i \in [0,1]^d$ that fully describes the agents opinion. Temporally, the model operates in discrete time steps t. The length of a run is defined by the *Number of Time Steps*. As will be discussed more in Section 3.2.3, elections happen periodically such that there are some constant number of movement rounds before an election.

3.2.3 Process Overview and Scheduling

As previously mentioned, the model operates in discrete time steps, starting at 0, which can either be election rounds, denoted T_e or movement rounds, denoted T_m. Election rounds happen periodically as determined by the *Number of Rounds Before Election*. At each time step $t \in T_m$, the following processes happen in order:

1. A single Voter is randomly selected to activate.
2. The activated Voter moves according to either a Bounded Confidence (Eq. (3.1)) or Attraction-Repulsion (Eq. (3.2)) opinion update.
3. Every 20^{th} movement round, all candidates $c_i \in C$ move.

At each time step $t \in T_e$, the following processes happen in order:

1. A movement round happens.

2. Candidates who have reached their term limit are removed from the model.
3. A winner is chosen according the election system specified by the *Election System*.
4. Candidates stochastically exit the model.
5. Candidates stochastically enter the model.
6. Data is collected.

The mathematical description of each of these processes is detailed in Section 3.2.7.

3.2.4 Design Concepts

While Voter agents are not adaptive in the sense of changing their behavior based on the environment, Candidate agents are. This is discussed in detail in Section 3.2.7, but the general idea is that Candidates move in a way to increase their expected performance in the election, which is a function of the Voter opinion distribution. Candidate movement is therefore a direct objective-seeking adaptive trait, as they move to increase an objective function. Here, we call this an adaptive trait because it depends on the environment, i.e. the position of all other Voters in the system, so Candidate movements adapt to their environment.

Candidates do not learn to improve their adaptive traits over the course of the run, but they do experience a change in their traits as a result of model processes, specifically elections. This will again be discussed more in Section 3.2.7, but elections have consequences for the candidates in terms of changing traits that can be thought of as political power. Candidates have an *Exit Probability* attribute that describes their probability of staying in the race, and a *Threshold* attribute that describes how far away in opinion space they are able to attract Voters. Their performance in the election determines whether these attributes experience a positive or negative change.

Voter agents interact stochastically with both Voter and Candidate agents based on the Euclidean distance between them in opinion space. While interacting agents know each other's exact opinion vector, the opinion updates themselves are noisy. The model is therefore stochastic in interactions, as the agents that interact and the results of interactions are determined, in whole or in part, randomly. The model is also stochastic in elections, as Voters do not vote deterministically, but instead by sampling a probability distribution. As a result, Candidate movements are not to increase their performance, but rather their expected performance. This reflects the idea that in real world political processes, candidates rarely, if ever, know the true opinions of voters. Instead, they must infer them based on public polling and adapt their platform accordingly.

These concepts were designed to show that the emergent behavior of the model, i.e. polarization, arises from agent level interactions and not system level rules. Polarization, which we measure as the variance among opinions, depends on agent locations in opinion space. Where agents are located in opinion space depends on their movement functions, which are detailed in Section 3.2.7, and are not explicitly defined at the system level. Therefore, polarization is an emergent behavior that results from low level agent interactions.

3.2.5 Initialization

Table B.2 lists the default initial values for model parameters at the start of the run, which were chosen in line with previous literature when possible. When there was not adequate previous literature for a given parameter, it was chose heuristically. Some agent attribute values may change over the course of a run, such as the *Threshold* value for Candidates. In these cases, the attributes are initialized to the initial value listed in Table B.2, e.g. *Initial Threshold*. Table B.3 lists the modified parameters used for specific runs discussed in Chapter 4.

3.2.6 Input Data

Our model does not use any input data. In this context, input data refers to information gathered and stored outside of the model that the model references and uses during its processes.

3.2.7 Submodels

This section details the specific functioning of each of the processes laid out in Section 3.2.3.

Agent Activation

The first process to happen at each time step is a single Voter is randomly chosen to be activated for that round. The purpose of selecting these subsets is for the next process, agents movements, detailed in Section 3.2.7.

Agent Movements

The second process to happen at each time step t is agent movement. The single activated Voter moves first, followed by all Candidates in a random order. The activated Voter moves by updating their opinion in two stages. In the first stage, they update their opinion according to an interaction with another randomly selected Voter, and in the second stage they update their opinion according to an interaction with a randomly selected Candidate. How they update their opinion as a result of each interaction depends on the function specified in the *Voter-Voter Interaction Function* and *Voter-Candidate Interaction Function* model parameters. In the bounded confidence model, as developed by Deffuant et al. [7], Krause [21], and Weisbuch et al. [38], Voters update their opinion according to:

$$x_i^{t+1} = \begin{cases} x_i^t + \mu(x_j^t - x_i^t), & \text{if } ||x_i^t - x_j^t|| \leq T \\ x_i^t, & \text{otherwise} \end{cases} \tag{3.1}$$

where x_i^t is the opinion vector of agent i at time step t, μ is a model parameter controlling the magnitude of opinion updates and T is the *Initial Threshold*. When applied to Voter-Candidate interactions, the update is slightly different. First, these updates are one sided in that only the Voter updates their opinion; the Candidate does not move. Second, instead of using the *Initial Threshold* value as T, each Candidate has their own threshold value which is used instead. This is meant to represent political power, whereby more powerful Candidates are able to reach larger audiences. In the Attraction-Repulsion Model, as developed by Axelrod, Daymude, and Forrest [3], Voters update their opinion according to:

$$x_i^{t+1} = \begin{cases} x_i^t + \mu(x_j^t - x_i^t), & \text{if } p \leq (\frac{1}{2})^{d/E} \text{ and } ||x_i^t - x_j^t|| \leq T \\ x_i^t - \mu(x_j^t - x_i^t), & \text{if } p \leq (\frac{1}{2})^{d/E} \text{ and } ||x_i^t - x_j^t|| > T \\ x_i^t, & \text{otherwise} \end{cases} \tag{3.2}$$

where E is the model parameter *Exposure* and $p \sim U(0, 1)$. Similarly to Eq. (3.1), when applied to Voter-Candidate interactions, the value of T in Eq. (3.2) is the specific Candidate's threshold.

Voter movements are therefore stochastic processes which are determined by the random order in which Voter agents are activated, the random order in which they interact with Candidates, and the random order in which they interact with other Voters. After Voters move, Candidates move strategically to maximize their expected performance in the election. This is done as a function of their expected votes. To calculate vote probabilities, we first define the function:

$$f(x_i,\ y_j) = \frac{1}{1 + e^{\gamma(||x_i - y_j|| - r)}} \tag{3.3}$$

to represent the amount of support Voter i gives to Candidate j. Here, x_i and y_j are the opinion vectors of i and j, respectively, γ is the model parameter *Gamma* and r is the model parameter *Radius*. This function describes an area of support around v_i, dependent on r and γ, where candidates within this area receive more support while candidates outside this area receive less. The parameter γ controls the steepness of the decrease in support, while r controls the inflection point, i.e. the radius of the area. We apply the *softmax* function to Eq. (3.3) to generate the vote probability distribution between each Voter and Candidate:

$$\mathbb{P}(\mathrm{v}_i \text{ votes for } \mathrm{c}_j) = \frac{e^{\beta f(x_i,\, y_j)}}{\sum_{j' \in C} e^{\beta f(x_i,\, y_{j'})}} \tag{3.4}$$

where C is the set of all Candidates and β is a model parameter. Here, β is an inverse temperature parameter which controls the variation of the probabilities. The probability that Voter i votes for Candidate j is therefore higher when i and j are closer in opinion space. The higher the value of β, the higher the variation in vote probabilities, and therefore the more the vote resembles deterministic voting based on Eq. (3.3). The lower the value of β, the lower the variation, and therefore the more the vote resembles taking a random sample of C with equal probability.

To calculate candidate movements that increase their winning probability, we define Candidate objective functions for each election system that represent how well a Candidate is expected to perform in an election. For plurality, the objective function is defined as:

$$g_p(X,\, y_j) = \sum_{i \in V} \mathbb{P}(\mathrm{v}_i \text{ votes for } \mathrm{c}_j) \tag{3.5}$$

$$= \sum_{i \in V} \frac{e^{\beta f(x_i,\, y_j)}}{\sum_{j' \in C} e^{\beta f(x_i,\, y_{j'})}} \tag{3.6}$$

where X is the matrix of all Voter opinion vectors and y_j is the opinion vector of Candidate c_j. Eq. (3.5), therefore, gives the expected votes of Candidate j given their current opinion y_j and all current Voter opinions X. The ranked choice objective function is defined as:

$$g_{rc}(X,\ y_j) = \sum_{i \in V} \mathbb{P}(\text{v}_i \text{ votes for } \text{c}_j) + s \sum_{i \in V} \mathbb{P}(\text{v}_i \text{ ranks } \text{c}_j \text{ second}) \tag{3.7}$$

$$= \sum_{i \in V} \left(\frac{e^{\beta f(x_i,\, y_j)}}{\sum_{j' \in C} e^{\beta f(x_i,\, y_{j'})}} + s \sum_{j' \in C \setminus \{j\}} \frac{e^{\beta f(x_i,\, y_{j'})}}{\sum_{\tilde{j} \in C \setminus \{j\}} e^{\beta f(x_i,\, y_{\tilde{j}})}} \frac{e^{\beta f(x_i,\, y_j)}}{\sum_{\tilde{j} \in C \setminus \{j'\}} e^{\beta f(x_i,\, y_{\tilde{j}})}} \right) \tag{3.8}$$

where s is the *Second Choice Weight Factor*. Similar to the plurality objective function, Eq. (3.7) gives the expected first choice votes of Candidate j, as well as a weighted value of the expected second choice votes, given their current opinion y_j and all current Voter positions X.

Candidate movements are defined to heuristically increase these objective functions, and therefore maximize their expected performance in the election. Candidates move by updating their opinion according to:

$$y_j^{t+1} = y_j^t + \alpha \nabla_{y_j} g(X, y_j) \tag{3.9}$$

where α is the *Candidate Learning Rate* and $\nabla_{y_j} g(X, y_j)$ is the gradient of the candidate objective function with respect to the Candidate's position vector y_j. We calculate the gradient using automatic differentiation provided by the Python package *PyTorch* [30]. Candidates move, therefore, by performing gradient ascent on the objective function, which itself gives their expected performance in an election. In the plurality objective function, Candidates move towards areas of higher expected numbers of first choice votes. When moving with respect to the ranked-choice objective function, Candidates move towards areas of lower expected numbers of first and second choice votes. Due to computational resource limitations, Candidates move every 20 rounds instead of every round.

Elections

The first thing to happen during an election round is all candidates who have reached their term limit are removed from the model. Then, an election takes place according to the specific election system, either *plurality* or *instant-runoff ranked choice*. Finally, after the election, Candidates stochastically enter and exit the race. In plurality, the election works as follows:

1. Voters probabilistically cast a single vote for a single Candidate by sampling the probability distribution for all Candidates given by Eq. (3.4).
2. Votes are tallied and the candidate with the most number of votes is chosen as the winner.
3. Candidates update their *Exit Probabilities* and *Thresholds* based on election results.

Candidate places are determined by the number of votes they receive. In instant-runoff ranked choice, the election works as follows:

1. Voters probabilistically rank each candidate by sampling without replacement from the probability distribution for all Candidates given by Eq. (3.4).
2. Votes are tallied based on the number of first choice rankings each candidate receives.

