EVOLUTIONARY DYNAMICS OF COLLECTIVE ACTION PROBLEMS

A Thesis

Submitted to the Faculty

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

in

Mathematics

by

Matthew Isaiah Jones

DARTMOUTH COLLEGE

Hanover, New Hampshire

May 2022

Examining Committee:

Feng Fu, Chair

Scott Pauls

Dan Rockmore

Nicholas Christakis

F. Jon Kull, Ph.D.
Dean of the Guarini School of Graduate and Advanced Studies

Abstract

In the study of a single rational individual, it is often straightforward to design a framework of rules and rewards to encourage a particular outcome that maximizes the individual's welfare. When it comes to groups of individuals, on the other hand, this is not guaranteed. Incentives for individuals do not always align with an optimal outcome for society as a whole, and it is misguided to treat a group as a single entity that thinks and behaves like its constituent members. This thesis uses mathematical tools to study aspects of collective action problems in three contexts: the distributed graph coloring problem, polarization and voting, and fake news in social networks.

Many scenarios require group members to specialize or differentiate themselves from those around them to maximize group effectiveness. In such situations, reaching a state of maximum global effectiveness may require individuals to make short-term sacrifices for the greater good when the group becomes "gridlocked." We use the mathematical concept of a graph coloring problem as a proxy for such coordination problems, which allows us to draw new conclusions and parallels with other problems that require consensus instead of specialization.

One of the classic group decision-making problems is the question of leadership through voting. The outcome of an election is determined by the properties of the underlying electorate, and we use a spatial model of voting to examine polarization and the relationship between rising voter extremism and extremism in the political elite class.

One of the commonly-cited drivers of rising polarization is the recent explosion in misinformation driven by online social media. We examine how fake news spreads through a social network, test the effectiveness of relying on citizen fact-checkers, and measure the effect social network structure has on our fact-checking efforts.

This work demonstrates the utility of using agent-based models when studying social dilemmas and we hope that these techniques will continue to be applied to critical problems of the modern era including public health, climate change, and democracy.

Preface

Almost half of my time at Dartmouth has been marred by a global pandemic, but these have been wonderful times despite that.

I am grateful to everyone that I have met in my time here, especially my advisors, Feng Fu and Scott Pauls. Feng has been an inspiration and a guiding light without whom this thesis would not exist, and Scott's wisdom has informed my decisions in my research and in my teaching. Special thanks also go to the other members of my committee, Dan Rockmore and Nicholas Christakis, for their curiosity and interest in my work.

This small department has been nothing except kind and encouraging to me, and I will cherish everyone in it as friends for the rest of my life. Special recognition goes to the other members of my cohort, Sam Tripp, Zach Winkeler, and Yao Xiao, for helping me survive all five years of graduate school.

Last but not least, I would like to thank my parents for their (uncountably) infinite love and support.

Contents

List of Figures

Chapter 1

Introduction

Section 1.1

Collective Action Problems

It began with a cow. At a livestock show in England in the early 1900s, 800 or so people, many with little or no experience around cows, placed bets on the weight of an ox. After examining all the votes, Sir Francis Galton made the remarkable discovery that the median guess was off by only 0.8% [123]! Galton recognized the importance of examining the ability of large groups to organize, coordinate, and make rational decisions, writing "In these democratic days, any investigation into the trustworthiness and peculiarities of popular judgments is of interest."

Since then, a tremendous amount of effort has gone into studying such "collective intelligence," and numerous examples of a group being wiser than the sum of its part have been found in all areas of life (see [178] for a survey). In "The Wisdom of Crowds," James Surowiecki identified three types of problems faced by groups: cognition, coordination, and cooperation [284]. In cognition problems, the group tries to answer a question like "How much does that cow weigh?" or "Who would make the best president?" or "Which of these news stories is real and which is fake?" This

requires compiling the thoughts and intuitions of individual group members into a coherent whole. Coordination problems require groups to determine how to work together to solve a problem like selecting individual labor specializations to increase the productivity of the economy or neighboring radio stations selecting different broadcast frequencies to minimize the amount of radio interference. Finally, cooperation problems require individuals in a group to put aside their self-interest and contribute to some common good. Cooperation problems are not specifically addressed in this work, but methods for fostering cooperation in a population are highly sought after and well-studied [73, 94, 118, 119, 175, 176, 213, 249]. Surowiecki identified collective intelligence in all walks of life, from game shows and sports betting to internet search engines and intelligence agencies.

However, group decision-making can have serious drawbacks. It has been well documented that behavior that is optimal for the individual is not always optimal for the group [142, 219]. This makes coordination and cooperation problems difficult because actions that are incentivized on the individual level are often harmful to the group's success. Pollution and over-population are two examples that Hardin identified as self-interested behavior that is detrimental to the group as a whole in one of the landmark papers introducing this "Tragedy of the Commons" [141]. Cognition problems also have unique complications in the group setting. There are statistical and psychological effects that can derail the wisdom of the crowds as individuals abandon their correct personal beliefs in favor of the incorrect beliefs of those around them [26, 177].

Making decisions as a group does have advantages, though. One critical idea in collective action is the notion of decentralization, where any central decision-makers are removed and the group is left to organize itself effectively with no outside direction. Decentralization has been studied in fields ranging from economics [302]

and cryptocurrency [201] to control engineering [262] and the organization of intelligence agencies [138], and while decentralization offers many benefits depending on context, it can also make coordination difficult. This happens even when solving problems that are trivial for a central decision-maker with complete information and authority [44, 299]. Despite these difficulties, in certain areas, groups seem remarkably well-adapted to decentralized coordination. Famously, Adam Smith's "Invisible Hand" guided workers to specialize and coordinate with others in a way that benefited the group as well as the individual without any sort of outside perspective or top-down instruction [275]. Since then, much work has gone into the ability of markets to self-organize into effective specialized units that all complement each other [33, 104].

Malone et al. [183] identified four key questions to keep in mind when examining collective intelligence:

- Who is performing the task?
- Why are they doing it?
- What is being accomplished?
- How is it being done?

The focus of this thesis is the "why" and the "how." Why do individuals behave the way they do? What are the incentives driving them to act, and how will changing those incentives affect their behavior? And how do the many small choices of individuals manifest themselves as behavior of the entire group? What are the systems that turn actions into trends, and how can these systems be manipulated to impact the decision-making abilities of the entire group? The "who" and "what" are much more context dependent and not the focus of this work, but we will see models of voters electing officials in an increasingly polarized environment and news consumers deciding between real and fake news.

This work is fundamentally applied, meaning the math is almost entirely motivated by and focused on real-world scenarios and questions. However, every attempt to study the world using math requires a conversion from reality to abstract mathematical structures. Such models appear in many disciplines, including physics [20], biology [136], psychology [1], economics [49], and finance [182]. In some cases, the mathematical models are so ingrained in the field that experts in the field make no distinction between model and reality; the model is uncontested as a complete and accurate representation of the situation. However, this is often not the case (particularly in the social sciences), and models are understood to be approximations designed to take into account one or two important factors while ignoring the rest in hopes of deriving a novel result that can be one piece among many in our understanding of the system in question.

This thesis develops several models of collective action and then uses diverse mathematical tools to drawn meaningful conclusions about the systems in question. I begin by introducing the necessary mathematical paradigms: game theory, evolutionary dynamics, network theory, and graph colorings.

Section 1.2

Basic Mathematical Frameworks and Concepts

1.2.1. Game Theory

Modern game theory is usually said to have begun with John von Neumann. In 1928, he published a paper that proved the Minimax Theorem [304] for two-player zero-sum games with perfect information, and 16 years later, he wrote a book with Oskar Morganstern [206] which greatly expanded their previous work and laid the foundation for the field.

Although originally applied to economics, game theory has found applications in

all fields of science. Any scenario where agents interact can be modeled as a game. We typically evaluate these games analytically, but there is also empirical work which has exposed interesting behaviors and irrationalities in the way humans behave [222].

The foundation of game theory is the *payoff matrix*, which contains information about the benefits and costs of choosing specific strategies, depending on the actions of one's opponent. In a two-player game where each player has two strategies, the payoffs can be placed in the matrix as follows:

		Player 2	
		C	D
Player 1	A	(a, b)	(c, d)
	B	(e, f)	(g, h)

Table 1.1: A payoff matrix for a two-player asymmetric game. The ordered pairs represent the payoffs for Player 1 and Player 2, respectively. For example, if Player 1 chooses B and Player 2 chooses C, Player 1 gets a payoff of e and Player 2 gets payoff f.

Of course, this can be generalized to have any number of strategies (by adding rows and/or columns) and any number of players (by using an n-dimensional lattice), but in this thesis, we only need to consider two-player games.

There are many classes of games. The most famous by far is the Prisoner's Dilemma [18, 94], a symmetric game with two choices for each player, to cooperate (C) or defect (D).

		Player 2	
		C	D
Player 1	C	(R, R)	(S, T)
	D	(T, S)	(P, P)

Table 1.2: The payoff matrix for the Prisoner's Dilemma. The four payoffs follow the relationship $T > R > P > S$ and $2R > S + T$.

The Prisoner's Dilemma has a dominant strategy, D. Regardless of Player 2's

strategy, it is better for Player 1 to play D (see Table 1.2). It is not difficult to see that this results in both players choosing D, even though both players choosing C would increase all payoffs equally and give the maximum payoff for the group as a whole. Many classical games like the Prisoner's Dilemma focus on two players attempting to get the better of each other; in the Prisoner's Dilemma, both players would like to defect while their opponent cooperates, thus reaping rewards and avoiding punishments. A great body of work is focused on discovering conditions where cooperation can survive or even out-compete defection [94, 213], including games on networks [62, 216] and iterated games [18].

In contrast to the Prisoner's Dilemma, coordination games are a well-studied class of games with no dominant strategies in which all players receive the most benefit when they work together [273]. The optimal behavior for all players can usually be determined and agreed upon if all players can meet, exchange information, and strategize beforehand, but we typically require players to choose strategy simultaneously with no prior communication. In such games, the difficulty comes not from attempting to take advantage of one's opponent, but predicting what strategy one's partner will play before choosing one's own strategy [212, 299]. Coordination games can result in non-optimal outcomes for the group similar to the Prisoner's Dilemma, but instead of a failure to cooperate, now there is a failure to coordinate. However, there can still be a "defecting" component, in which one's opponent can unilaterally choose a strategy with lower maximum payoff but also less risk [107].

For the purposes of this thesis, collective action games fall into two broad categories: games where individuals coordinate to pick the same strategies (referred to as coordination games) [34, 156, 187], and games where individuals coordinate to pick different strategies (referred to as anti-coordination games) [46, 52, 53, 185, 205]. Symmetric coordination games can usually be resolved if the players are allowed to

communicate, but asymmetries in anti-coordination games can make cooperation difficult and highly dependent on network structure [46].

In general, there are substantial qualitative differences between coordination and anti-coordination games. In a coordination game, it is relatively simple to assign every individual the same strategy, reaching a maximum payoff for the entire group. However, if three individuals play an anti-coordination game, then two of them must choose the same strategy and thus not get the maximum payoff possible. In Chapter 3, we will see that these problems can be dual problems under very specific conditions, but in general, these two games give very different dynamics.

Of course, a rational individual in any of these scenarios may decide that the best course of action is to play a mixture of strategies with different probabilities, either to be less predictable to her opponent or to take advantage of different choices her opponent may make. These are called *mixed strategies*, as opposed to *pure strategies* where an individual chooses a single strategy to play with probability 1.

A classical way to evaluate these games is with the *Nash equilibrium*. A Nash equilibrium is a profile of strategies, one for each player, so that no single individual can improve his expected payoff by changing to a different strategy. A wealth of research has gone into studying Nash equlibria since John Nash first proved that at least one equilibrium must always exist [203, 204]. Perhaps the most famous result is Wilson's Oddness Theorem, which states that in almost all finite games, there are an odd number of Nash equilibria [311]. The Prisoner's Dilemma (Table 1.2) has a single Nash equilibrium, (D, D). In many coordination games, there are three Nash equilibria: two with pure strategies and one with mixed strategies.