3. If a candidate receives more than a majority of votes, they are declared the winner. If not, the candidate with the fewest number of first choice votes is eliminated, and each ballot that had that candidate as their first choice changes so that the second choice becomes the first choice.

4. Steps 2-3 are repeated until a candidate reaches a majority of votes.

5. Candidates update their *Exit Probabilities* and *Thresholds* based on election results.

Candidate places in the election are determined by runoff-round. The first candidate eliminated places last, the second eliminated places second last, and so on. Then, on the winning round, the candidate with the fewest first choice votes in that round is next last, and so on with the winner placing first. Candidates then update their exit probabilities and thresholds based on the places they finish in the election according to Eq. (3.10).

Candidates update their *Exit Probabilities* and *Thresholds* as a function of their current exit probability and threshold. A Candidate's *Exit Probability* controls whether they remain in the race or exit, and, as discussed in Eq. (3.1), each Candidate's *Threshold* controls how far away they are able to attract Voters. Together, these Candidate attributes represent a form of political power, and updating them according to election results is a form of election consequences. It is politically beneficial for each Candidate to have a low exit probability and a high threshold, since that gives them a higher probability of staying in the race and attracting more Voter's to their position. Candidates update these attributes according to the function:

$$S_o(k,\ p,\ c) = \frac{1}{1 + e^{-(kp - \ln(\frac{1-c}{c})}} \tag{3.10}$$

where k is the factor to change political power by (either the *Exit Probability Decrease Factor* or *Threshold Increase Factor*), c is their current political power attribute, (either *Exit Probability* or *Threshold*), and p is a function of how they performed in the election. When applied to exit probabilities, p is the Candidate's place in the election (starting from 0) minus the *Number of Candidates to Benefit*. When applied to thresholds, p is the *Number of Candidates to Benefit* minus the candidate's place in the election. The following three statements hold (see Appendix A for proofs) for Eq. (3.10) for some constant $c \in (0, 1),\ k \in \mathbb{R}^+, p \in \mathbb{Z}$:

1. When $p = 0$, $S_o(k,\ p,\ c) = c$
2. When $p > 0$, $S_o(k,\ p,\ c) > c$
3. When $p < 0$, $S_o(k,\ p,\ c) < c$

Therefore, the *Number of Candidates to Benefit* parameter controls which places in the election experience negative results, which place experiences no change, and which places experience positive results.

Data Collection

The last thing to happen at each round, both movement and election, is data is collected from that round. Table B.4 lists the data points collected for each agent at each election round. This data is then exported as a CSV file to be analyzed.

Hardware

The election simulations were run on Middlebury College's high performance computing cluster (HPCC), called Ada.

CHAPTER 4

RESULTS

To quantify the difference in polarization between plurality and instant-runoff ranked choice election systems, we ran and analyzed models with the parameters listed in Table B.2. For each election system, we ran both Bounded Confidence and Attraction-Repulsion opinion updates, giving a total of four run types. For each type, we ran the model 1,000 times and the variance at each step was averaged across all 1,000 runs of that type. Similar to Axelrod, Daymude, and Forrest [3], we define the polarization of the model at time step t as the variance between all agent opinions at t. Since variance can only be calculated in 1-dimension, each of these runs was done in 1-dimensional opinion space. Fig. 4.1 shows the average variance at each time step for each run type.

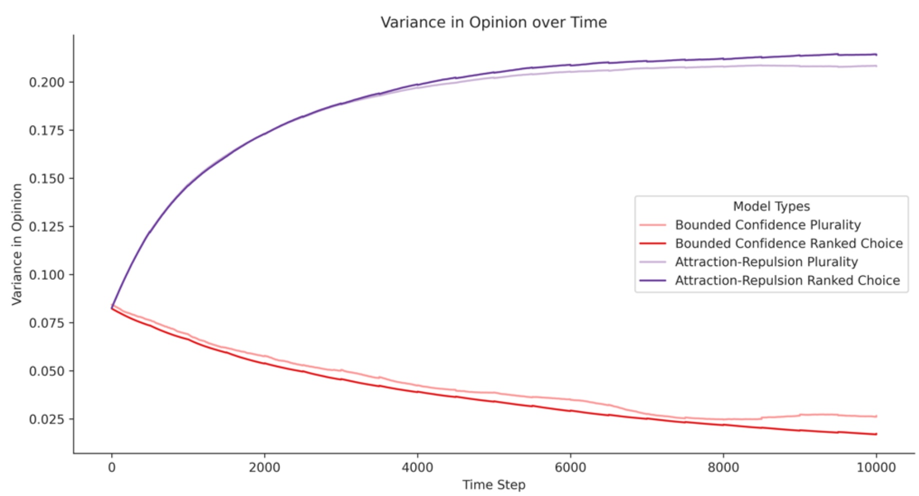


Figure 4.1: The variance at each time step for the four run types with parameters listed in Table B.2, averaged across 1000 runs.

The average Attraction-Repulsion run ended in polarization, whereas the average Bounded Confidence run ended near consensus. Due to Voter's noisy opinion updates and the stochastic introduction of new Candidates at random opinion values after each election, runs won't end in perfect consensus or perfect polarization. Table 4.1 lists the average ending variance for each run type. In Bounded Confidence opinion update models, ranked choice led to a 35% decrease in variance compared to plurality. In attraction-repulsion opinion update models, ranked choice led to a 3% increase in variance compared to plurality.

Run Type	**Ending Variance**
Bounded Confidence Plurality	0.0266
Bounded Confidence Ranked Choice	0.0173
Attraction-Repulsion Plurality	0.208
Attraction-Repulsion Ranked Choice	0.214

Table 4.1: The average ending variance over 1000 runs for each run type. The parameters used are listed in Table B.2.

To emphasize the differences in model behavior between election system, 6 parameters were changed: *Number of Time Steps*, *Term Limit*, *Threshold Increase Factor, Candidate Learning Rate, μ, and Second Choice Weight Factor*. Their values are listed in Table B.3. Fig. 4.2 shows the behavior of the model with these changed parameters. Table 4.2 lists the average ending variance for each run type, showing that instant runoff ranked choice election systems with Attraction-Repulsion opinion updates were, on average, 95% less polarized than plurality systems.

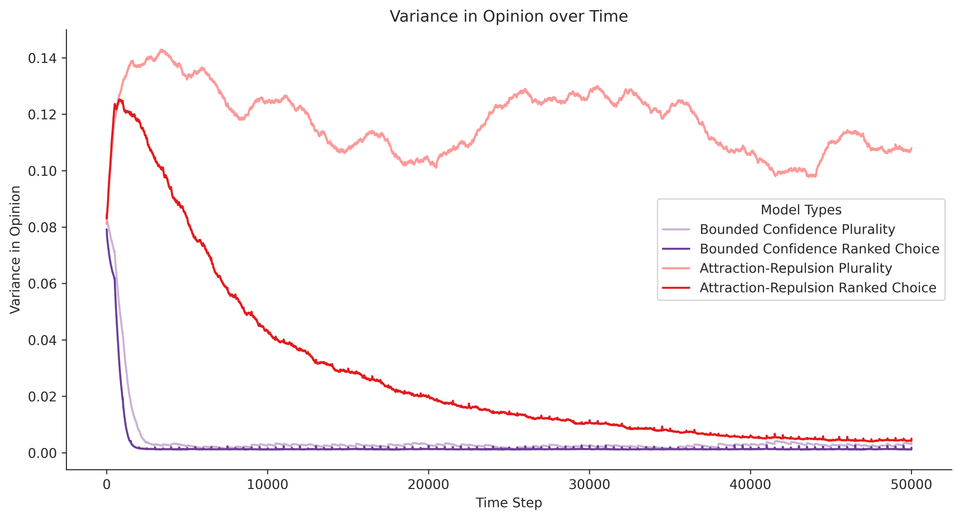


Figure 4.2: The variance at each time step for the four run types with parameters listed in Table B.3, averaged across 1000 runs.

Run Type	Ending Variance
Bounded Confidence Plurality	0.00290
Bounded Confidence Ranked Choice	0.00220
Attraction-Repulsion Plurality	0.1104
Attraction-Repulsion Ranked Choice	0.00939

Table 4.2: The average ending variance over 100 runs for each run type. The parameters used are listed in Table B.3.

4.1 Single Run Examples

In this section, we will highlight and define different agent and model behaviors by visually analyzing opinion dynamic plots from single runs. These opinion dynamics plots show how the distribution of agent opinions changes over the course of a run. Voter agents are colored gray, Candidate agents are colored purple, and winning Candidate agents are colored orange. The size of the line corresponds to a scaled version of that agent's *Threshold* (Eq. (3.1), Eq. (3.2)); larger Candidates can be considered more "powerful". Since this analysis comes from single run examples, it should be considered as anecdotal evidence of agent and model behaviors. That is, while it is true that these models can produce this behavior, it may or may not be true that these behaviors are common in runs or represent the average behavior.

4.1.1 Intermediate Candidate Effect

In some runs, particularly Bounded Confidence opinion update runs, Candidates are able to connect two disparate opinion clusters. Fig. 4.3 shows such a run, where starting around $t = 2000$, Voters begin to diverge into two disparate opinion groups. Some agents do crossover, but by around $t = 6000$, all movement between groups has stopped. Shortly after that, however, a Candidate enters the model with an opinion near the middle of the two groups. This Candidate facilitates movement between the two groups by interacting with Voters and attracting them to within the threshold of the other group, at which point the two groups converge. This is not always the case, however, as Fig. C.1a and Fig. C.1b show.

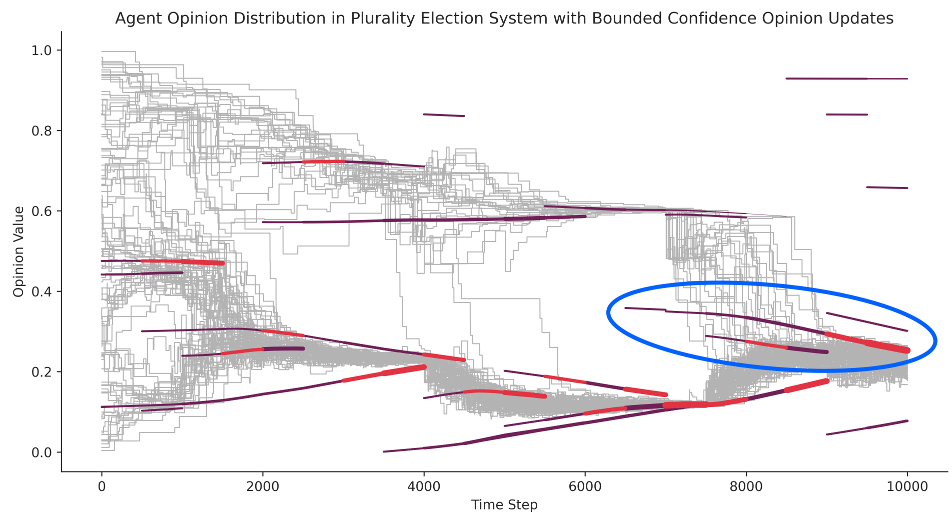


Figure 4.3: An intermediate Candidate is able to connect two disparate opinion clusters. Voters are colored gray, Candidates are colored purple, winning Candidates are colored orange, and the size represents the Candidates threshold (Eq. (3.1), Eq. (3.2))

.