1.2.2. Evolutionary Dynamics

Often, we are interested in observing how a population evolves. In the simple case of an infinite, well-mixed population, the population can be effectively modeled with or-

dinary differential equations that give a closed-form solution to the change in strategy distribution over time. We assign each of the i strategies an integer value between 1 and i and utilize a vector $\mathbf{x}$ where x_i is the fraction of the population playing strategy i. The population's change is governed by the replicator equation

$$\dot{x_i} = x_i \left(f_i(\mathbf{x}) - \phi(\mathbf{x}) \right) \tag{1.1}$$

where $f_i(\mathbf{x})$ is the fitness of every individual playing strategy i, and $\phi(\mathbf{x}) = \sum_i x_i f_i(\mathbf{x})$ is the average fitness of the population [153, 212, 293]. The replicator equation is inspired by the Lotka-Volterra equations which modeled predator-prey systems [179, 291, 301].

Keeping with this population dynamics approach, *evolutionarily stable strategies* [276] are closely related to Nash equilibria and reflect similar ideas, but from the perspective of competition between strategies instead of between individuals. An evolutionarily stable strategy is a strategy that is secure against invasion by mutants playing a different strategy.

The replicator equation is an excellent tool for studying infinite populations, but other methods are needed to handle the subtleties of finite systems. A class of models called birth-death processes are the standard tool for modeling population change in this thesis [215,287,294]. These processes take many forms, but in general, individuals have some fitness (often a function of the payoff from playing games) which determines the likelihood of being chosen to reproduce after a random individual is removed from the population. Just like the replicator equation in infinite populations, this causes strategies with high fitness to grow and those with low fitness to head toward extinction. The specific modeling choices made in an update rule can have a profound impact on the system dynamics, and analytic results can be either straightforward or impossible to obtain [217, 317].

Evolutionary game theory, the use of game theory to determine how systems change over time, requires us to make many assumptions when modeling real-world populations. Our models frequently have features that do not match what is really happening, but can be explained to be good approximations of reality by shifting perspective. For example, individuals may change strategy if they see others playing strategies that have more success in the current state of the system. Alternatively, we may trace how a reproducing population changes over the course of many generations by simulating deaths and births over a long period of time. Birth-death models cover both of these processes.

Many of these models make the tenuous assumption that population size is fixed. However, by shifting perspective, we can view our fixed population as showing the frequency of each type of behavior instead of showing the true number of individuals in the population. We also typically assume that the population is made up of individuals that all have the same payoffs, fitness functions, etc. This makes the model tractable, and the hope is that any variation among the population will be insignificant when using average values for all individuals.

Our goal when developing models is to find the middle ground that is complex enough to learn something new and interesting while also being simple enough to understand and analyze. We will see an example of a birth-death process that meets both criteria in Chapter 5.

1.2.3. Networks

The assumption that a population is well-mixed, i.e. that all pairs of individuals are equally likely to interact, allows for very clean analytic work, but is clearly false in a wide range of scenarios where networks create a more restrictive interaction architecture. Social networks (including but not limited to networks arising from online social media) have a profound on impact on how we function as a society

[32, 46, 64, 74, 77, 78, 85, 97, 131, 132, 167, 194, 248, 265, 287, 289, 313].

From a mathematical perspective, a network (also referred to as a graph) is a set V of vertices and a set E of edges, which are pairs (sometimes ordered and sometimes accompanied by weights) of vertices. (For an introduction to directed and weighted networks and their implications, see [209]). In a social network, the vertices represent individuals and the edges represent some connection between people. Friendship, acquaintanceship, familial relations, and physical proximity are all common ways of connecting two individuals in a network, and two vertices that are connected by an edge are said to be adjacent and a vertex's *degree* is the number of connected edges.

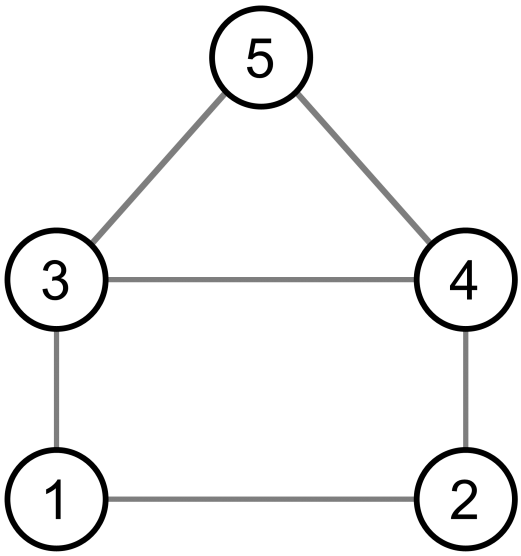

Figure 1.1: A small network to demonstrate different methods of network representation, with five vertices and six edges.

One way of describing a network mathematically is just to list all the elements in V and E. For example, in the network in Figure 1.1, we have the following sets:

$$V = \{1, 2, 3, 4, 5\}$$

$$E = \{(1,2),(1,3),(2,4),(3,4),(3,5),(4,5)\}$$

This method of listing all the edges as pairs is known as an edge list, and while it has the advantage of being directly tied to the definition of a network, there are other network representations that are more appropriate for particular applications.

A more pragmatic way of representing this network is to consider each individual separately and list their neighbors. This is called an adjacency list [270], and the adjacency list for the same small network is below.

1	2	3	
2	1	4	
3	1	4	5
4	2	3	5
5	3	4	

Because it lists the neighbors of each individual, the adjacency list is very useful for simulations and agent-based models. However, there is another method of network representation that is extremely useful for the mathematical analysis of networks, called the adjacency matrix [209]. In an unweighted network, we say that the i,jth entry of A is given by

$$A_{ij} = \begin{cases} 1 & \text{there is an edge from } j \text{ to } i \\ 0 & \text{otherwise} \end{cases} \tag{1.2}$$

In our small example, the adjacency matrix is

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \tag{1.3}$$

Notice that because our particular network is undirected, the matrix is symmetric. Representing the network as a matrix allows us to use all the tools of linear algebra to study the network. In particular, the spectrum of the adjacency matrix (and other related matrices) contains an abundance of information about the properties of the network [41, 75, 109, 199, 244].

Broadly, the goal of network science is to show that the structure of a network has some impact on the behavior of the system. To make such claims, it can be useful to be able to create large *families* of networks that have a given property. In Chapter 5, we describe several common network models, how they are created, and what useful properties they have.

After creating networks with desirable properties that "look real" in some sense, we still have the question of how to use them to draw conclusions about the system we are interested in. A natural question to ask about a given network is which vertices are most important. This obviously depends on the system we are modeling as well as the notion of importance, and the various methods of determining importance are known as *centrality measures*. Each centrality measure gives a different measure of importance and may be relevant in a different scenario. The simplest centrality measure is just degree, which determines a vertex's importance by the number of neighbors the vertex has. Eigenvector centrality [41] and Google Pagerank [50, 127] are both defined recursively, meaning a vertex has a high centrality if its neighbors

also have high centrality. Other centralities are defined more combinatorially. The closeness centrality [72] measures how close a vertex is to all the other vertices in the network and betweenness centrality [30] measures how important a vertex is in passing messages around the network.

In addition to determining static properties of the network, we may also wish to model how network structure affects dynamic systems. This may be as simple as determining how much material can flow across a transportation network like travellers moving from one city to another by rail [114].

However, the most versatile models that can be implemented on networks are game theoretic. and use the ideas presented in 1.2.2. In these *spatial games*, each vertex is an individual in a population, and the edges represent connections between the individuals whose interactions are modeled by games [8,119,130,216,233,248,258, 265,266]. In the broadest possible sense, each individual chooses some strategy, and pairs of individuals connected by edges play the game. The resulting payoff is used to change the strategy of one or more individuals who copy the strategy of successful neighbors, and this process is repeated to see how the population's strategy choices change over time. These models can be studied analytically [168,215,216,292], with computer simulations [47,53], or with human subjects [52,250].

1.2.4. Graph Colorings

Networks were first studied by mathematicians in the context of graph theory, which began, as many interesting mathematical things do, with Leonhard Euler. In 1735, he wrote a paper about the "geometry of position" of seven bridges and two islands in a river that would inspire generations of mathematicians to come [7,105]. Since then, graph theory has proved to be a very fruitful branch of mathematics, with many interesting results on topics including planar graphs [171], cut size to separate vertices [191], matching problems [180], and graph coloring.

A *proper graph coloring* is a labeling of each vertex with a color so that no two adjacent vertices have the same color. For example, in a *bipartite network* [54, 103, 135, 226, 278], the vertices can be divided into two groups, where every edge connects vertices of different groups. If each of these groups is assigned a different color, then neighbors will always have different colors.

The study of graph colorings began over a century ago, inspired by maps where adjacent regions were always colored with different colors. The most famous result about graph colorings is easily the Four Color Theorem, which states that every map can be colored with four colors so that no region is the same color as any of its neighbors. After years of traditional proofs being proposed and subsequently debunked, the Four Color Theorem became the first major theorem to be successfully proven with computer assistance [14]. However, there are many other interesting results about graph colorings, including results about computing the chromatic number, which is the smallest number of colors needed to color a graph [55], and the chromatic polynomial, which determines how many ways a graph can be colored [36].

Strictly speaking, the graphs studied by pure mathematicians and the networks encountered in the real world are the same objects, collections of vertices and edges, and the terms are used interchangeably in this thesis. In practice, however, real-life networks tend to be much larger and messier than the clean graphs studied by mathematicians. Graph theory and network science also differ in the types of questions asked and how to answer them. Graph theorists tend to prove theorems in complete generality with the full rigor of mathematics, while network science tends to be more scenario-dependent, attempting to answer a specific question about a specific situation playing out on a specific type of network. Of course, this is not to say that the two are easily separated; many results from graph theory can be extremely useful in network science and vice versa. Chapters 2 and 3 use graph colorings as an analog for

studying group coordination problems on networks, a question that would normally fall under the classification of network science.

In contrast to proper graph colorings, we define a *uniform coloring* as a labelling of the vertices so that every individual selects the same color. Mathematically, such colorings are trivial and uninteresting. However, they can be useful when studying coordination games, as we will see below.

Section 1.3

Societally Important Collective Action Problems

In Chapters 4 and 5, we examine two particular social coordination problems that are particularly relevant today: polarization/voting and fake news.

1.3.1. Polarization and Voting

In [261], Sen makes a strong argument that political needs, specifically democratic representation, must be satisfied before economic needs like poverty, unhappiness, and inequality can be satisfied. If we assume this premise holds, the basic building blocks of democracy, voting and elections, are of the utmost importance. There are a range of methods for holding elections, each with their pros and cons. We do not offer a comparative analysis of election systems here, but for an overview of the voting systems used around the world, see [211].

As a note, in most of this work, we consider political candidates to be rational, vote-seeking individuals who are able to change position to maximize votes. We assume candidates will move through the ideological space according to adaptive dynamics to maximize their vote share [152, 312]. However, we could also consider candidates who are ideologically motivated and will not change position to gain more votes [169].

The work in Chapter 4 is in the spirit of this quote from Davis, Hinich and Ordeshook, "The fundamental process of politics is the aggregation of citizens' preferences into a collective - a social - choice" [87]. We consider voting as a collective action problem where the entire population must decide on a single leader that will best serve the needs of the electorate. This can also be considered a collective action problem from the perspective of a political party. Party officials may believe that one candidate can best represent them in a general election, but another candidate may decide to challenge the party favorite in a primary election for selfish reasons.

This is tangentially related to work studying a general version of elections called *social welfare functions*, which attempt to consider individual value judgements and measure the total amount of value to society for each possible choice [35]. However, Arrow's Impossibility Theorem has shown us that there is no possible social welfare function that satisfies a few sensible axioms [17]. As a result, this work does not attempt to reinvent or even critique the most prevalent of voting systems, called plurality or first-past-the-post voting, as all social welfare functions have deficiencies. Instead, we will study how the properties of a population impact the types of politicians that win public office when using plurality voting as our method of group decision-making. Of course, this does not mean that the exact system of voting is irrelevant [252, 259]. We will also ignore the effects of spatial sorting [184] and gerrymandering [279].

Spatial Models of Voting. Our main tool for studying voting and polarization is the *spatial model of voting*. Spatial models were first introduced by Hotelling in 1929 [155], applied to political science by Downs in 1957 [95], and solidified by Davis, Hinich, and Ordeshook in the late 60s [87]. The essence of spatial models is to represent political beliefs as points in $\mathbb{R}^n$. This embedding into a metric space allows us to convert what was previously a vague and poorly-defined sense of political

orientation into a clean mathematical space.