4.1.2 Ranked Choice Allows for Compromise

Ranked choice systems allow for compromising policies in a way that plurality systems may not. Fig. 4.4 provides an example of this, where in the Attraction-Repulsion ranked choice system it is possible for a centrist Candidate to win the election in a highly polarized system. In this run, the Voter population quickly polarizes into two groups at opposite ends of opinion space, and the initial Candidates follow them. At around $t = 4,500$, several new Candidates enter the race occupying centrist positions. They are not inclined to move much towards either group as they are in an area of high numbers of expected second choice votes. Despite not being near any Voter agents, a centrist Candidate is able to win election at $t = 6,500$, while occupying a position between the two polarized groups.

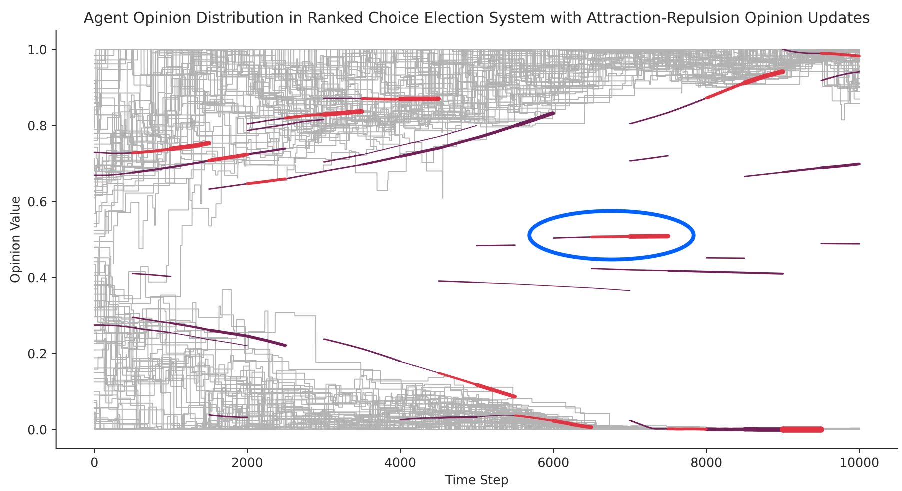


Figure 4.4: A centrist Candidate with a centrist position between two polarized groups is able to win election in a ranked choice system. Voters are colored gray, Candidates are colored purple, winning Candidates are colored orange, and the size represents the Candidates threshold (Eq. (3.1), Eq. (3.2))

.

4.1.3 Extreme and Powerful Candidate Effect

In some cases, a Candidate occupying an ideologically extreme position is able to become politically powerful, despite not winning an election, through a series of second place finishes. Once they become powerful, they exercise a large influence on the Voter's nearby them, which is shown through their ability to change the opinions of those Voters. In one example of this, shown in Fig. 4.5, the Voters have formed two opinion clusters, around 0.4 and 0.8 by $t = 7{,}000$. At around $t = 8{,}000$, the Voters begin to converge. However, instead of converging around a mean opinion of the two groups as would usually be the case, they instead converge around the mean opinion of the more ideologically extreme group due to the political influence of this extreme Candidate.

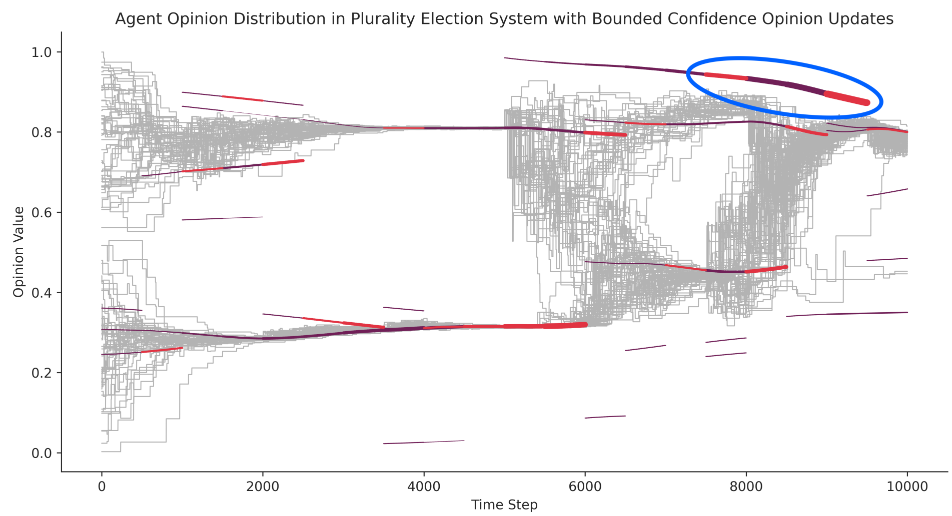


Figure 4.5: An ideologically extreme and powerful candidate is able to exert a large influence on nearby Voters, effectively changing the the dynamics of that group. Voters are colored gray, Candidates are colored purple, winning Candidates are colored orange, and the size represents the Candidates threshold (Eq. (3.1), Eq. (3.2))

.

4.1.4 Centrist Candidates Hold the Middle

In Attraction-Repulsion systems, which tend to polarize quickly and acutely, it is possible for a group of centrist Voters to not polarize if there are also centrist

Candidates present. Fig. 4.6 shows one such example. Up to around time step 5,000, there is a group of four Candidates, three of which are relatively influential, all centered around an opinion value of 0.6. Despite polarized opinion groups forming on either side of them, this centrist group remains in the middle with a core voter bloc. However, once all but one of these Candidates have exited the race ($t = 5000$) the centrist Voter group begins to dissolve. The single remaining Candidate is unable to hold the middle group any longer, and by the end of the run, all Voters have moved to one of the polarized groups. The ability of this centrist group of Candidates to hold the middle for that long is noteworthy, as most Attraction-Repulsion runs have reached polarization by time step 5,000 (Fig. 4.1).

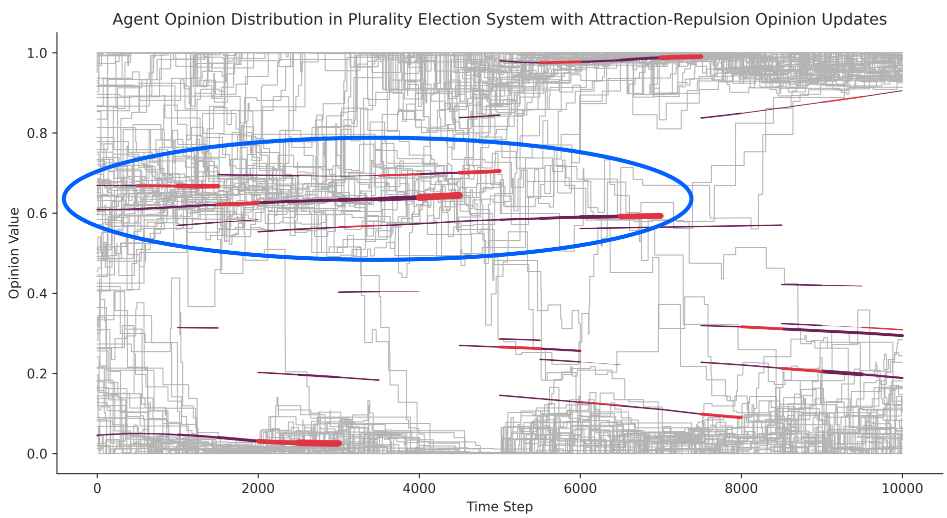


Figure 4.6: An influential group of centrist Candidates is able to extend the existence of a centrist group of Voters. Voters are colored gray, Candidates are colored purple, winning Candidates are colored orange, and the size represents the Candidates threshold (Eq. (3.1), Eq. (3.2)).

4.2 Parameter Sweeps

To study the effect of varying parameter values on the ending opinion variance, we chose four parameter combinations based on their perceived impact on model variance: *Initial Threshold* and μ, β and *Second Choice Weight Factor*, *Exit Probability Decrease Factor* and *Threshold Increase Factor*, and γ and *Radius*. Each parameter sweep was done with the four model types, and the ending variance for each parameter combination was averaged between 20 runs. It is important to note that Bounded Confidence runs have a different color scale than Attraction-Repulsion runs since the average Bounded Confidence run ends with low variance, whereas the average Attraction-Repulsion run ends with high variance.

4.2.1 *Initial Threshold* and μ

First, to understand the role of parameters controlling agent opinion updates, a sweep was done between *Initial Threshold* and μ, shown in Fig. 4.7. *Initial Threshold* controls the area around an agent in opinion space within which they positively interact with other agents, and μ controls the magnitude of the opinion update (Eq. (3.1), Eq. (3.2)). The ending variance was largely the same across election system, but differed between opinion update function, dependent mostly on the *Initial Threshold* value rather than μ. Bounded Confidence runs ended with near-zero variance for *Initial Thresholds* greater than about 0.25. A plausible explanation as to why the models variance is almost entirely dependent on the *Initial Threshold* and not μ is that what the model converges to depends on the *Initial Threshold*, whereas the rate of convergence depends on μ, and the runs are sufficiently long enough to see the convergence.

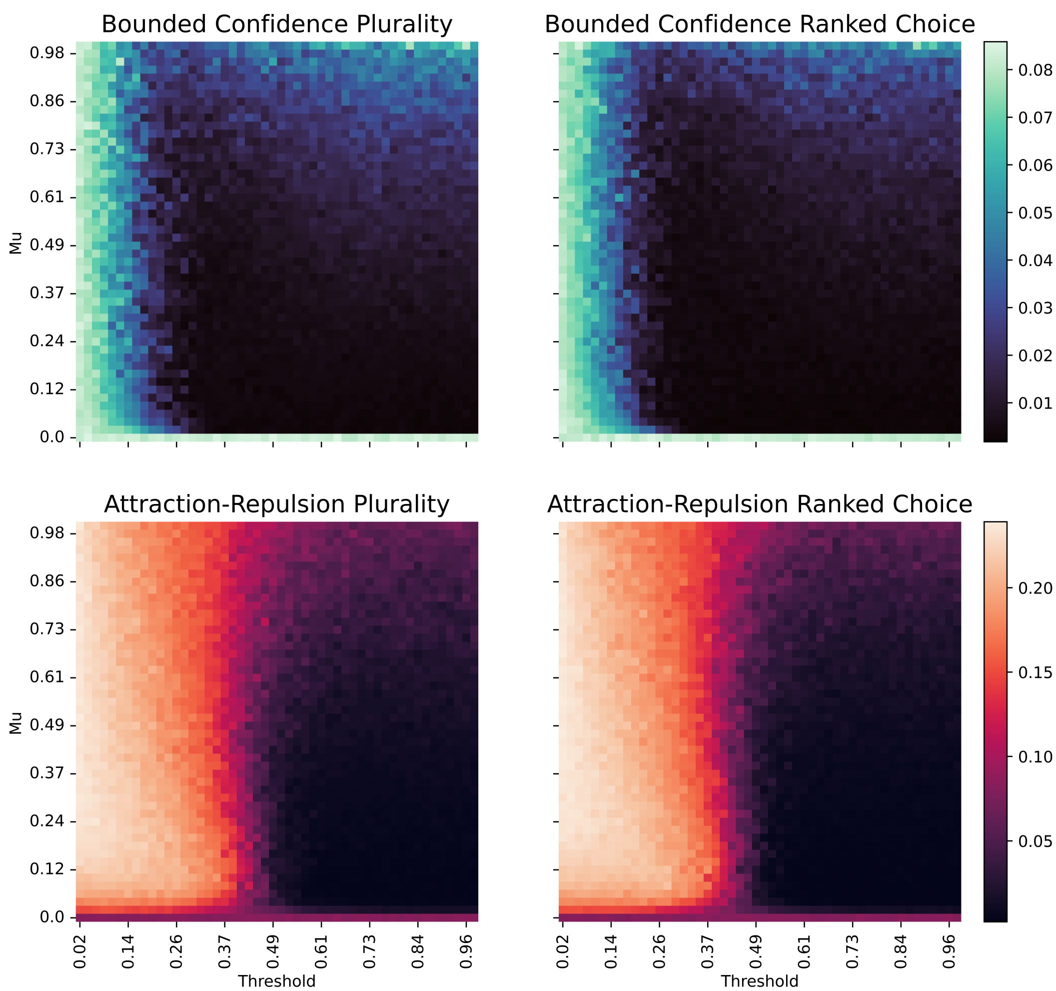


Figure 4.7: The ending variance for different combinations of the *Initial Threshold* and μ parameters, averaged across 20 runs for each combination. Note the different color scale for each opinion update function. All other parameters are listed in Table B.2.