The impact of these spatial models has been immense. The NOMINATE and DW-NOMINATE voting scores by Poole and Rosenthal are two related methods to determine the position of a member of Congress in n-dimensional political space using nothing but their roll call votes [237, 239–241]. Remarkably, the NOMINATE and DW-NOMINATE algorithms needs no information about the content of the bills and instead can determine a liberal/conservative score simply by comparing the votes of all members of Congress. These "NOMINATE scores" have been instrumental in studying the history of Congress as well as evaluating its current state. In fact, according to Robert Franzese, this work has "revolutionized the manner in which political scientists measure and think about ideology. One can say perfectly correctly, and without any hpyerbole: the modern study of the U.S. Congress would be simply unthinkable without NOMINATE legislative-roll-call-voting scores" [116].

A key consideration when building a spatial model is how many dimensions to use. Obviously, lower-dimensional models are easier to work with, but at the cost of fidelity. On the other hand, higher-dimensional models may be taking into account more factors than the low-dimensional models, but they are much more unwieldy from an analytic perspective [86, 149]. Interestingly, in many spatial models, much of modern political action is effectively one-dimensional [239]. The one-dimensional political axis ranging from liberal to conservative is a simple and effective tool for "quick and dirty" analysis but there is evidence that additional dimensions are needed for a more complete understanding of how individuals vote [144]. In this work, we stick with the one-dimensional model that accounts for an ever-growing proportion of voting behavior [239] and lends itself nicely to mathematical analysis.

A key result from spatial models of voting is the Median Voter Theorem [79, 81, 146]. There are many different statements of the theorem, but they all are variations

on the theme of "the median voter chooses the winner, so candidates will compete for the median voter." Hotelling observed something like this [155], although not in the context of politics, but Black was the first one to apply it to voting in the context of single-peaked preference profiles [38]. The Median Voter Theorem is often invoked as a moderating force in politics, forcing politicians to maintain even-keeled positions despite most voters holding more extreme views [43]. In Chapter 4, we consider variations of the spatial model in which the Median Voter Theorem does not hold.

1.3.2. Fake News

Putting all this effort into the study of elections and voting only makes sense if people know who or what they are voting for. Thomas Jefferson said that one of the foundational requirements for a lasting democratic government was "general education to enable every man to judge for himself what will secure or endanger his freedom. [157]" Agreeing with Jefferson, there is a general consensus among political scientists that a well-informed voting population is a prerequisite to a well-functioning government. However, it has always been a struggle to keep voters at an appropriate level of education; a lack of factual knowledge in the American electorate has been well documented (see Ref [151] for some examples). While Jefferson was concerned primarily with literacy and the formation of a public education system, there is another complicating factor that is threatening voters' ability to make well-informed decisions. A voter cannot reasonably be expected to determine what will "secure or endager his freedom" when being exposed to a deluge of misinformation, particularly on social media.

Malicious and deliberately false posts, commonly called "fake news", burst into public consciousness in the mid 2010s, around the time of the 2016 U.S. presidential election and the U.K. Brexit vote. In a social media environment that encourages

sharing and reposting flashy headlines instead of researching and developing a nuanced understanding of a topic, social media platforms seem to facilitate the spread of fake news [89, 172, 263, 271, 305]. Social influence, following, and unfollowing can create polarized and segregated structure in social media like Twitter [307]. These insular communities, colloquially known as "echo chambers" [196, 298], can serve as hotbeds for fake news to fester and create conditions for confirmation bias and selection bias [239] and thus can facilitate the spread of misinformation [89]. During the COVID-19 pandemic, misinformation has severely impacted our efforts to control the pandemic ("misinfodemics") [59, 192, 229].

Related to the polarization studied in Chapter 4, in the context of modern politics, partisanship has come to dominate the political sphere and stall political consensus, both amongst the political elite and the general population [11, 65–68]. This, in turn, has led to a rise in politically motivated falsehoods being spread online. It has become a major research concern to effectively understand circumstances that will lead to consensus of opinion and others that will lead to divergence of opinion and a weakening of information transfer [13, 117, 121, 154, 202, 306, 314].

While it is hard to measure the exact factors that contribute to the spread of fake news, it may be that something about the structure of social media (e.g. the reposting/retweeting network) is allowing their spread. To attempt to quantify the effect network structure has on the proliferation of fake news, we develop a mathematical model of fake news sharing and test it on a variety of social networks. There is an established tradition of using spatial game theory to study problems of coordination and collective action, particularly the Prisoner's Dilemma, and the structure of the population playing games has been found to drive the behavior of the system, encouraging or discouraging good behavior depending on the network's properties [216, 292]. The evolution of the system can also exhibit interesting spatial phenomenon that is

not present in the well-mixed case [214]. We use a similar strategy to study the spread of fake news through a social network.

In Chapter 5, we view fake news as another group coordination problem. As individuals, both Democrats and Republicans are quite good at identifying which news sources are accurate and which are not [101, 230], but misinformation is remarkably resilient in large social networks. It is possible that the population breaks into two groups through echo chambers, with echo chambers carrying on a different narrative than the population at large [298]. Inspired by work studying public goods games [145,267,268], we will examine how positive and negative feedback can be useful tools to stop the spread of fake news and help the population reach an accurate consensus opinion.

Chapter 2

Random Choices Facilitate Solutions to Collective Network Coloring Problems by Artificial Agents

Section 2.1

Introduction

There is a rich history of playing games to model interactions in the presence of some social structure [46, 52, 53, 74, 167, 187, 248, 265, 287]. For example, studying how this structure impacts player behavior has been a particularly useful area for those interested in fostering certain kinds of behavior like cooperation by allowing punishment [73] or partner choice [119], among others. To simulate this spatial structure, we can treat the population as a network where each node is an individual, and individuals play games if they are connected by an edge [130, 216, 233, 258, 266].

On such a network, many coordination games can be cast as network coloring

problems [162]. On the surface, graph coloring is an abstract question about different ways to label vertices. However, if we think of vertices as individuals (which we refer to as artificial agents in this work) and the color choice representing the strategy of that individual, a graph coloring can have game theoretic meaning [162, 167, 186].

In the Radio game (Table 2.1), two radio station owners must each choose to broadcast on one of two frequencies. If they choose different frequencies, both radio stations have good sound quality, but if they choose the same frequency, they interfere with each other and no one can listen to either station.

		Player 2	
		A	B
Player 1	A	$(0,0)$	(a,a)
	B	(a,a)	$(0,0)$

Table 2.1: The payoff matrix for the radio game. If both players broadcast on the same frequency, they get no benefit. Instead, they both get a benefit a when they anti-coordinate, so one individual plays A and one plays B.

When playing the Radio Game (Table 2.1) on a network, a valid coloring of the network where adjacent vertices have different colors represents a social optimum; there are no longer any neighbor pairs playing the same strategy and getting decreased payoffs.

2.1.1. Coordination Problems as Distributed Graph Coloring Problems

Graph colorings have been applied to theoretical problems, inluding register allocation in computer science [70] and network clustering problems [140], but there are also many real-world social systems that can be studied through graph colorings. Deciding on a time table for various classes with shared classrooms [88,256], assignment of radio frequencies [224, 316], and contrarians or "hipsters" who make choices specifically to distinguish with those around them [21, 163] are a few examples of anti-coordination games that manifest naturally as network coloring problems. When the nodes of a

network are properly colored, all the individuals are playing an optimal strategy. In this sense, the network coloring problem, if assigned with a proper payoff structure for the coloring outcome, can be considered broadly as collective action problem [15,170].

Just like proper colorings can be used to study anti-coordination games, uniform colorings, in which each individual chooses the same color, can be a useful model of the coordination game. The voter model is a classic example of individuals in a networked population playing a coordination game using myopic update rules in an attempt to reach consensus with those around them using only limited local information [174, 277, 300].

This connection between real-life anti-coordination games and graph colorings allows us to bring all the techniques for finding graph colorings to bear on real scenarios [37, 158, 310]. Unfortunately, the graph coloring problem is NP-hard [124]. Many difficult mathematical problems cannot be solved by a simple, direct approach, but it can sometimes help to apply a small degree of randomness to any algorithms searching the solution space. This approach has been used to all sorts of problems, including the Traveling Salesman Problem [42] and the graph coloring problem [158] with which we are concerned in Chapters 2 and 3. It is noteworthy that, more broadly, effects of noises on phase transitions and collective outcomes have been studied in diverse contexts, including consensus in opinion formation and evolution [235, 282], ordering in Kawasaki dynamics [165], cooperation in evolutionary games [231, 290], and convergence in combinatorial optimization problems [60], to name a few.

Attempts to solve the network coloring problem typically use information about the entire network to make decisions about the colors of nodes. This is a good idea, since having all the information simultaneously leads to better informed decisions. For example, in [158], Johnson, Aragon, McGeoch, & Schevon use a notion of temperature to gradually reduce stochastic behavior. As the system "cools," random behavior

decreases so the system can settle into the global solution after fully exploring the state space in the early stages. This requires some central information unit that instructs each node on color choice, but if we are using the network as a model of a population in which edges represent interactions, such a central "brain" may not exist. Instead, individuals may be forced to make decisions based on nothing except the color of their neighbors at any given moment. Thus, solving the distributed network coloring problem, in which each node decides its color with only the local information about its neighbors, is more difficult, as we lose the ability to make decisions based on the state of the entire network. This introduces new complications to the classic problems, and stochastic behavior is often needed to successfully find an n-coloring of the graph [160].

In recent years, there has been a growing interest in studying *distributed coloring problems*. One line of work involves deterministic algorithms that require more colors than necessary for the network [71, 110]; the additional available colors make the problem much more tractable. There has also been work involving experiments with human subjects who have been given control of the color of a single node, and are asked to choose colors to eliminate conflicts with their neighbors. In [167], Kearns, Suri, and Montfort observed that individuals would frequently choose colors that temporarily increased the total number of color conflict, but ultimately lead to a global coloring. Following this, in [264], Shirado and Christakis found that by adding a small number of bots (namely, artificial agents as opposed to humans) to the system who periodically made random changes "decreased both the number of conflicts and the duration of unresolvable conflicts" when finding network colorings. However, they also found that the bots could be detrimental if not properly tuned with the appropriate levels of randomness. Along this line, a recent related modeling work has incorporated reinforcement learning algorithms (q-bots) into agent-based simulations

of the distributed coloring problem [246]. Despite these developments, there still is a lack of analytical insights into the optimal level of random behavior needed when solving network coloring problems.

To provide further analytical insights into the role of behavioral randomness in finding solutions to the distributed coloring problem, Chapters 2 and 3 focus on "myopic artificial agents" attempting to solve network coloring problems using decision update rules that are only based on local information but allow random choices at various stages of their heuristic reasonings. While most previous work used three-colorable graphs [167, 264], we assume agents are situated on networks that can be colored with only two colors, often called bipartite networks [135]. This specific network structure simplifies the number of possible colorings (exactly two for a connected network) and offers analytical insights that would be formidable to obtain otherwise. For an omniscient observer that can view the entire graph and dictate colors to vertices, finding one of these 2-colorings is a trivial matter. However, things become more difficult when there is no central decision-maker, and instead each vertex represents an individual who must choose her own color with no information except the colors of her neighbors. The results reported below come from an entire population of artificial agents (in the fashion of simulated bots as in Ref. [264]), some of whom are behaving deterministically and some stochastically. Our work in Chapter 2 sheds some light on the appropriate levels of randomness to optimize solving the distributed coloring problem.

This study was published in [160].

Section 2.2

Results

2.2.1. Random Network Construction

As we will see, different network topologies will be easier or harder to color. Even with global information, finding network colorings becomes exponentially more difficult as the number of nodes increases [124]. On the other hand, as average degree increases, individuals will have more neighbors and therefore be able to make more informed decisions when choosing a color. Throughout this chapter, we simulate artificial agents that attempt to find 2-colorings of random bipartite networks. The exact structure of these networks will vary, as will the decision update rules agents use to solve the network colorings.

We construct a random network with n nodes and expected degree k by first assigning each node to group A or group B with probability $\frac{1}{2}$. Then, we add an edge between any two nodes in different groups with probability $\frac{2k}{n}$. Thus, the resulting network is guaranteed to have a 2-coloring by assigning every node in group A one color and every node in group B the other color. However, there may be different numbers of nodes for each color, as the sizes of groups A and B are binomially distributed in our bipartite network model.

2.2.2. Decision Update Rules for Agents

In this chapter, we consider multiple decision update rules to account for a variety of artificial agents' behavior, each with their own strengths and weaknesses. In the following, an *acceptable* local coloring at a node is the choice of color such that none of the node's neighbors have that color (no color conflicts with neighbors).