For these higher thresholds, increasing μ had the effect of slightly increasing the variance. Attraction-Repulsion runs transitioned from high variance to low variance around a threshold of 0.5, and similarly to Bounded Confidence runs, increasing μ for higher thresholds slightly increased the ending variance. For higher threshold values, however, larger values of μ increase the models polarization slightly. One possible explanation is that large, noisy opinion updates cause agents to overshoot the update, moving *past* the agent they are supposed to be moving closer to in opinion space. When this is the case, changing opinions in smaller increments (smaller μ) causes less polarization.

4.2.2 β and *Second Choice Weight Factor*

Then, to understand the role of parameters controlling Candidate movement, a sweep was done between β and *Second Choice Weight Factor*, s. As discussed in Section 3.2.7, β controls the variation of the probability distribution for a Voter's vote. Higher values of β create more variation, so when a vote is taken as a sample from this distribution it more often resembles deterministic voting based on distance in opinion space (or more specifically based on Eq. (3.3)). *Second Choice Weight Factor* controls how much value a Candidate's expected number of second choice votes contributes to their objective function (Eq. (3.7)). Since the Candidate's ranked choice objective function depends on both of these parameters, the gradient computed with respect to a Candidate's position also depends on these parameters. Therefore, different values will cause different gradients to be computed, which will cause different Candidate movements (Eq. (3.9)). However, since the candidates plurality objective function depends only on β, changing the *Second Choice Weight Factor* in this system has no effect.

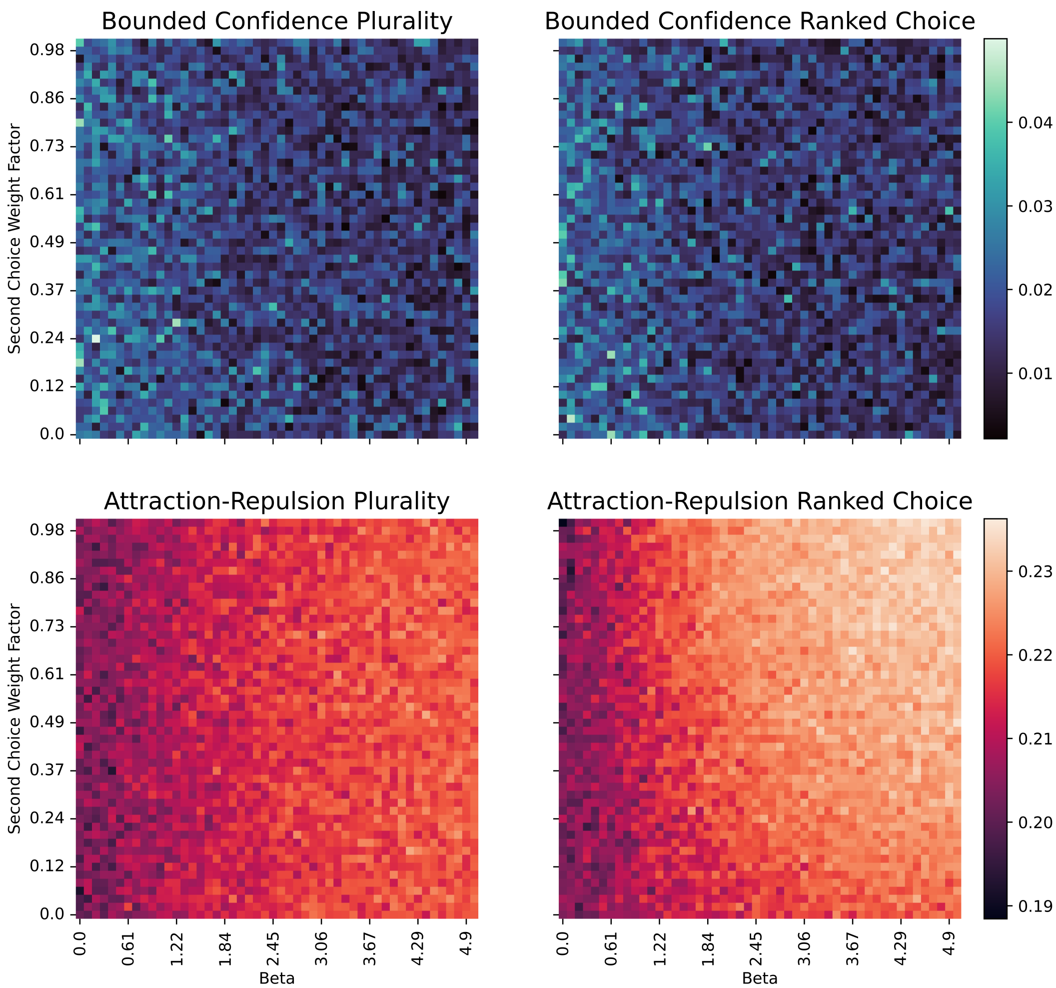


Figure 4.8: The ending variance for different combinations of the *Beta* and *Second Choice Weight Factor* parameters, averaged across 20 runs for each combination. Note the different scale for each opinion update function. All other parameters are listed in Table B.2.

Fig. 4.8 shows the average ending variance for different parameter combinations in each run type. In plurality systems, the ending variance is invariant with respect to the *Second Choice Weight Factor*, as changing that parameter has no effect on Candidate movement in plurality systems. In both Bounded Confidence runs, the ending variance was also largely invariant with respect to β. In Attraction-Repulsion runs, however, increasing β increased the ending variance, and in the ranked choice system, increasing the *Second Choice Weight Factor* also slightly increased the variance, though to a lesser extent than β. This phase change can be explained by the fact that β controls the variation in vote probabilities. Higher values of β create more variation, so when Candidates move by performing gradient ascent on these probabilities (Eq. (3.9)), the gradients they compute will be larger, so the movements themselves will also be larger. Therefore, in Attraction-Repulsion models, which polarize under the default parameters (Table B.2 and Fig. 4.1), higher β values allow Candidates to more quickly join the polarized groups.

4.2.3 *Exit Probability Decrease* and *Threshold Increase Factors*

To understand the role of parameters controlling Candidate political power, a sweep was done between *Exit Probability Decrease Factor* and *Threshold Increase Factor*. As discussed on Section 3.2.7, Candidates each have an *Exit Probability* and *Threshold* which are updated after elections take place according to Eq. (3.10). Fig. 4.9 shows the average ending variance for different combinations in each run type.

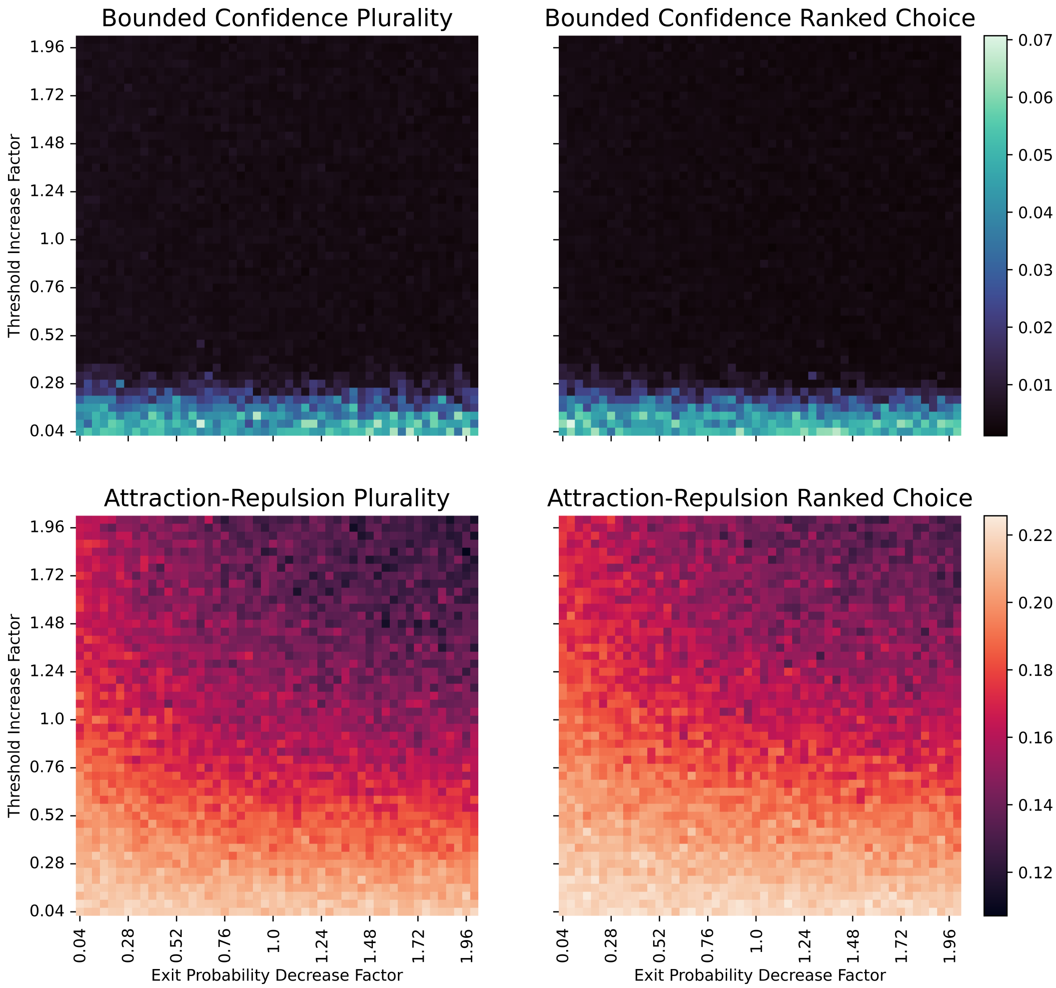


Figure 4.9: The ending variance for different combinations of the *Exit Probability Decrease Factor* and *Threshold Increase Factor* parameters, averaged across 20 runs for each combination. Note the different scale for each opinion update function. All other parameters are listed in Table B.2.

In Bounded Confidence runs, the ending variance was largely the same between plurality and ranked choice election systems. In such runs, the variance transitioned from low (0.05), to near-zero around a threshold increase factor of 0.3. The ending variance was invariant with respect to the *Exit Probability Decrease Factor*. In Attraction-Repulsion runs, the transition pattern between high and low variance was also largely the same between election systems, except that in ranked choice it happened at slightly higher values. The transition from high variance to low variance was mostly dependent on the threshold increase factor, with values below around 0.5 in plurality and 0.8 in ranked choice defining the phase change. For sufficiently high values of threshold increase factor, larger exit probability decrease factors led to lower variance.

The phase change seen for Bounded Confidence runs when varying the *Threshold Increase Factor* and the *Exit Probability Decrease Factor* is explained mostly by the *Threshold Increase Factor*. In these runs, a sufficiently high *Threshold Increase Factor* allows the winning Candidate to become powerful enough to attract everyone to their position, creating consensus. In Attraction-Repulsion models, the phase change is less clear. The general trend, however, shows that the consequences that the degree of change of Candidate political power has on the polarization of the model is similar in shape between election systems, though plurality systems see a greater effect that begins at lower values.

4.2.4 γ and *Radius*

Finally, to understand the role of parameters defining the radius of support function (Eq. (3.3)), a sweep was done between γ and *Radius*. The radius of the support function, as discussed in Section 3.2.7, defines an area around a Voter v_i where

Candidates within this area receive more support. This function is used to generate the vote probability distribution between each Voter and Candidate (Eq. (3.4)), which in turn defines the Candidates objective functions. γ controls the "steepness" of the transition between the area of high support and low support, and the *Radius* controls the size of this area, i.e. the location of the inflection point. A high value of γ means that the rate of change of support is high near the radius but low further away. Therefore, for high γ values, the support Candidate c_j receives from v_i depends mostly on whether they are inside or outside the radius. For low γ values, the support Candidate c_j receives from v_i depends mostly on the distance between c_j and v_i.

As Fig. 4.10 shows, in Bounded Confidence runs, intermediate values of *Radius* with higher values of γ cause low variance, while low and high values of *Radius* cause higher variance, regardless of γ. The shape of the variance pattern was similar between plurality and ranked choice systems, but the area of low variance was larger in ranked choice. Specifically, $0.36 < \textit{Radius} < 0.7$, and $\gamma > 4.5$ caused low polarization in ranked choice systems, while $0.4 < \textit{Radius} < 0.55$, and $\gamma > 6$ caused low polarization in plurality systems. In Attraction-Repulsion plurality systems, lower *Radius* and higher γ values generally cause more variance, while higher *Radius* and higher γ values generally cause less variance. The pattern is again similar in ranked choice, albeit slightly more varied. As well, there exists an island of higher variance at intermediate radii and high γ.

The phase change seen in Bounded Confidence runs when varying the γ and the *Radius* parameters in Eq. (3.3) can be explained by how γ changes the func-

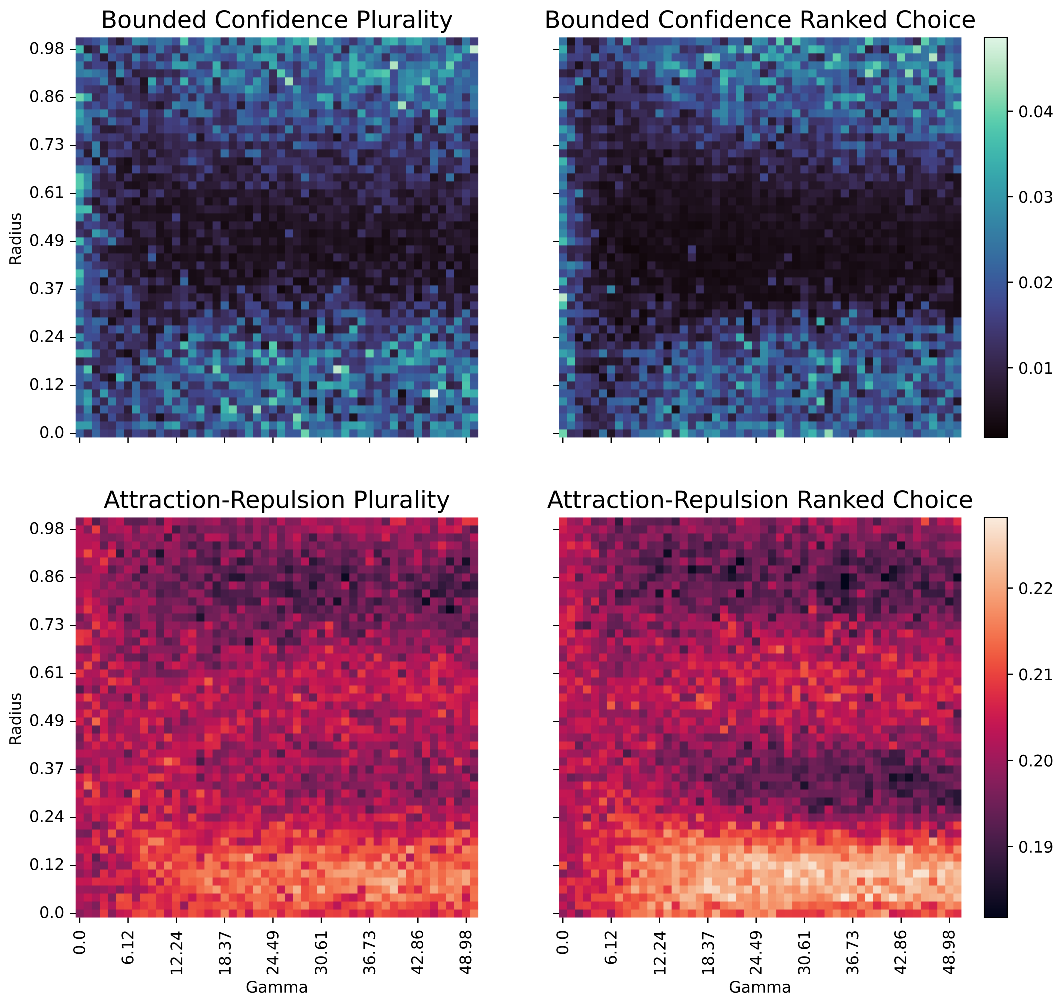


Figure 4.10: The ending variance for different combinations of the γ and *Radius* parameters, averaged across 20 runs for each combination. Note the different scale for each opinion update function. All other parameters are listed in Table B.2.

tion defining the support a Voter gives to a Candidate (Eq. (3.3)). For small γ values, the support a Candidate receives from a Voter is largely dependent on the euclidean distance between them in opinion space. Due to that, the support Candidates receive changes even with slight movements, so Candidates will compute a large enough gradient for all values of *Radius* to be able to move to the group of Voters that have formed consensus. Conversely, for large γ values, the support a Candidate receives from a Voter is largely dependent whether they are inside or outside the *Radius*. For low and high *Radius* values, once the Voters have formed consensus, as they do in these models (Fig. 4.1), most, if not all, Voters will be either inside or outside the *Radius*. Since the amount of support a Voter gives a Candidate is almost constant inside and outside the *Radius* for large γ values, Candidate movements will cause little to no change in their expected support. This will make the computed gradients near zero, limiting the ability of Candidates to move towards the single group of Voters and form consensus.

Similarly, the phase change seen in Attraction-Repulsion runs when varying the γ and the *Radius* parameters in Eq. (3.3) can be explained by how the *Radius* parameter changes Candidate movements. Under the default parameters (Table B.2), Attraction-Repulsion runs polarize (Fig. 4.1). This means that, on average, the Voter clusters are further apart than Bounded Confidence. For low *Radius* and high γ values, Candidates are able to follow the Voters to the poles, creating extreme variance (Fig. C.2). For medium-low *Radius* and high γ values, Candidates are still able to follow the Voters to the poles. However, because of the larger *Radius* and high γ values, the Candidates move close to poles but are not incentivized to move all the way to the extreme since the Voters are all within the *Radius* and giving almost maximal support, creating lower variance

(Fig. C.3). For medium-high *Radius* and high γ values, Candidates are actually incentivized to move towards the middle, as most Voters are already within the *Radius*. The Voters still move to the poles, however, and since the Candidates are too far away they cannot attract the Voters away from the poles, creating high variance (Fig. C.4). Finally, for high *Radius* and high γ values, Candidates are not incentivized to move at all since all Voters are already in their *Radius*. Therefore, since some are already near the poles, they are able to attract some of the extremely polarized Voters away, reducing the variance (Fig. C.5).

CHAPTER 5

DISCUSSION

Understanding the role election systems play in polarizing political systems is important for finding ways to reduce ideological polarization. While the relationship has been studied before [25, 24], that work has largely looked at the relationship through the lens of elector rule disproportionality. While this can be an effective proxy measurement of the relationship, it does not provide a direct link. To better study the relationship directly, we developed an ABM to simulate elections and political system opinion dynamics.

Our ABM simulates both Voter and Candidate agents in the same system. Voters interact with other agents according to different interaction rules, specifically Attraction-Repulsion [3] and Bounded Confidence [7, 21, 38], which cause them to update their opinions. Candidates update their opinions to improve their expected performance in the election. The model also simulates periodic elections, which are either instant-runoff ranked choice or plurality. The emergent behavior of this model, polarization, arises from the low level interactions of agents. The results of these elections change Candidate attributes, which can be thought of as changing their political power. The different interaction functions and election systems, therefore, give four possible types of simulations.

We analyzed the opinion dynamics of each of these simulation types to quantify the degree of polarization present. Similar to Axelrod, Daymude, and Forrest [3], we defined the degree of polarization present in a system as the variance among all agent opinions. With the parameters listed in Table B.2, ranked choice election system models with Bounded Confidence opinion updates were, on average, 35% less polarized than plurality election system models. However, ranked choice election system models with Attraction-Repulsion opinion updates were, on average, 3% more polarized than plurality election system models (Fig. 4.1). With the parameters listed in Table B.3, instant runoff election systems with Attraction-Repulsion opinion updates were, on average, 95% less polarized than plurality systems (Fig. 4.2). Again, we define polarization as the variance all agent opinions, as is done in Axelrod, Daymude, and Forrest [3]. As will be discussed later in Section 5.1, there are different polarization metrics that could have been used which would change the degree of polarization measured.

To get the results seen in Fig. 4.2, 5 parameters were changed, listed in Table B.3. Increasing the *Threshold Increase Factor* allowed Candidates to experience greater changes in political power due to elections, and increasing the *Term Limit* allows those changes to influence the model for longer. Increasing the *Second Choice Weight Factor* from 0.5 to 100, though unrealistic, causes a drastic change in Candidate movements in only ranked choice systems (Eq. (3.7), Eq. (3.9)), and increasing the *Candidate Learning Rate* allows the Candidates to move faster (Eq. (3.9)). Finally, increasing the *Number of Time Steps* allows the long term model behavior to play out. Therefore, changing the model behavior to show a greater difference between election systems depends largely on changing Candidate movements and the effects Candidates experience as a result of election performance. While these changes may be unrealistic (a *Second Choice Weight Factor* of 100 means that Candidates value second choice votes 100 times more than first choice votes), they may also be specific to this model and not generalize to real world election systems.

As Table B.2 shows, there are 22 real-valued parameters for the model. Understanding the complete dynamics of how all of these parameters interact to change the model behavior is not within the scope of this thesis; we focused on the parameters we believed to be most impactful. One relationship we did not investigate, largely due to computational resource limitations, was that between the *Number of Candidates* and the *Number of Voters*. With 100 Voters and a minimum of 3 Candidates, there were never more than 34 Voters per Candidate. In real world political systems, this ratio can often be millions to one. Perhaps catering to more opinions would account for the Candidates unrealistic vote valuations needed to see opposing behavior in election systems.