We first consider a basic *greedy* update rule of agents:

I: Basic greedy update rule

Step 1: Check if the current color is already an acceptable local coloring. If yes, keep the current color for this update step. If not, advance to Step 2.

Step 2: Check if the other color would make an acceptable local coloring. If yes, change to that color. If not, advance to Step 3.

Step 3: Choose whichever color will minimize the number of color conflicts. If both colors will create the same number of color conflicts with neighbors, randomly choose one color.

As the goal of each agent is to reduce and ultimately eliminate color conflicts with their neighbors, the greedy update rule can be seen as the *rational* strategy for an agent playing every single round of the coloring game, and is therefore implemented as the "default" strategy in our population. We incorporate random behavior in various decision stages in the following modified update rules based on the basic greedy update rule above. Notably, these simple yet natural update rules based on intuitive heuristics, combined with the bipartite network structure which simplifies the possible colorings, enable analytical insights that are unobtainable in the more complicated systems put forward in other work [246].

II: Randomness-first update rule

Step 1: With probability p, choose a color uniformly at random. With probability $1-p$, advance to Step 2.

Step 2: Check if the current color is already an acceptable local coloring. If yes, keep the current color for this update step. If not, advance to Step 3.

Step 3: Check if the other color would make an acceptable local coloring. If yes, change to that color. If not, advance to Step 4.

Step 4: Choose whichever color will minimize the number of color conflicts. If both colors will create the same number of color conflicts with neighbors, randomly choose one color.

III: Memory-0 update rule

Step 1: Check if the current color is already an acceptable local coloring. If yes, keep the current color for this update step. If not, advance to Step 2.

Step 2: Check if the other color would make an acceptable local coloring. If yes, change to that color. If not, advance to Step 3.

Step 3: With probability p, choose a color uniformly at random. With probability $1-p$, advance to Step 4.

Step 4: Choose whichever color will minimize the number of color conflicts. If both colors will create the same number of color conflicts with neighbors, randomly choose one color.

IV: Memory-N update rule

Step 1: Check if the current color is already an acceptable local coloring. If yes, keep the current color for this update step. If not, advance to Step 2.

Step 2: Check if the other color would make an acceptable local coloring. If yes, change to that color. If not, advance to Step 3.

Step 3: If no neighbors have changed colors in prior N cycles, with probability p, choose a color uniformly at random, and with probability $1-p$, advance to Step 4.

If any neighbors have changed colors in prior N cycles, advance to Step 4.

Step 4: Choose whichever color will minimize the number of color conflicts. If both colors will create the same number of color conflicts with neighbors, randomly choose one color.

2.2.3. Initialization of Agent Behavior

Each artificial agent, located at a node in the network, behaves according to one of the aforementioned update rules. Specifically, we consider scenarios where the population may be using two different update rules. A certain fraction ρ_r of randomly-selected agents adopt one of the randomness-first, memory-0, or memory-N update rules where the propensity of random behavior is p (as defined in the update rules), and the rest of agents use the basic greedy update rule.

The color choice of agents is updated in a random sequential manner [287]. Agents update one at a time, and the order in which agents update is random. Each agent begins with a randomly chosen color.

2.2.4. Difficulty Metrics

We use three different metrics to quantify how successful a given decision update rule is in solving coloring problems by artificial agents: the number of unsolved networks, the number of update cycles, and the number of player updates.

The number of unsolved networks metric is simply the probability that a given network will reach a coloring given certain initial conditions including update order, update rules for each agent, and initial coloring.

The number of update cycles measures the number of times each agent goes through the update process, and the number of updated agents measures the total number of color changes. Roughly, the number of update cycles measures how long it will take the system to reach a coloring in real time, and the number of updated agents measures how involved the process is for all agents involved. Because

some combinations of networks and initial conditions may never end up with a complete coloring solution, these metrics have the possibility to be infinite in these cases. Therefore, the average of difficulty metrics across model parameter combinations may be heavily skewed by some of the unsolved network coloring cases. Nevertheless, these difficulty metrics provide a practical means to compare the efficacy of resolving color conflicts across simulated scenarios and can help reveal interesting results to some extent.

2.2.5. Bowties and Gridlock

To see how local minima arise, we show a small network in which each agent occupying a network node uses the greedy update rule in Fig. 2.1a. The dashed edges are "bowties," small subgraphs consisting of a central edge whose end nodes both have at least three edges. Motif structures like this can lead to gridlock and the failure of the greedy update rule, as demonstrated in Fig. 2.1b. If the central agents are playing the same color, they can become locked in by their other neighbors, and as a consequence, the greedy update rule becomes trapped at this local minimum, unable to explore the entire space and find a global minimum of color conflicts. Without random behavior, the network will never reach a global coloring once this happens. The smallest possible network structure that can become gridlocked is the six-node bowtie, as shown in Fig. 2.1b.

This simple case demonstrated in Fig. 2.1b can yield an interesting insight. Consider the case where there is no random behavior and each agent is playing the greedy update rule. There are $6! \cdot 2^6$ random initial conditions for the update order and initial colors. Using exhaustive search to work out each case, we find that the simple bowtie results in gridlock with probability $\frac{29}{120}$. In each case, either gridlock or a global coloring is always reached after two update cycles.

Of course, brute-force computation quickly becomes untenable for large network

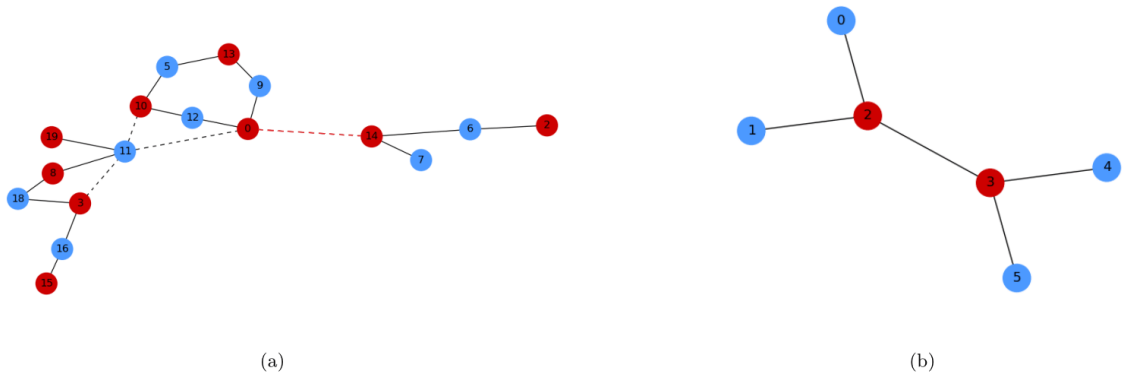

Figure 2.1: **Overcoming local minima is often needed to solve collective action problems.** (a) shows a small network that did not find a valid coloring using only greedy behavior. The four dashed edges represent bowties, subgraphs where the greedy update rule can become gridlocked. The red edge shows a color conflict that cannot be resolved by greedy behavior. In (b), we see how the interior nodes of a bowtie are both forced to keep the same color by the exterior nodes, creating gridlock.

sizes, but we can still develop helpful intuition from this simple example (Fig. 2.1b). With the randomness-first update rule, if at least one agent has random behavior (occurs with probability $1-(1-\rho_r)^6$), the network will eventually find a global coloring. However, in the memory-N update rule, the peripheral nodes already have a locally acceptable color, and will not change even if they have the potential for random behavior. One of the middle two nodes must have random behavior to find a coloring, which happens with probability $1-(1-\rho_r)^2$, a much less likely event than in the randomness-first update rule. Thus, the gridlock probabilities for the randomness-first and memory-N update rules respectively are approximately

$$P_{\text{rand-first}}(\text{Gridlock}) = \frac{29}{120}(1-\rho_r)^6 \tag{2.1}$$

$$P_{\text{memory-}N}(\text{Gridlock}) = \frac{29}{120}(1-\rho_r)^2 \tag{2.2}$$

We see excellent agreement between these equations and simulations in Fig. 2.2. We note that these probabilities are less accurate when p is large, because individuals

could behave randomly before the system reaches gridlock, disrupting the earlier computation for $\frac{29}{120}$ which assumed no random behavior takes place in the first two update cycles.

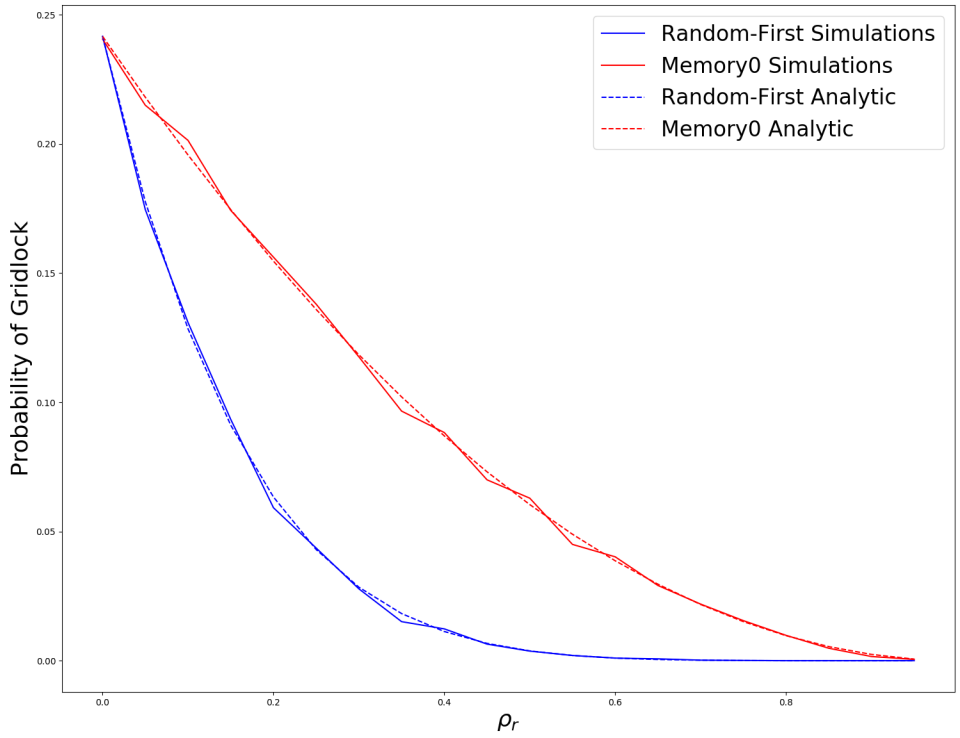

Figure 2.2: **The probability of gridlock in the six-node bowtie for varying the fraction of agents with random behavior, ρ_r.** We see that the simulations (using $p = 0.5$) match well with the analytic results in Eqs. 2.1 and 2.2. Here we compare the randomness-first rule with the memory-0 rule. Simulation results are averaged over $1,000$ independent runs.

Similarly, we see that the memory-N update rules require larger ρ_r than the randomness-first rule to reach the same efficacy of resolving color conflicts. When using the former update rule, only agents with a color conflict are allowed to make random choices, unlike the latter randomness-first update rule. Because random behavior is limited to individuals with a color conflict, large ρ_r values are less likely

to result in too much randomness when most agents are already in a local coloring without conflicts and hence will not behave randomly in any given time step. We shall see this difference between randomness-first and memory-N update rules manifest itself in simulations on larger networks in the following section.

2.2.6. Monte Carlo Agent-Based Simulations

Having defined the model parameters for the problem, we now can ask a basic question: What is the optimal amount of randomness to have in the system so as to reach a coloring solution? It turns out that the answer varies, depending on the specific update rule used, the size of the network, and the average degree of the underlying network. Typically, we will consider large and small networks with 50 and 500 nodes, and vary with average network degree values of 2 and 20, respectively. Figure 2.3 shows how noisy agents using different update rules succeed at reducing the total number of conflicts in different situations. Notice that no update rule alone can beat the greedy update rule in the short term, but eventually the randomness-based update rules begin under-performing the greedy rule only to eventually surpass it and completely eliminate color conflicts.

There are two sources of difficulty for coloring networks using any randomness-based update rule. If there is not enough randomness, the decision update rule is unable to break away from the local minimum found by agents using the greedy update rule. If there is too much randomness, the probability that at least one agent will be picking the wrong color every turn is so high that the network will not find a coloring in a reasonable number of time steps. Methods like simulated annealing avoid this problem by cooling the system and decreasing the amount of randomness over time [158]. However, in a distributed system (where each agent is using only local information to choose color) with no global information like temperature, we are limited to very simple local update rules that simply cannot evolve over time.