The parameter relationships that were investigated, however, provided some insights into model behaviors, as Section 4.2 showed. Varying μ and the *Initial Threshold* produced phase transitions, or locations in parameter space where the model transitions between low and high variance, similar to those seen in Axelrod, Daymude, and Forrest [3], who introduced the Attraction-Repulsion model. The consequence of this phase change (Fig. 4.7) is that whether agents polarize depends on their tolerance for considering dissimilar opinions positively, not on how much they change their opinion by. When opinion updates are noisy, smaller opinion updates reduces the chance of overshooting an update, which can increase polarization. Varying β and the *Second Choice Weight Factor* shows that polarization can depend on the variation in vote probabilities the each Voter has for the Candidates.

The phase changes for the *Threshold Increase Factor* and the *Exit Probability Decrease Factor* parameters implies that polarization can depend on how powerful Candidates become after an election. If a Candidate performs well, they increase their threshold dependent on *Threshold Increase Factor*, and decrease their exit probability dependent on *Exit Probability Decrease Factor*. These updates represent a change in the political power of a Candidate. A high threshold means they are able to attract Voters to their position from farther away in opinion space, and a low exit probability means they are more likely to stay in the race even if they don't win the election. Fig. 4.9 shows that sufficiently high threshold increases can drive polarization to near-zero, as they become powerful enough to bring everyone to their position, creating consensus.

Finally, the phase changes seen for the γ and *Radius* model parameters show the importance that varying levels of support have on polarization. When Voters give different levels of support for each Candidate, the models tend towards low polarization. When Candidates receive similar levels of support, however, the models can tend towards higher polarization as Candidates are disincentivized to move towards the Voters.

These results were, of course, obtained through agent based model simulations. While such models are internally valid, in that their results come from explicit rules and formulas, their main emergent patterns are often only observable across a narrow ranges of conditions [32]. Moreover, in order to abstract a complex dynamic process, such as a real world political system, to an ABM, assumptions were made. Each decision about model implementation was itself an assumption. For example, we simulated 100 Voters and 3 to 5 candidates, whereas many real world political systems often include millions of Voters. We also assumed that Candidates move in elections in ways that maximize the number of votes the expect to receive, which may not always be the case. The applicability of these results to real world political systems depends on the generability of the assumptions made.

5.1 Future Work

One important feature of our model that caused varying behaviors between election systems are the consequences Candidates experience as a result of their election performance. Our model advantaged or disadvantaged Candidates by increasing or decreasing their influence over changing Voter opinions and their likelihood of staying in the race. There are, however, many other ways to advantage or disadvantage Candidates based on their performance. Deciding on a heuristic to model some observable political effect, such as a Candidates charisma, could cause interesting and unexpected changes in the model behavior.

The modeling choice that is arguably most influential to the results we obtained was the decision to quantify polarization as the variance in agent opinions. While this is not a novel approach [3], it is also not the only choice. DeMarzo, Vayanos, and Zwiebel [8] offered one such different method to measure the polarization group opinion formation based on ideological alignment. Musco et al. [28] extended this idea, to provide a class of group-based polarization metrics.

Here, *group-based* refers to the idea of measuring the separation of distinct groups of individuals, rather than just the spread of individuals themselves, as is the case with using variance. The clear advantage of this metric is that it is able to capture the existence of distinct opinion clusters, which variance may not.

Devia and Giordano [9] provides another approach to analyzing the behavior of opinion formation models. Their approach relies on a histogram-based classification algorithm and transition tables, which can classify an opinion distribution as perfect consensus, consensus, polarization, clustering, or dissensus. An advantage of this approach is that this algorithm can be applied to a model generated opinion distribution as well as real world data representing an opinion distribution to determine how well the model matches the process it was intended to simulate. Applying any of these different polarization metrics to our model's results may lead to different polarization measurements for each model.

APPENDIX A

PROOFS

Consider the function:

$$S_o(k,\ p,\ c) = \frac{1}{1 + e^{-(kp - \ln(\frac{1-c}{c}))}} \tag{A.1}$$

A.1 Proof that $S_o(k,\ p,\ c) = c$ when $p = 0$

Let $c \in (0, 1)$ and $k \in \mathbb{R}^+$ be some constants. Let $p = 0$.

$$S_o(k,\ p,\ c) = \frac{1}{1 + e^{-(k(0) - \ln(\frac{1-c}{c}))}} \tag{A.2}$$

$$= \frac{1}{1 + e^{-(0 - \ln(\frac{1-c}{c}))}} \tag{A.3}$$

$$= \frac{1}{1 + e^{\ln(\frac{1-c}{c})}} \tag{A.4}$$

$$= \frac{1}{1 + (\frac{1-c}{c})} \tag{A.5}$$

$$= \frac{1}{\frac{1}{c}} \tag{A.6}$$

$$= c \tag{A.7}$$

A.2 Proof that $S_o(k,\ p,\ c) > c$ when $p > 0$

Let $c \in (0,1)$ and $k \in \mathbb{R}^+$ be some constants. Let $p > 0$.

$$S_o(k,\ p,\ c) = \frac{1}{1+e^{-(kp-\ln(\frac{1-c}{c}))}} \tag{A.8}$$
$$= \frac{1}{1+e^{-(m-\ln(\frac{1-c}{c}))}}, \text{ for some } m > 0 \tag{A.9}$$
$$= \frac{1}{1+e^{-m+\ln(\frac{1-c}{c})}} \tag{A.10}$$
$$= \frac{1}{1+\frac{e^{\ln(\frac{1-c}{c})}}{e^m}} \tag{A.11}$$
$$= \frac{1}{1+\frac{\frac{1-c}{c}}{e^m}} \tag{A.12}$$
$$= \frac{1}{1+(\frac{1-c}{c}\cdot e^{-m})} \tag{A.13}$$
$$= \frac{1}{1+(\frac{1}{c}-\frac{c}{c})e^{-m}} \tag{A.14}$$
$$= \frac{1}{1+(\frac{e^{-m}}{c}-\frac{ce^{-m}}{c})} \tag{A.15}$$
$$= \frac{1}{1+(\frac{e^{-m}}{c}-e^{-m})} \tag{A.16}$$
$$= \frac{1}{1+\frac{x}{c}-x}, \text{ for some } 0 < x < 1 \tag{A.17}$$
$$= \frac{1}{\frac{c}{c}+\frac{x}{c}-\frac{cx}{c}} \tag{A.18}$$
$$= \frac{1}{\frac{c+x-cx}{c}} \tag{A.19}$$
$$= \frac{c}{c+x-cx} \tag{A.20}$$

Since $0 < x < 1$, then $c + x - cx < 1$ for all $c \in (0,1)$. Therefore $\frac{c}{c+x-cx} > c$.

A.3 Proof that $S_o(k,\ p,\ c) < c$ when $p < 0$

Let $c \in (0, 1)$ and $k \in \mathbb{R}^+$ be some constants. Let $p > 0$.

$$\begin{aligned}
S_o(k,\ p,\ c) &= \frac{1}{1+e^{-(kp-\ln(\frac{1-c}{c}))}} && \text{(A.21)}\\
&= \frac{1}{1+e^{-(m-\ln(\frac{1-c}{c}))}},\ \text{for some } m < 0 && \text{(A.22)}\\
&= \frac{1}{1+e^{-m+\ln(\frac{1-c}{c})}} && \text{(A.23)}\\
&= \frac{1}{1+\frac{e^{\ln(\frac{1-c}{c})}}{e^m}} && \text{(A.24)}\\
&= \frac{1}{1+(\frac{1-c}{c}\cdot e^{-m})} && \text{(A.25)}\\
&= \frac{1}{1+\frac{1-c}{c}e^{-m}} && \text{(A.26)}\\
&= \frac{1}{1+(\frac{1}{c}-\frac{c}{c})e^{-m}} && \text{(A.27)}\\
&= \frac{1}{1+(\frac{e^{-m}}{c}-\frac{ce^{-m}}{c})} && \text{(A.28)}\\
&= \frac{1}{1+(\frac{e^{-m}}{c}-e^{-m})} && \text{(A.29)}\\
&= \frac{1}{1+\frac{x}{c}-x},\ \text{for some } x > 1 && \text{(A.30)}\\
&= \frac{1}{\frac{c}{c}+\frac{x}{c}-\frac{cx}{c}} && \text{(A.31)}\\
&= \frac{1}{\frac{c+x-cx}{c}} && \text{(A.32)}\\
&= \frac{c}{c+x-cx} && \text{(A.33)}
\end{aligned}$$

Since $x > 1$, then $c + x - cx > 1$ for all $c \in (0, 1)$. Therefore $\frac{c}{c+x-cx} < c$.

APPENDIX B

MODEL PARAMETERS AND AGENT ATTRIBUTES

Agent Parameters

Attribute	**Voter**	**Candidate**
Type	fixed	fixed
Unique ID	fixed	fixed
Voted For	variable	fixed
Number of Votes	N/A	variable
Is Winner	N/A	variable
Number of Wins	N/A	variable
Exit Probability	N/A	variable
Threshold	N/A	variable
Opinion	variable	variable
Voter Interaction Function	fixed	N/A
Candidate Interaction Function	fixed	N/A

Table B.1: An attribute that is variable means that it may change over the course of the run, whereas a *fixed* attribute retains its initial value over the course of the run. *N/A* means that agent type does not possess that attribute.

Default Model Parameters

Attribute	**Value**
Number of Time Steps	10,000
Number of Voters	100
Number of Voters to Activate	1
Initial Number of Candidates	3
Minimum Number of Candidates	3
Maximum number of Candidates	5
Term Limit	2
Number of Opinions	1
Voter Noise Factor	0.001
Initial Exit Probability	0.33
Exit Probability Decrease Factor	0.25
Initial Threshold	0.35
Threshold Increase Factor	0.1
Number of Candidates to Benefit	2
Number of Rounds Before Election	500
Election System	Plurality or Ranked Choice
Voter-Voter Interaction Function	Bounded Confidence or Attraction-Repulsion
Voter-Candidate Interaction Function	Bounded Confidence or Attraction-Repulsion
μ	0.25
Radius	0.1

Candidate Learning Rate	0.000001
Number of Steps before Ascent	20
γ	10
β	1
Second Choice Weight Factor	0.5
Exposure	0.2

Table B.2: The initial values for the default model parameters used in parameter sweeps and Fig. 4.1.

Modified Model Parameters

Attribute	Value
Number of Time Steps	50,000
Term Limit	5
Threshold Increase Factor	1
Candidate Learning Rate	0.0005
γ	0.5
Second Choice Weight Factor	100

Table B.3: Only the parameters that were changed from Table B.2 are listed with their initial value. These parameters were used in the runs that produced Fig. 4.2.

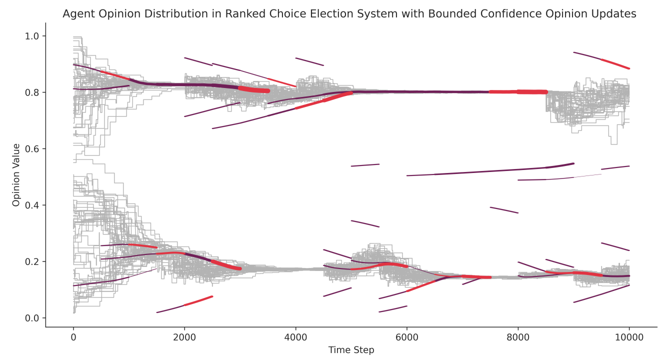
Data Collected from the Model

Data Point	**Type**	**Description**
Voted For	Integer	The ID of the candidate that this agent voted for.
Number of Votes	Integer	The number of votes that this candidate received. NaN value for Voters.
Is Winner	Boolean	A boolean representing whether this candidate was the winner of this election round. NaN for Voters.
Number of Wins	Integer	The number of wins this candidate has up to and including this round. NaN for Voters.
Threshold	Float	Value for the threshold of this candidate. NaN for Voters.
Exit Probability	Float	Value for the exit probability of this candidate. NaN for Voters.
Type	String	'candidate' if the agent is a Candidate, 'voter' if the agent is a Voter.
Opinion(s)	Float(s)	The opinion value of the agent. There are n columns named 'opinion$k$' for each opinion dimension k of opinion space.

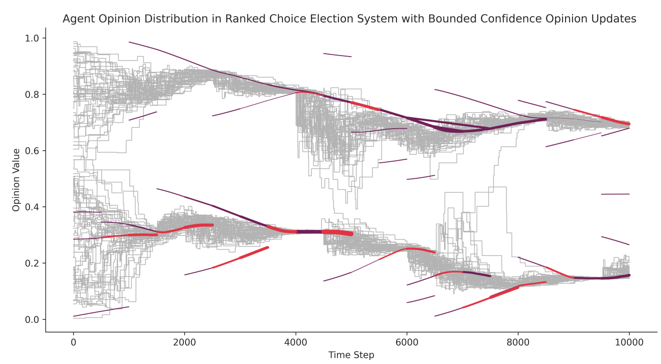
Table B.4: The full list of data points collected at the end of each round.

APPENDIX C

EXTRA FIGURES



(a) If the opinion clusters are too far apart, or if the intermediate Candidate is not powerful enough, then the groups will remain divergent.



(b) A powerful Candidate in one group may be able to attract some Voters from the other group, but that may not be enough to cause the groups to converge.

Figure C.1: Two examples of disparate opinion clusters that do not converge. Voters are colored gray, Candidates are colored purple, winning Candidates are colored orange, and the size represents the Candidates threshold (Eq. (3.2), Eq. (3.1).)

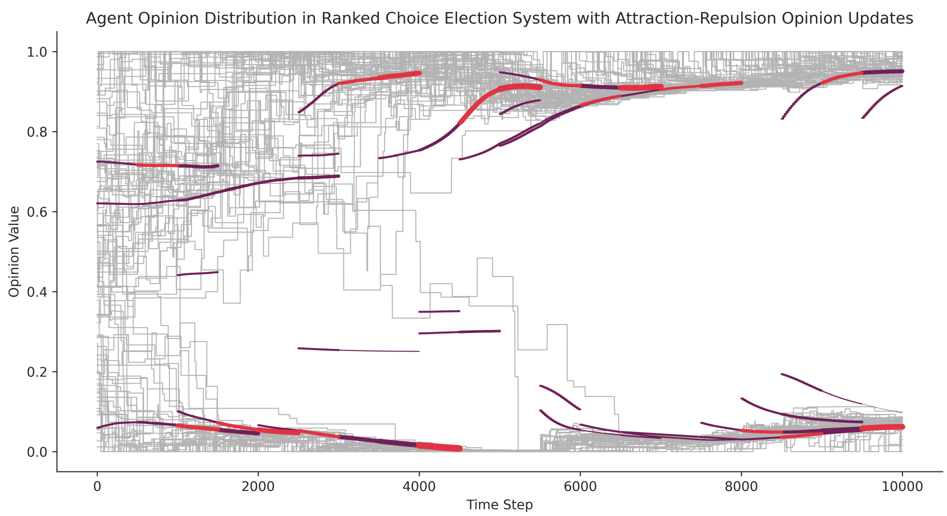


Figure C.2: Opinion dynamics when $\gamma = 42$ and *Radius* $= 0.1$. Voters are colored gray, Candidates are colored purple, winning Candidates are colored orange, and the size represents the Candidates threshold (Eq. (3.2), Eq. (3.2)). All other parameters are listed in Table B.2.

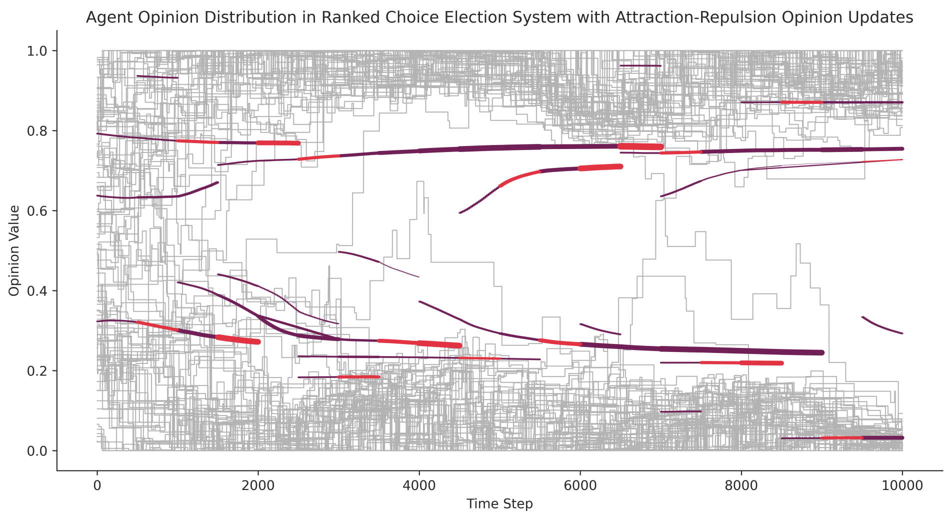


Figure C.3: Opinion dynamics when $\gamma = 42$ and *Radius* $= 0.35$. Voters are colored gray, Candidates are colored purple, winning Candidates are colored orange, and the size represents the Candidates threshold (Eq. (3.2), Eq. (3.2)). All other parameters are listed in Table B.2.

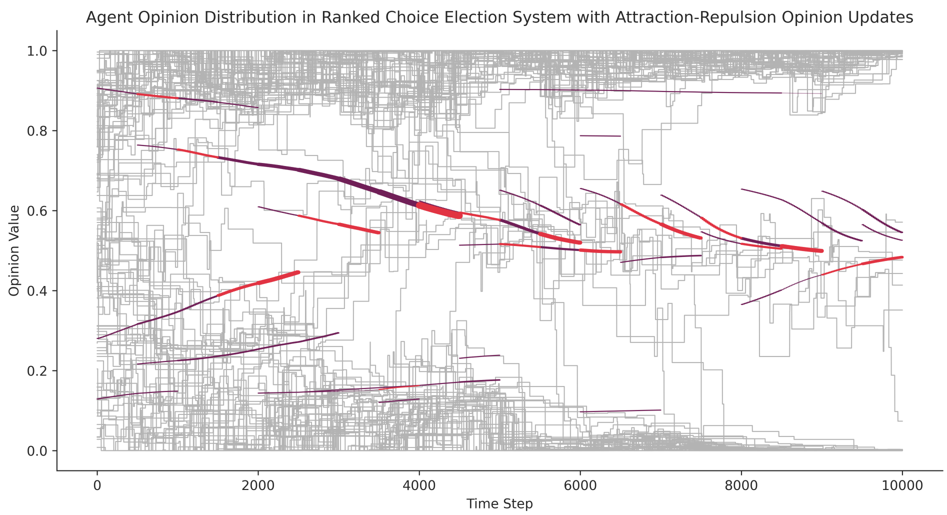


Figure C.4: Opinion dynamics when $\gamma = 42$ and *Radius* $= 0.6$. Voters are colored gray, Candidates are colored purple, winning Candidates are colored orange, and the size represents the Candidates threshold (Eq. (3.2), Eq. (3.2)). All other parameters are listed in Table B.2.

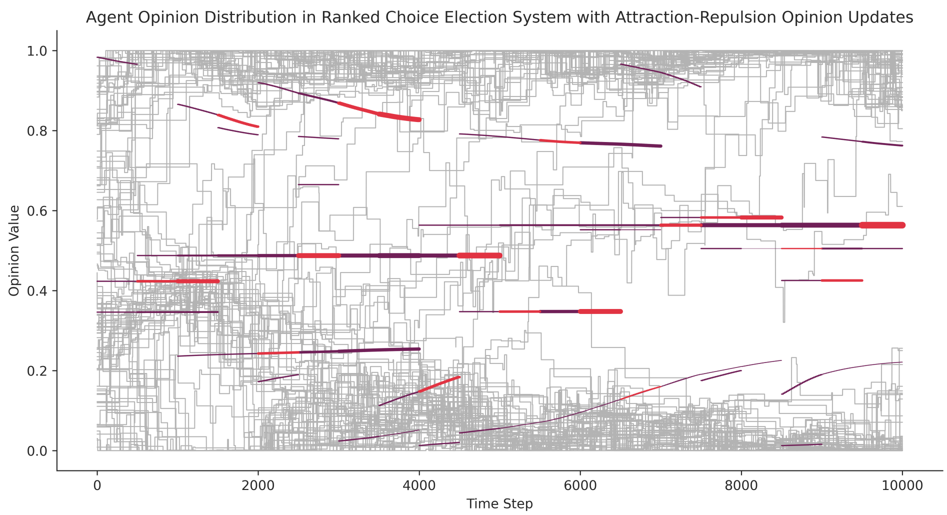


Figure C.5: Opinion dynamics when $\gamma = 42$ and $Radius = 0.85$. Voters are colored gray, Candidates are colored purple, winning Candidates are colored orange, and the size represents the Candidates threshold (Eq. (3.2), Eq. (3.2)). All other parameters are listed in Table B.2.

BIBLIOGRAPHY

[1] Alan I. Abramowitz and Kyle L. Saunders. "Is Polarization a Myth?" In: *The Journal of Politics* 70.2 (Apr. 2008), pp. 542–555. ISSN: 0022-3816, 1468-2508. DOI: 10.1017/S0022381608080493. URL: https://www.journals.uchicago.edu/doi/10.1017/S0022381608080493 (visited on 03/10/2024).

[2] Chloe Ahn and Diana C Mutz. "The Effects of Polarized Evaluations on Political Participation: Does Hating the Other Side Motivate Voters?" In: *Public Opinion Quarterly* 87.2 (July 13, 2023), pp. 243–266. ISSN: 0033-362X, 1537-5331. DOI: 10.1093/poq/nfad012. URL: https://academic.oup.com/poq/article/87/2/243/7161083 (visited on 04/22/2024).

[3] Robert Axelrod, Joshua J. Daymude, and Stephanie Forrest. "Preventing extreme polarization of political attitudes". In: *Proceedings of the National Academy of Sciences* 118.50 (Dec. 14, 2021), e2102139118. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2102139118. URL: https://pnas.org/doi/full/10.1073/pnas.2102139118 (visited on 01/19/2024).

[4] Emily Clough. "Strategic Voting Under Conditions of Uncertainty: A Re-Evaluation of Duverger's Law". In: *British Journal of Political Science* 37.2 (Apr. 2007), pp. 313–332. ISSN: 1469-2112, 0007-1234. DOI: 10.1017/S0007123407000154. URL: https://www.cambridge.org/core/journals/british-journal-of-political-science/article/abs/strategic-voting-under-conditions-of-uncertainty-a-reevaluation-of-duvergers-law/B1CAA3B613F1A133C5EFA9F89BCB3F0B (visited on 03/10/2024).

[5] Emily Clough. "Talking Locally and Voting Globally: Duverger's Law and Homogeneous Discussion Networks". In: *Political Research Quarterly* 60.3 (Sept. 2007), pp. 531–540. ISSN: 1065-9129, 1938-274X. DOI: 10.1177/1065912907305681. URL: http://journals.sagepub.com/doi/10.1177/1065912907305681 (visited on 03/10/2024).