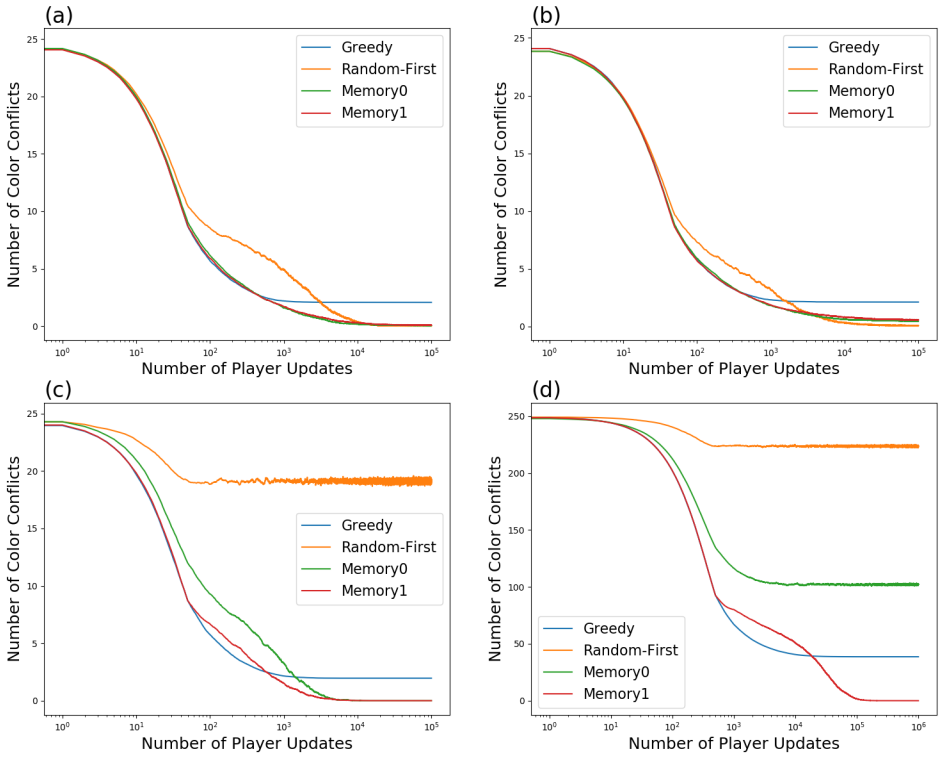

Figure 2.3: **Plots of total conflicts vs time.** Each curve is the average of $1,000$ simulations, and each run consists of $2,000$ update cycles. Observe that the x-axis is log-scale, to show the short and long term behavior of each update rule. All networks have average degree 2, and the other network properties are as follows: a) $n = 50, p = 0.1, \rho_r = 0.9$ b) $n = 50, p = 0.1, \rho_r = 0.5$ c) $n = 50, p = 0.6, \rho_r = 1$ d) $n = 500, p = 0.6, \rho_r = 1$

2.2.7. Randomness-first rule

For the randomness-first update rule, we ran simulations for 20 combination values of ρ_r and p between 0 and 1. Networks that found a coloring within $10,000$ update cycles by agents were considered solved, and those that did not find a coloring within $10,000$ cycles were considered unsolved. In Fig. 2.4, we show the results of these simulations.

We see the difficulty of too much and too little randomness in Fig. 2.4. In all

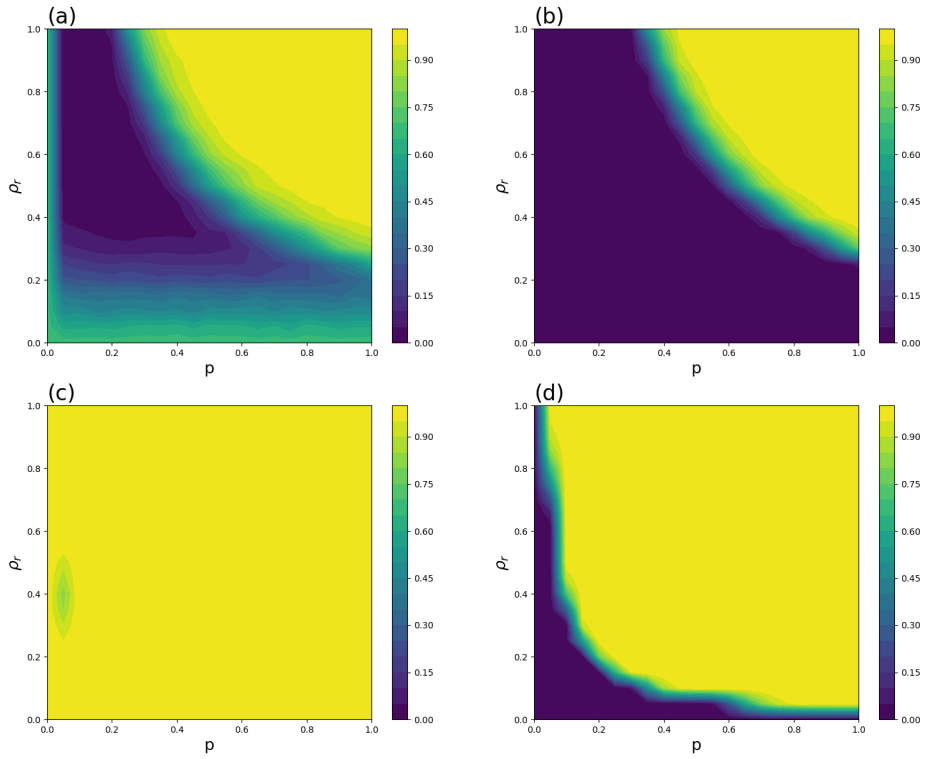

Figure 2.4: **For the randomness-first update rule, simulation results of the probability of not solving the network in** $10,000$ **time steps using four different types of networks as a function of the level of randomness** p **and the fraction of agents with random behavior** ρ_r**.** The bipartite network parameters including the size N and the average degree k used for the underlying networks are as follows: a) $n = 50, k = 2$ b) $n = 50, k = 20$ c) $n = 500, k = 2$, d) $n = 500, k = 20$.

four regions of the network parameter space (small/large size, low/high edge density) the probability of solving the network goes to zero because agents are always making random decisions, even when the rest of the network has found a local coloring. When average degree is two, we also see unsolved networks when there is very little randomness. Here, there are too few random agents to break out of the local minimum.

These results demonstrate how the randomness-first update rule's success varies

depending on the properties of the network (Fig. 2.4). When average degree is high, randomness is actually a hindrance; the fewer random actions there are, the better. However, when average degree is low, a large fraction of the population using the randomness-first update rule with a low p is best. Unfortunately, for large networks with small average degree, there seems to be no good p and ρ_r when using the randomness-first rule.

Notice that in general, as network size goes up and/or average degree goes down, there are more unsolved networks. This makes intuitive sense, as additional nodes means more colors that need to be correct, and smaller average degree means the nodes have less information and make poorer decisions.

2.2.8. Memory-N rules

We first study the memory-0 update rule that differs from the randomness-first rule in that agents only take random actions if they are in conflict with at least one of their neighbors. Thus, there are fewer needless random actions, and we would expect this decision update rule to perform better where excess randomness is an issue. This is partially confirmed by simulations in Fig. 2.5.

Generally, we see an improvement of performance over the randomness-first update rule. The memory-0 rule does very well when ρ_r is close to one, even for large networks with low average degree. However, it still struggles with excess randomness, particularly when network size and average degree are large. A higher average degree means that a single random color choice creates more color conflicts and therefore makes it more difficult for the system to settle into a global coloring. However, if we assume agents with a longer memory (i.e., $N \geq 1$), this issue vanishes, as demonstrated in Fig. 2.6.

This compelling evidence suggests that the memory-1 update rule is most effective at resolving color conflicts as compared with the basic greedy update rule and the

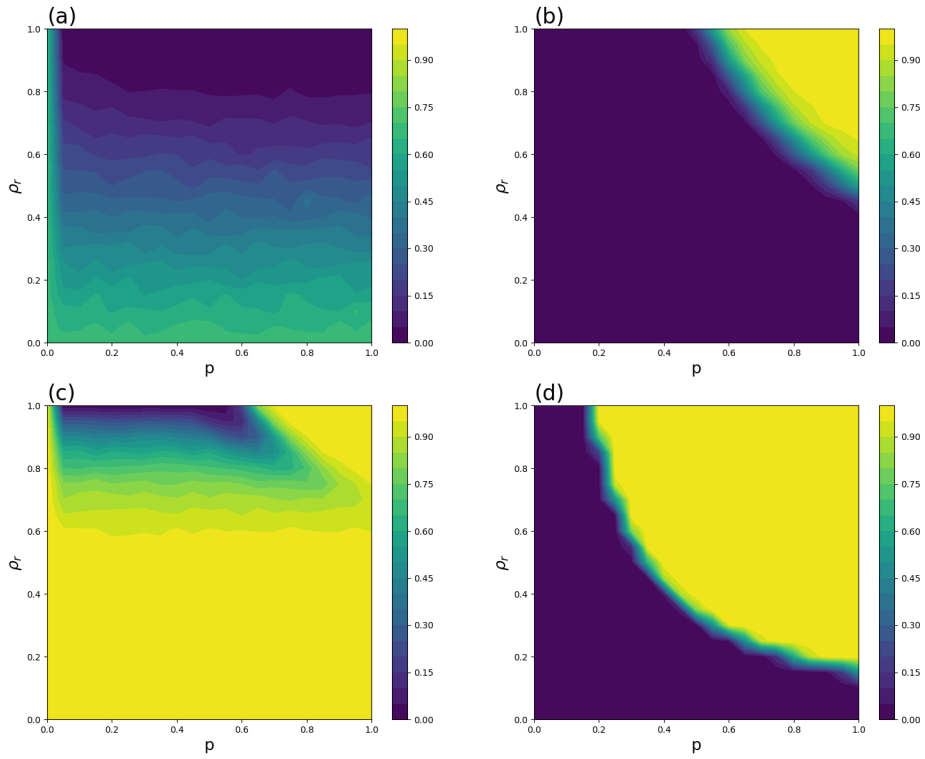

Figure 2.5: **For the memory-0 update rule, simulation results of the probability of not solving the network in $10,000$ time steps using four different types of networks as a function of the level of randomness p and the fraction of agents with random behavior ρ_r.** The bipartite network parameters including the size N and the average degree k used for the underlying networks are as follows: a) $n = 50, k = 2$ b) $n = 50, k = 20$ c) $n = 500, k = 2$, d) $n = 500, k = 20$.

memory-0 update rule (cf. Figs. 2.4, 2.6, and 2.5). If ρ_r is close to one, networks are almost always able to find a global coloring, regardless of network size or average degree. However, if for some reason only a rather small fraction of the agents are allowed to use randomness-based update rules, the randomness-first update rule will have more success, as seen in the simple bowtie example in Fig. 2.1b.