[6] Russell J. Dalton. "Modeling ideological polarization in democratic party systems". In: *Electoral Studies* 72 (Aug. 2021), p. 102346. ISSN: 02613794. DOI: 10.1016/j.electstud.2021.102346. URL: https://linkinghub.elsevier.com/retrieve/pii/S0261379421000652 (visited on 01/21/2024).

[7] Guillaume Deffuant et al. "Mixing beliefs among interacting agents". In: *Advances in Complex Systems (ACS)* 3 (2001), p. 11. DOI: 10.1142/S0219525900000078. URL: https://hal.inrae.fr/hal-02580278 (visited on 01/22/2024).

[8] P. M. DeMarzo, D. Vayanos, and J. Zwiebel. "Persuasion Bias, Social Influence, and Unidimensional Opinions". In: *The Quarterly Journal of Economics* 118.3 (Aug. 1, 2003), pp. 909–968. ISSN: 0033-5533, 1531-4650. DOI: 10.1162/00335530360698469. URL: https://academic.oup.com/qje/article-lookup/doi/10.1162/00335530360698469 (visited on 03/10/2024).

[9] Carlos Andres Devia and Giulia Giordano. "A framework to analyze opinion formation models". In: *Scientific Reports* 12.1 (Aug. 4, 2022), p. 13441. ISSN: 2045-2322. DOI: 10.1038/s41598-022-17348-z. URL: https://www.nature.com/articles/s41598-022-17348-z (visited on 01/29/2024).

[10] Anthony Downs. *An economic theory of democracy*. Nachdr. Boston: Addison Wesley, 1957. 310 pp. ISBN: 9780060417505.

[11] Brian R. Flay. "Catastrophe theory in social psychology: Some applications to attitudes and social behavior". In: *Behavioral Science* 23.4 (1978), pp. 335–350. ISSN: 00057940, 10991743. DOI: 10.1002/bs.3830230404. URL: https://onlinelibrary.wiley.com/doi/10.1002/bs.3830230404 (visited on 03/30/2024).

[12] James H. Fowler and Oleg Smirnov. "Dynamic Parties and Social Turnout: An Agent-Based Model". In: *American Journal of Sociology* 110.4 (Jan. 2005), pp. 1070–1094. ISSN: 0002-9602, 1537-5390. DOI: 10.1086/426554. URL: http://www.journals.uchicago.edu/doi/10.1086/426554 (visited on 03/10/2024).

[13] Noah E. Friedkin et al. "Network science on belief system dynamics under logic constraints". en. In: *Science* 354.6310 (Oct. 2016), pp. 321–326. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aag2624. URL: https://www.science.org/doi/10.1126/science.aag2624 (visited on 03/30/2024).

[14] Rainer; Krause Hegselmann. *Opinion dynamics and bounded confidence models, analysis and simulation*. June 30, 2002. URL: https://www.jasss.org/5/3/2.html (visited on 01/17/2024).

[15] Melvin Hinich. "The mean versus the median in spatial voting games". In: *Game Theory and Political Science New York. NYU Press: New York* (1978).

[16] Melvin J. Hinich. "Equilibrium in spatial voting: The median voter result is an artifact". In: *Journal of Economic Theory* 16.2 (Dec. 1, 1977), pp. 208–

219. ISSN: 0022-0531. DOI: 10.1016/0022-0531(77)90005-9. URL: https://www.sciencedirect.com/science/article/pii/0022053177900059 (visited on 04/24/2024).

[17] Melvin J. Hinich and Peter C. Ordeshook. “Abstentions and Equilibrium in the Electoral Process”. In: *Public Choice* 7 (1969), pp. 81–106. ISSN: 0048-5829. URL: https://www.jstor.org/stable/30022611 (visited on 04/24/2024).

[18] Melvin J. Hinich and Peter C. Ordeshook. “Plurality Maximization vs Vote Maximization: A Spatial Analysis with Variable Participation”. In: *American Political Science Review* 64.3 (Sept. 1970), pp. 772–791. ISSN: 0003-0554, 1537-5943. DOI: 10.2307/1953462. URL: https://www.cambridge.org/core/journals/american-political-science-review/article/abs/plurality-maximization-vs-vote-maximization-a-spatial-analysis-with-variable-participation/E20672E0912D10C22845C062B72213D9 (visited on 04/24/2024).

[19] Yoshihisa Kashima et al. “Ideology, communication and polarization”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 376.1822 (Apr. 12, 2021), p. 20200133. ISSN: 0962-8436, 1471-2970. DOI: 10.1098/rstb.2020.0133. URL: https://royalsocietypublishing.org/doi/10.1098/rstb.2020.0133 (visited on 03/10/2024).

[20] Sung-youn Kim, Charles S. Taber, and Milton Lodge. “A Computational Model of the Citizen as Motivated Reasoner: Modeling the Dynamics of the 2000 Presidential Election”. In: *Political Behavior* 32.1 (Mar. 1, 2010), pp. 1–

28. ISSN: 1573-6687. DOI: 10.1007/s11109-009-9099-8. URL: https://doi.org/10.1007/s11109-009-9099-8 (visited on 03/10/2024).

[21] Rainer Hegselmann {and} Ulrich Krause. *Truth and Cognitive Division of Labour: First Steps Towards a Computer Aided Social Epistemology*. June 30, 2006. URL: https://www.jasss.org/9/3/10.html (visited on 03/30/2024).

[22] Emily Kubin and Christian Von Sikorski. "The role of (social) media in political polarization: a systematic review". In: *Annals of the International Communication Association* 45.3 (July 3, 2021), pp. 188–206. ISSN: 2380-8985, 2380-8977. DOI: 10.1080/23808985.2021.1976070. URL: https://www.tandfonline.com/doi/full/10.1080/23808985.2021.1976070 (visited on 04/22/2024).

[23] Geoffrey C. Layman, Thomas M. Carsey, and Juliana Menasce Horowitz. "PARTY POLARIZATION IN AMERICAN POLITICS: Characteristics, Causes, and Consequences". In: *Annual Review of Political Science* 9.1 (June 1, 2006), pp. 83–110. ISSN: 1094-2939, 1545-1577. DOI: 10.1146/annurev.polisci.9.070204.105138. URL: https://www.annualreviews.org/doi/10.1146/annurev.polisci.9.070204.105138 (visited on 01/21/2024).

[24] Konstantinos Matakos, Orestis Troumpounis, and Dimitrios Xefteris. "Electoral Rule Disproportionality and Platform Polarization". In: *American Journal of Political Science* 60.4 (2016), pp. 1026–1043. ISSN: 0092-5853. URL: https://www.jstor.org/stable/24877470 (visited on 01/22/2024).

[25] Konstantinos Matakos, Orestis Troumpounis, and Dimitrios Xefteris. "Turnout and Polarization Under Alternative Electoral Systems". In: *The Political Econ-*

omy of Governance: Institutions, Political Performance and Elections. Ed. by Norman Schofield and Gonzalo Caballero. Studies in Political Economy. Cham: Springer International Publishing, 2015, pp. 335–362. ISBN: 9783319155517. DOI: 10.1007/978-3-319-15551-7_18. URL: https://doi.org/10.1007/978-3-319-15551-7_18 (visited on 01/28/2024).

[26] Jennifer McCoy et al. *Reducing Pernicious Polarization: A Comparative Historical Analysis of Depolarization*. Carnegie Endowment for International Peace. May 2022. URL: https://carnegieendowment.org/2022/05/05/reducing-pernicious-polarization-comparative-historical-analysis-of-depolarization-pub-87034 (visited on 01/29/2024).

[27] Shane T. Mueller and Yin-Yin Sarah Tan. "Cognitive perspectives on opinion dynamics: the role of knowledge in consensus formation, opinion divergence, and group polarization". In: *Journal of Computational Social Science* 1.1 (Jan. 1, 2018), pp. 15–48. ISSN: 2432-2725. DOI: 10.1007/s42001-017-0004-7. URL: https://doi.org/10.1007/s42001-017-0004-7 (visited on 03/10/2024).

[28] Christopher Musco et al. *How to Quantify Polarization in Models of Opinion Dynamics*. Oct. 25, 2021. DOI: 10.48550/arXiv.2110.11981. arXiv: 2110.11981[cs]. URL: http://arxiv.org/abs/2110.11981 (visited on 03/10/2024).

[29] Sergey E. Parsegov et al. "Novel Multidimensional Models of Opinion Dynamics in Social Networks". In: *IEEE Transactions on Automatic Control* 62.5 (May 2017). arXiv:1505.04920 [cs, math], pp. 2270–2285. ISSN: 0018-9286,

1558-2523. DOI: 10.1109/TAC.2016.2613905. URL: http://arxiv.org/abs/1505.04920 (visited on 03/30/2024).

[30] Adam Paszke et al. "Automatic differentiation in PyTorch". In: (Oct. 28, 2017). URL: https://openreview.net/forum?id=BJJsrmfCZ (visited on 05/15/2024).

[31] *Political Polarization in the American Public*. Pew Research Center, June 12, 2014. URL: https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/ (visited on 01/29/2024).

[32] Lin Qiu and Riyang Phang. "Agent-Based Modeling in Political Decision Making". In: *Oxford Research Encyclopedia of Politics*. Oxford University Press, Feb. 28, 2020. ISBN: 978-0-19-022863-7. DOI: 10.1093/acrefore/9780190228637.013.913. URL: https://oxfordre.com/politics/view/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-913 (visited on 01/17/2024).

[33] Steven F. Railsback and Volker Grimm. *Agent-based and individual-based modeling: a practical introduction*. Princeton: Princeton University Press, 2012. 329 pp. ISBN: 9780691136738 9780691136745.

[34] Charles K. Rowley. "The Relevance of the Median Voter Theorem". In: *Zeitschrift für die gesamte Staatswissenschaft / Journal of Institutional and Theoretical Economics* 140.1 (1984), pp. 104–126. ISSN: 0044-2550. URL: https://www.jstor.org/stable/40750679 (visited on 04/24/2024).

[35] Simon Schweighofer, David Garcia, and Frank Schweitzer. "An agent-based model of multi-dimensional opinion dynamics and opinion alignment". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 30.9 (Sept. 1, 2020), p. 093139. ISSN: 1054-1500, 1089-7682. DOI: 10.1063/5.0007523. URL: https://pubs.aip.org/cha/article/30/9/093139/1059669/An-agent-based-model-of-multi-dimensional-opinion (visited on 03/10/2024).

[36] Pawel Sobkowicz. "Quantitative Agent Based Model of Opinion Dynamics: Polish Elections of 2015". In: *PLOS ONE* 11.5 (May 12, 2016), e0155098. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0155098. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0155098 (visited on 03/10/2024).

[37] Abraham Tesser. "When Individual Dispositions and Social Pressure Conflict: A Catastrophe". In: *Human Relations* 33.6 (June 1980), pp. 393–407. ISSN: 0018-7267, 1741-282X. DOI: 10.1177/001872678003300604. URL: http://journals.sagepub.com/doi/10.1177/001872678003300604 (visited on 03/30/2024).

[38] Gerard Weisbuch et al. *Interacting Agents and Continuous Opinions Dynamics*. Nov. 26, 2001. DOI: 10.48550/arXiv.cond-mat/0111494. arXiv: cond-mat/0111494. URL: http://arxiv.org/abs/cond-mat/0111494 (visited on 03/30/2024).