The memory-1 update rule is extremely effective in networks with high connectivity. Similar effects of connectivity on graph colorablity, albeit using the Brélaz's

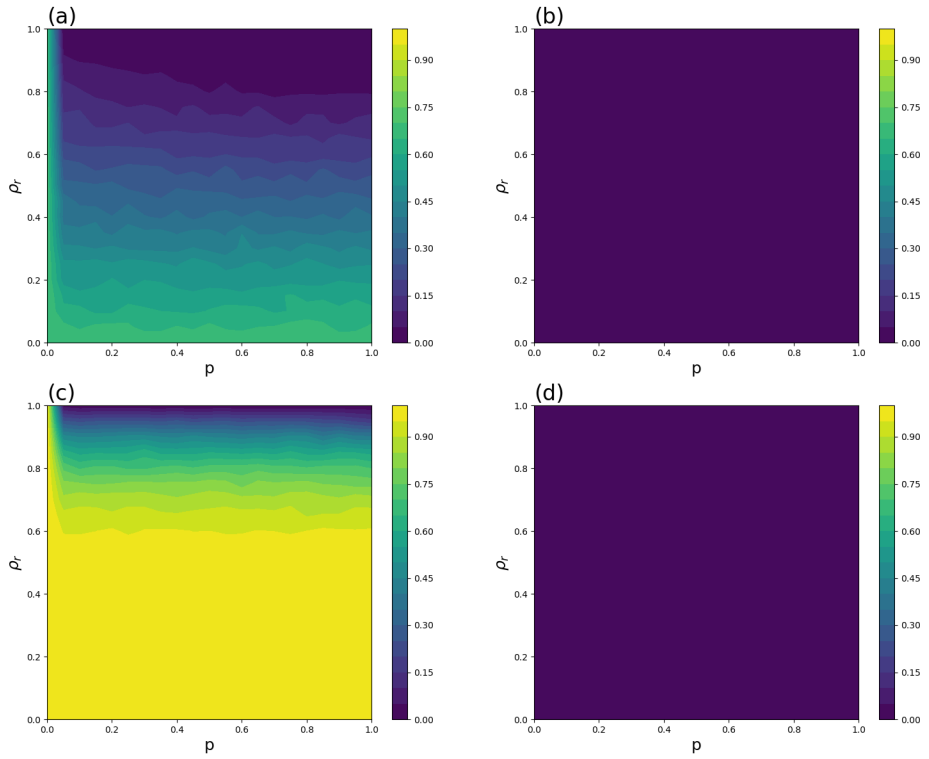

Figure 2.6: **For the memory-1 update rule, simulation results of the probability of not solving the network in 10,000 time steps using four different types of networks as a function of the level of randomness p and the fraction of agents with random behavior ρ_r.** The bipartite network parameters including the size N and the average degree k used for the underlying networks are as follows: a) $n = 50, k = 2$ b) $n = 50, k = 20$ c) $n = 500, k = 2$, d) $n = 500, k = 20$.

heuristic algorithm [48], have been observed in coloring small-world networks [285]. When average degree is $k = 20$, every individual's local information encompasses the color of a large number of neighbors, allowing individuals to make very informed decisions. Additionally, individuals with a high number of edges are able to see a lot of potential color changes. If the population is large and over-randomness is a concern, this is lessened when individuals can see a wide range of neighbors, and will not randomly change color if they see one of their many neighbors changing. Thus,

the system will be allowed to settle into the global solution, even when individuals playing random update rules are likely to choose a random color.

Section 2.3

Discussion

The 2-coloring problem, while trivial on a global scale, represents new challenges when solved by a population of agents that have only limited local information. When an agent only sees a small fraction of the entire network, they can be led astray into making myopic decisions that are non-optimal for the population at large. On the other hand, agents making random decisions, however infrequently, can serve to perturb a system that is stuck at a local minimum, thereby breaking up gridlock and moving the population toward the desired global coordination.

Among others, an important insight stemming from this work is that the type of decision update rule used by agents is at least as important as the amount of random behavior. The randomness-first and memory-N update rules require different conditions to be successful. This gives us two different update rules that are useful in different settings, and should be thought of as complementary instead of one being superior to the other. For example, in a scenario where all agents are able to use a randomness-based update rule, a memory-N update rule can be used to great success. However, if only a few agents in the population can be persuaded to take on the personal risk of behaving randomly (or a small number of bots prescribed with random behavior have been introduced into the population like [264]), a randomness-first update rule with a low p will have a higher chance of success.

Section 2.4

Limitations of the Study

The work presented in this chapter most closely relates to previous work involving human subjects playing the coloring game with random bots [264]. While random behavior was observed coming from human players [167], it is not clear if this behavior was closer to the randomness-first or the memory-N update rule. The noisy bots themselves in Ref. [264] played a randomness-first update rule, which may explain how such a small fraction ($\rho_r = 0.15$) of random actors had such a profound impact on the network coloring game. While the artificial agents in this work may not fully capture sophisticated human behavior, they indeed encompass the essence of random exploration ubiquitous in human decision-makings, as demonstrated in prior observations of human decision choices in game theoretical interactions [296]. It is thus promising for future work to leverage existing data such as in Ref [264] to further validate and refine the stochastic decision update rules presented in this chapter.

Our work demonstrates that the solving of challenging distributed network coloring problems can be achieved by entirely using myopic artificial agents without human subjects. We find that it is necessary to have enough randomness to ensure that the system is able to find the global coloring, but without having so much random behavior the system never settles down. That said, certain randomness-based update rules can be particularly successful, depending on the underlying population characteristics (see Table 2.2). In this regard, our findings as summarized in Table 2.2 can be used to inform future hybrid experiment design.

Of particular note, here we only consider the simplest possible 2-colorings of bipartite networks, which is surely an over-simplification of the more general coloring problems. Introducing even one more color adds all sorts of complications. For exam-

Population characteristics	Optimal update rule
Small ρ_r	Randomness-first update rule
Large ρ_r, small n	Memory-0 update rule
Large ρ_r, large n	Memory-N update rule, $N \geq 1$

Table 2.2: **Population characteristics determine the efficacy of each decision update rule in solving collective network coloring problems.** This table summarizes how ρ_r, the proportion of agents with random behavior, and n, the size of the population, can impact which stochastic update rule used by noisy agents will work best together with the remaining greedy agents.

ple, the bowtie analysis completely falls apart, as the subgraphs to result in gridlock in a 3-colorable network are significantly larger and more complex. Besides, this study also only considers populations that play a mix of two decision update rules: a fraction of the agents use greedy decision rule and the rest use randomness-based rule. It is possible that other potential combinations, such as a mixed population of agents using the randomness-first rule and the memory-N rule, could succeed in places where neither update rule succeeds alone. Future work taking into account these extensions will be of interest and improves our understanding of collective decision-makings in the presence of noises [83, 84], and more generally, machine behavior [247].

Chapter 3

The Dual Problems of Coordination and Anti-coordination on Random Bipartite Graphs

Section 3.1

Introduction

The main result of this chapter is that the distributed coloring problem has an unexpected property: in the context of bipartite graphs, finding a 2-coloring of the graph, which models an anti-coordination game, is equivalent to getting all individuals in the graph to choose the same color, which is a coordination game. Thus, anti-coordination games and coordination games are dual problems, and a whole new class of coordination games where everyone wants to opt for the same strategy can also be modeled as a graph coloring problem. We show this novel result by defining two Markov chains on the space of colored graphs, one where individuals are playing

the anti-coordination game and one where individuals are playing the coordination game, and showing that they are isomorphic.

3.1.1. Coordination Games

In Chapter 2, we presented the Radio game (Table 2.1) as one example of an anti-coordination game. Here we introduce two of the most famous and well-studied coordination games: the Stag Hunt and the Bach-Stravinsky game.

		Player 2	
		S	H
Player 1	S	$(\frac{s}{2}, \frac{s}{2})$	$(0, h)$
	H	$(h, 0)$	$(\frac{h}{2}, \frac{h}{2})$

Table 3.1: The payoff matrix for the Stag Hunt game, where a large payoff s is split between both players if both choose S. Otherwise, a payoff of h is divided between the players that choose H.

Table 3.1 gives the payoff matrix for the Stag Hunt [34, 272, 273]. In this game, two hunters would like to work together to hunt a large stag (S) but each hunter can decide to hunt hares (H) instead, which can be done alone. A hunter who tries to hunt the stag alone fails and gets nothing. In this game, both players would prefer to hunt the stag together, but choosing S runs the risk of getting a payoff of zero if the other hunter chooses H. Therefore, it is sometimes called the "Assurance Game" because both players would like some sort of assurance that the other will also choose to hunt the stag before making their choice.

Table 3.2 shows the payoff matrix for the Bach-Stravinsky game (originally called Battle of the Sexes [181, 221]) in which two players must decide between going to a Bach (B) or a Stravinsky (S) concert. Player 1 prefers Bach and Player 2 prefers Stravinsky, but they both would rather go to the same concert than listen to their preferred composer. The difficulty here for Player 1 is in persuading Player 2 to abandon their preference so that Player 1 gets a maximum payoff. Of course, the

		Player 2	
		B	S
Player 1	B	$(a+b, a)$	$(0, 0)$
	S	$(0, 0)$	$(a, a+b)$

Table 3.2: The payoff matrix for the Bach-Stravinsky game. Each player gets a benefit a if they go to the same concert, and the player who goes to their preferred concert gets an additional benefit b.

worse case scenario is when both individuals go to their less-preferred concerts, and both get zero payoffs.

Unlike the Stag Hunt, the Bach-Stravinsky game has symmetric strategies, meaning there is no "risky" and "safe" strategy. In this paper, we will consider a coordination game where each player maximizes payoff by choosing the same strategy as their neighbors, and no player has any inherent preference for one strategy or the other. This can be modeled with the payoff matrix in Table 3.2 with $a = 0$.

3.1.2. Markov Chains

Another tool that can be used to study social systems is the concept of a *Markov chain* from probability theory [98, 122, 133]. Formally, a discrete time Markov chain is a sequence of random variables X_n (each with the same state space Ω) whose probability distribution only depends upon the realization of the previous random variable X_{n-1}; i.e.

$$P(X_n = y_n | X_1 = y_1, X_2 = y_2, \ldots, X_{n-1} = y_{n-1}) = P(X_n = y_n | X_{n-1} = y_{n-1}). \quad (3.1)$$

Informally, a Markov chain is a system that evolves over time that is "memoryless", meaning only the present state is relevant for the next state of the system. Because many natural processes are memoryless, Markov chains have proved useful

for building models and analyzing systems in disciplines ranging from literature to physics, computer science to DNA [76, 143, 303].

A process that depends on a finite number of previous states can be turned into a Markov chain by expanding the state space to be ordered tuples of the previous states. In this way, for any finite k, a sequence of random variables where X_n depends on $X_{n-1}, X_{n-2}, \ldots X_{n-k}$ can still be modeled as a Markov chain by expanding the state space to be a list of the last k values.

If the number of possible states is a finite integer m, then the *transition matrix* P is the $m \times m$ where P_{ij} is the probability of moving from state i to state j. The transition matrix contains all the information that makes up the Markov chain, and just like the adjacency matrix of a network, it allows us to use all the tools of linear algebra to study the chain. For an introduction, see [133].

However, in systems with a large number of states (including those in this thesis), these transition matrices are simply too large. Enumerating all the states would take a tremendous amount of time and it would take an enormous amount of computing resources just to store such a large matrix. In such cases, simulations can be used to get approximate results with much less computational effort.

This work extends our study of the distributed coloring problem from Chapter 2 and has been published in [159].

Section 3.2

Theoretical Results

3.2.1. A Natural Bijection for Update Rules for 2-colorings and Uniform Colorings

In this chapter, the individuals located at each vertex will operate using a simple set of update rules. These rules can incorporate random behavior, but the update decisions

depend only on the color of an individual's neighbors. Consider the relationship between update rules for anti-coordination and coordination games. We will see that any update rule for an individual playing an anti-coordination game can be adapted to an update rule for playing a coordination game and vice versa. At its most basic, an anti-coordination rule aims to minimize the number of neighbors with the same color, and the goal of a coordination rule is to maximize the number of neighbors with the same color. Therefore, we can turn an anti-coordination update rule into a coordination update rule just by picking the opposite color every time.

Suppose we have an individual vertex with a neighbors playing color A and b neighbors playing color B, like in Figure 3.1a. When we define an anti-coordination rule where the central individual will select color A with probability $p(a, b)$ and color B with probability $1 - p(a, b)$, we can make the corresponding coordination rule as follows: choose A with probability $1 - p(a, b)$ and B with probability $p(a, b)$.

Consider an update rule (anti-coordination or coordination) that has a function $p(a, b)$ that gives the probability of choosing color A. If we switch the colors of all neighbors, the probability of choosing A is now $p(b, a)$ because now b neighbors are playing A and a neighbors are playing B. There is a natural restriction to impose on the possible update rules. If we switch the color of every neighbor, moving from Figure 3.1a to Figure 3.1b, the probabilities of the central vertex choosing color A, $p(a, b)$, and color B, $1 - p(a, b)$, should switch as well. This restriction gives us the following complementary condition, by setting the probability of choosing A equal to the probability of choosing B after switching all the neighbors' colors:

$$p(a, b) = 1 - p(b, a) \tag{3.2}$$

For any anti-coordination update rule, a vertex with a A neighbors and b B neighbors will choose A with some probability $p(a, b)$. If we switch the colors of

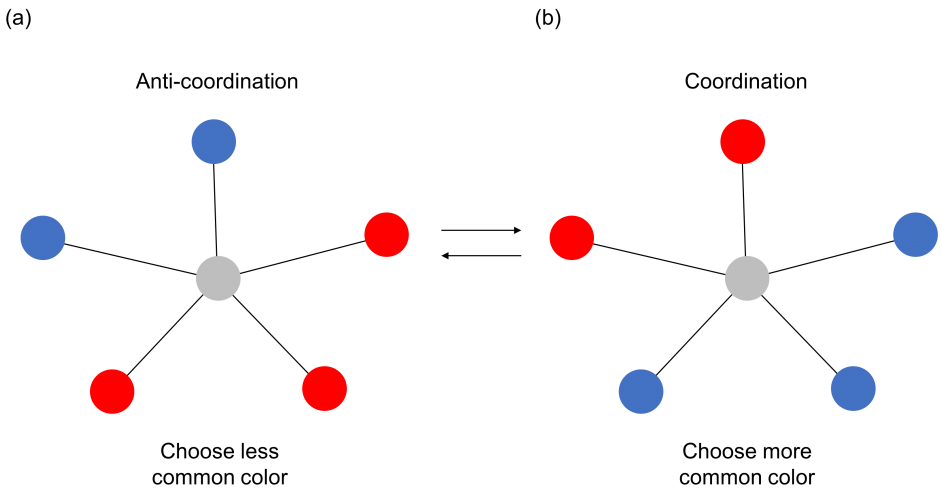

Figure 3.1: A simple case to demonstrate the bijection of update rules with two color choices. Making an anti-coordination decision in (a) will have the same outcome as making a coordination decision in (b), since all the colors of the neighbors have changed to the other color. If an individual would have chosen blue in (a) to match with as few neighbors as possible, that would correspond to choosing blue in (b), where the goal is to match with as many neighbors as possible.

all the neighbors, the vertex will choose A with probability $p(b,a) = 1 - p(a,b)$, but this is equal to the probability of a coordination player choosing A. Therefore, an anti-coordination algorithm can be converted into its dual algorithm for a coordination game by temporarily switching the colors of all the neighbors, using the anti-coordination update rule, and switching the neighbors' colors back. As an example, an anti-coordination update rule on Figure 3.1a will have the same behavior as a coordination update rule on Figure 3.1b.

The same process can be used to convert a coordination algorithm to an anti-coordination algorithm.

To put the above individual choice function $p(a,b)$ in context, it is worthwhile to introduce a few intuitive anti-coordination update rules. The first update rule, called randomness-first, involves making a random choice with probability r, and otherwise with probability $1-r$ makes a color choice that minimizes color conflicts. This update rule can be expressed as:

$$p(a,b) = \begin{cases} 1 - \frac{1}{2}r & a < b \\ \frac{1}{2} & a = b \\ \frac{1}{2}r & b < a \end{cases} \tag{3.3}$$

Under the second update rule, called memory-0, individuals first attempt to choose any color that eliminates all color conflicts. If that is not possible, the individual chooses randomly with probability r and otherwise with probability $1-r$ chooses the color minimizing conflicts with neighbors. In our terms, this algorithm is

$$p(a,b) = \begin{cases} 1 & a = 0 \\ 1 - \frac{1}{2}r & 0 < a < b \\ \frac{1}{2} & a = b \\ \frac{1}{2}r & 0 < b < a \\ 0 & b = 0 \end{cases} \tag{3.4}$$

The third main update rule, called memory-1, is like the memory-0 rule except that the agent only makes a random choice if no neighbors have changed color in the last round of updates. Since this is not a memory-less update rule, it does not have a corresponding $p(a, b)$ function, and the following proof would need to be slightly modified, particularly by significantly enlarging the state space of the Markov chains to include the last N colorings of the graph, to prove the equivalence for update rules with finite memory. While we do not go over all the details of proving that a finite-memory update rule also satisfies the isomorphism, we do show results of computer simulations to demonstrate that the duality of coordination and anti-coordination holds in Section 3.

This is only a small selection of all possible update rules. Any function that satisfies Equation (3.2) and returns values between 0 and 1 could be an update rule, although many would be very ineffective. The three update rules described above are all intuitively reasonable and simple to express, which made them excellent candidates for study in prior work on network graph coloring problems [160]. However, there are other natural update rules that we do not explicitly describe here. For example, an individual may wish to choose each color proportional to the number of neighbors playing that color.

In what follows, we demonstrate that an anti-coordination update rule is exactly

as effective at finding a 2-coloring as the corresponding coordination update rule is at finding a uniform color for the whole bipartite network.

3.2.2. Two Markov Chains

For a connected, bipartite graph G of size N, let $\text{col}(G)$ be the set of all possible labelings of the graph G. Note that here we refer to all ways of labeling the vertices of G with either color A or color B, not just 2-colorings in which no neighbors share the same color.

The system will update as follows: the graph is initialized by randomly assigning each vertex a color. An update order is created that describes the order in which the labelled vertices will update their color. The update order is represented as a list of the numbers 1 through N, which is just a permutation of N elements. The set of all permutations of N elements, called the symmetric group on N elements, is denoted S_N. The vertices continually update their colors in this order, one at a time, until the desired coloring (either a 2-coloring or uniform coloring) is found.

Now we can define our Markov chains. Let $\{X_i\}$ be a Markov chain using an anti-coordination update rule, and let $\{Y_i\}$ be the Markov chain using the associated coordination update rule, as described above. The state space Ω of both chains is the set of ordered triples (G^*, σ, m) where $G^* \in \text{col}(G)$, $\sigma \in S_n$, and $m \in \{1, 2, \ldots, n\}$. Unsurprisingly, G^* represents the colors of the vertices of the graph at some time i. σ is the order in which the vertices update, and m is the current position in the update step.

The state space is quite large, but for each state, there are exactly two states to which the Markov chains can move with non-zero probability, shown in Figure 3.2.

To begin, we initialize both Markov chains (anti-coordination and coordination) by sampling from the uniform distribution Π over Ω, so each starting coloring is equally likely.

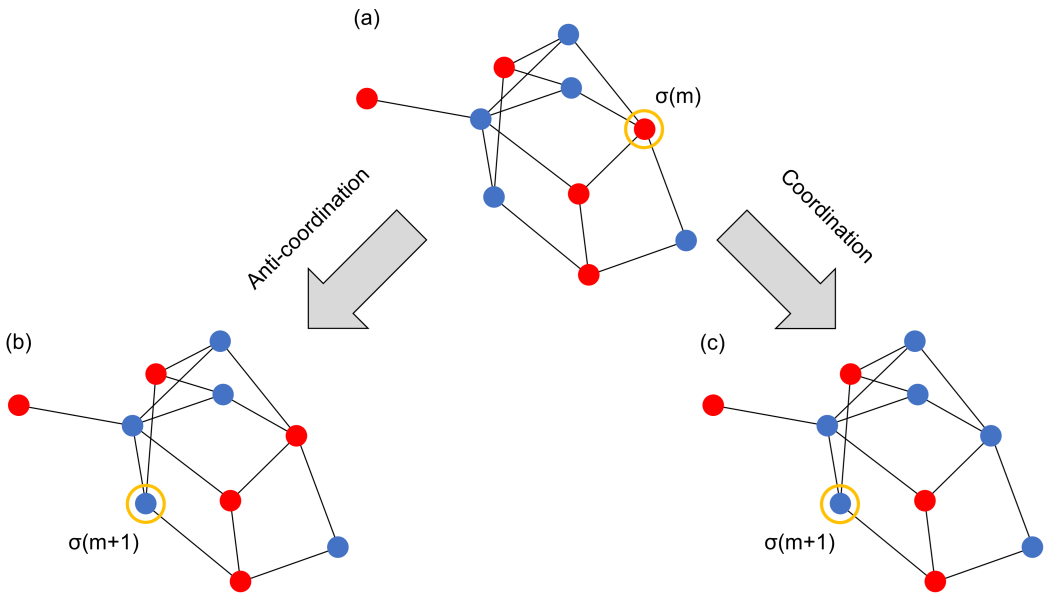

Figure 3.2: A demonstration of the possible transitions in both Markov chains. The next vertex to update is marked by a gold ring. Transitioning from (a) to (b) is minimizing matching with neighbors' colors, and is more likely to appear in an anti-coordination Markov chain, while transitioning from (a) to (c) is matching with as many neighbors as possible, and more likely in the coordination Markov chain.

Without loss of generality, let $X_j = Y_j = (G^*, \sigma, m)$. Here, $\sigma(m)$ is the vertex that is about to update. Let G^*_A be the colored graph that is the same as G^* except possibly $\sigma(m)$ which has color A, and G^*_B the same but for color B. In each step of the Markov chains, $\sigma(m)$ selects one of two colors and the position in the update cycle increases by one, resetting to 1 if necessary. The update order σ remains unchanged. Thus, if $\sigma(m)$ has a color A neighbors and b color B neighbors,

$$P(X_{j+1} = (G^*_A, \sigma, m \bmod(n) + 1)) = p(a, b) \tag{3.5}$$

$$P(X_{j+1} = (G^*_B, \sigma, m \bmod(n) + 1)) = 1 - p(a, b) = p(b, a) \tag{3.6}$$

$$P(Y_{j+1} = (G^*_A, \sigma, m \bmod(n) + 1)) = 1 - p(a, b = p(b, a)) \tag{3.7}$$

$$P(Y_{j+1} = (G^*_B, \sigma, m \bmod(n) + 1)) = p(a, b) \tag{3.8}$$

3.2.3. A Markov Chain Isomorphism

For bipartite graphs, we claim that these Markov chains $\{X_i\}$ and $\{Y_i\}$ are isomorphic. First, because G is a connected, bipartite graph, the vertices can be divided into two groups. In a 2-coloring, all the vertices in the same group will be the same color, and all vertices in different groups will be different colors. Let S be the set of vertices of one of these groups. Because we are working with 2-colorings of bipartite graphs, we can define a function $\phi : \mathrm{col}(G) \to \mathrm{col}(G)$ by switching the color of every vertex in S, and define $\psi_S : \Omega \to \Omega$ as the extension of ϕ in the natural way. We claim that this is a Markov chain isomorphism between X_i and Y_i. This requires proving two conditions hold. First, ψ_S must be bijective. Second, ψ_S must commute with the transition matrices of X_i and Y_i, i.e. the probability of X_i moving from x to y is the

same as Y_i moving from $\psi_S(x)$ to $\psi_S(y)$. More formally, for all $x, y \in \Omega$,

$$P(X_{i+1} = y | X_i = x) = P(Y_{i+1} = \psi_S(y) | Y_i = \psi_S(x)) \tag{3.9}$$

If Equation (3.9) holds, the two Markov chains are equivalent in that after relabelling the states in Ω (according to ψ_S), the Markov chains are identical.

3.2.4. Proof of isomorphism

That ψ_S is bijective is fairly obvious. For any colored graph G^*, because we are only working with 2-colorings on bipartite graphs, $\phi(G^*)$ is well-defined, and only $\phi(G^*)$ maps to G^*, so it is both one-to-one and onto, and therefore ψ is as well.

Now we will prove Equation (3.9). Since we are considering Markov chains moving from x to y (or $\psi_S(x)$ to $\psi_S(y)$), let $x = (G^*, \sigma, m)$. Let a and b be the number of color A and color B neighbors of $\sigma(m)$ in G^*, respectively.

We begin with the conditional statement $X_i = x = (G^*, \sigma, m)$. Equations (3.5) and (3.6) give the only two possibles states of X_{i+1} and their transition probabilities:

$$P\Big(X_{i+1} = (G^*_A, \sigma, m \bmod (n) + 1)\Big) = p(a, b) \tag{3.10}$$

$$P\Big(X_{i+1} = (G^*_B, \sigma, m \bmod (n) + 1)\Big) = 1 - p(a, b) \tag{3.11}$$

Once again, G^*_A and G^*_B are the same as G^* except $\sigma(m)$ which has color A or B, respectively.

Now we consider Y_{i+1} given that $Y_i = \psi(x) = \psi((G^*, \sigma, m)) = (\phi(G^*), \sigma, m)$. $\sigma(m)$ is the next vertex to update, and either it is in the subset S or it is not. These two cases must be handled separately.

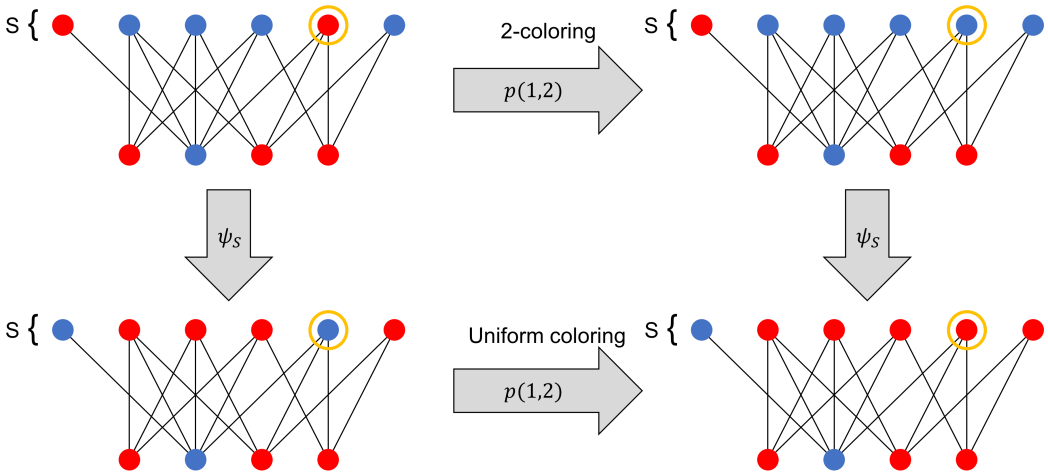

Figure 3.3: An example on a small bipartite graph showing that ψ_S commutes with the transition matricies, when $\sigma(m) \in S$. Color A is blue and color B is red. The top row shows the transition in the anti-coordination Markov chain, and the bottom is the transition in the coordination Markov chain. In both chains, this particular transition occurs with probability $p(1, 2)$.

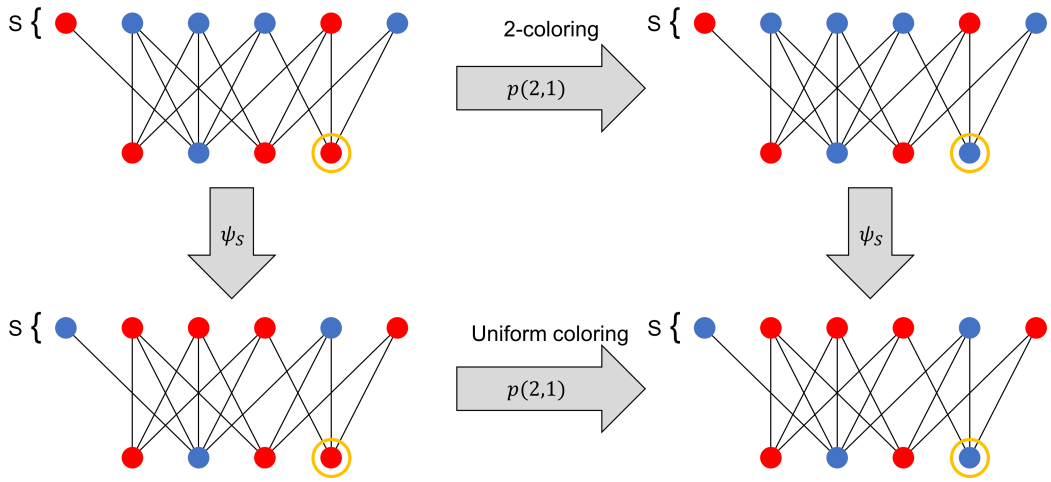

Figure 3.4: An example showing that ψ_S commutes with the transition matricies when $\sigma(m) \notin S$. Color A is blue and color B is red. The top is the anti-coordination Markov chain, and the bottom is the coordination Markov chain. This time, the transition occurs with probability $p(2,1)$.

3.2.5. Case 1: $\sigma(m) \in S$

If $\sigma(m) \in S$, none of $\sigma(m)$'s neighbors are in S, so $\sigma(m)$ still has a color A neighbors and b color B neighbors. Because we are now in the coordination Markov chain $\{Y_i\}$, $\sigma(m)$ chooses its color according to equations (3.7) and (3.8).

With probability $p(a,b)$, $\sigma(m)$ chooses color B. Because $\sigma(m) \in S$, $\phi(G^*)$ becomes $\phi(G^*_A)$ when $\sigma(m)$ chooses B. Thus, $Y_{i+1} = (\phi(G^*_A), \sigma, m \bmod(n) + 1) = \psi(G^*_A, \sigma, m \bmod(n) + 1)$.

With probability $1-p(a,b)$, $\sigma(m)$ chooses color A, and $Y_{i+1} = (\phi(G^*_B), \sigma, m \bmod(n)+1) = \psi(G^*_B, \sigma, m \bmod(n) + 1)$.

Thus, when $\sigma \in S$, Equation (3.9) holds.

3.2.6. Case 2: $\sigma(m) \notin S$

If $\sigma(m) \notin S$, then all of its neighbors are. So in $\phi(G^*)$, $\sigma(m)$ has b color A neighbors and a color B neighbors.

With probability $1-p(b,a) = p(a,b)$, $\sigma(m)$ chooses color A, and $Y_{i+1} = (\phi(G_A^*), \sigma, m \bmod(n)+1)$.

With probability $p(b,a) = 1-p(a,b)$, $\sigma(m)$ chooses B, and $Y_{i+1} = (\phi(G_A^*), \sigma, m \bmod(n)+1)$.

So Equation (3.9) holds when $\sigma(m) \notin S$. Therefore, ψ is a Markov chain isomorphism.

3.2.7. Equivalence of the 2-coloring and uniform coloring problems

Now we are prepared to state and defend the main claim of this work: when using local information, the anti-coordination and coordination problems are equivalent. Any result regarding the efficacy of an update rule $p(a,b)$ for an anti-coordination game can also be applied to a coordination game, and vice versa.

Since the initial distribution Π is the uniform distribution and ψ is bijective, $\psi(\Pi) = \Pi$ and both Markov chains begin from the same distribution. Furthermore, because ψ switches the color of the set S, for any state X_i that is a valid 2-coloring, $\psi(X_i) = Y_i$ is a uniform coloring. For all times i, applying Equation (3.9) i times tells us that moving the anti-coordination chain from a state X_0 to state X_i happens with the same probability as moving the coordination chain from $Y_0 = \psi(X_0)$ to $Y_i = \psi(X_i)$. Because Π is the uniform distribution, for all $x \in \Omega$ and for all times i:

$$P(X_i = x | X_0 \sim \Pi) = P(Y_i = \psi(x) | Y_0 \sim \psi(\Pi) = \Pi) \tag{3.12}$$

Critically, this says that the probability of solving the anti-coordination problem in i steps is the same as solving the coordination problem in i steps, for all i. Ad-

ditionally, the process is linked at each step, so the expected number of player color changes will be the same, for example.

This result also holds for any update rules with finite memory. Any stochastic process whose transition probabilities only depend on a finite number of previous states can be reexpressed as a Markov chain by defining the new state space to be lists of elements from the previous state space, and this works here with any update rule that considers the last n update steps.

Section 3.3

Simulation Results

This result has been confirmed with a variety of simulation results. First, we take a broad approach: We create a large number of different networks, and populating each with individuals playing a particular anti-coordination update rule. Then we repeatedly attempt to find a 2-coloring of the network, collecting data on probability of finding a 2-coloring, the number of update cycles needed, and the number of players updated. Then, using the same network with individuals playing the associated coordination update rule, we repeatedly search for a uniform coloring, collecting data on the same metrics. After repeating this on all the networks, we have a large data collection that, if anti-coordination and coordination games are equivalent, should be two samples of the same probability distribution.

And we see that this is the case using the two-sample Kolmogorov-Smirnov test on data collected from 1,000 different networks. For all three metrics (probability of solving the network, update cycles, and updated players), the K-S statistic is below 0.015 with a p-value greater than 0.999. This strongly suggests that the samples are drawn from the same distribution and the two problems are equivalent.

We can also consider a closer examination of the moment-to-moment behavior of

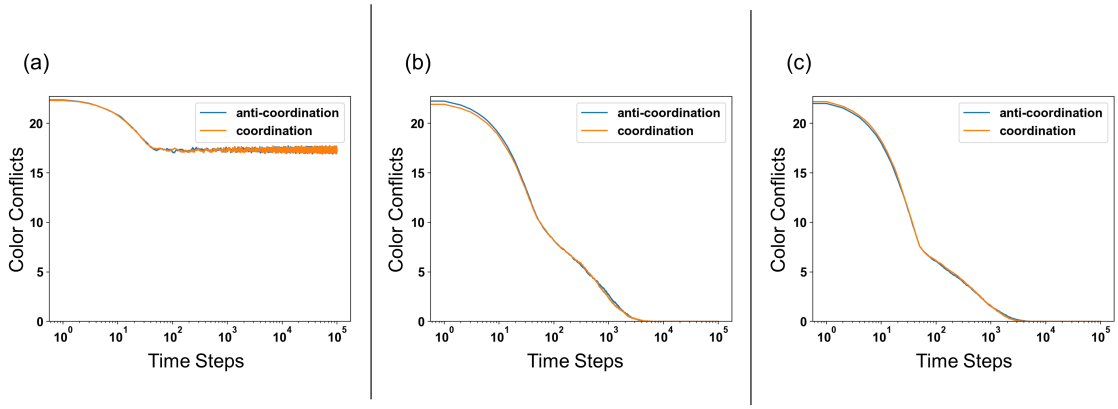

Figure 3.5: Plots showing the time evolution of the number of color conflicts using three reasonable update rules. (a) is randomness-first, (b) is memory-0, and (c) is memory-1. Crucially, the anti-coordination and coordination variants of the same update rule have the same behavior in all three plots. Curves are the average of 1000 simulations for each update rule. For randomness-first, the random behavior probability was 0.5. For memory-0 and memory-1, the random probability was 0.1.

each system by counting the number of color conflicts in the network at every time step, averaged over multiple runs. A color conflict is an edge who ends have the same color (in the case of an anti-coordination game) or different colors (in the case of a coordination game). Previous work [160] dealt mainly with three update rules: randomness-first, memory-0, and memory-1. In Figure 3.5, we see the result of many simulations on the same graph, with these three different update rules. The x axis is log scaled, to clearly show the behavior in the short and long term.

Although the proof given above doesn't strictly apply to the memory-1 update rule, it can be modified to work for any update rule that gives its agents finite memory by enlarging the state space to ordered tuples of network colorings. In Figure 3.5c, we see that finding uniform colorings and 2-colorings are equally difficult on random bipartite graphs.

These simulations confirm that the behavior when searching for a 2-coloring is the same as when searching for a uniform coloring, regardless of the specific update rule.

Section 3.4

Discussion & Conclusion

Studying the collective behavior of individuals in a large group has long been an important research area of statistical physics and relevant fields. The question of "collective action," the tendency for individuals in a group to forgo short-term selfish behavior in favor of long-term group benefit, has been extensively discussed and examined. Of particular interest is classifying the environmental factors that foster cooperation within group, particularly in the case of a public goods game and the Prisoner's Dilemma [142]. There are a plethora of studies that use networks to model a social structure on the group, and the exact topology of networks can have a profound impact on the cooperation inside a group [8, 226, 232, 288]. Additionally, empirical research uses human trials to examine how humans behave rationally (or irrationally) when actually playing public goods games with others [222].

Our results add to the study of collective action by approximating public goods games in that individuals sometimes need to make selfless actions (choosing colors that increase color conflicts) with the long-term goal of increasing success for the entire group (finding a 2-coloring or uniform coloring) [160]. Our present work shows that these two fundamentally different games behave in the same way on random bipartite networks.

Our finding is counter intuitive, but it is important to remember that it applies in a relatively narrow range of scenarios. A bipartite structure is unlikely in most social networks, which means anti-coordination and coordination are equivalent problems only in the small selection of populations that happen to be bipartite with an initial coloring sampled uniformly from all possible colorings. However, these bipartite networks do occur widely in real systems with two different types of individuals like

media producers and consumers [54, 278] or in a sexual contact network (that only considers heterosexual connections) [103]. More generally, there are also no parallels for n-colorings for $n > 2$.

Chapter 4

Polarization, Abstention, and the Median Voter Theorem

Section 4.1

Introduction

When is it rational for two strategically-motivated candidates to deviate from the ideological center in a general election? Spatial models of economic competition have long served as a baseline model for political agendas and electoral outcomes [95, 155]. In their simplest form, every voter's political preference is captured along a one-dimensional space, and each voter chooses the candidate (typically out of two) who is most proximate to them in the one-dimensional ideological space. Accordingly, each candidate rationally selects a point which maximizes their share of votes. The main result is well-known, two competing and self-interested candidates are at equilibrium when their political positions are equal to the opinion of the median voter. As anyone with even the slightest interest in politics will know, there are many examples of politicians in fully-functioning democracies that are not trying to appeal to a median voter, but instead take increasingly outrageous positions to appeal to fringe voter

groups. This chapter attempts to explain this departure from the Median Voter Theorem by focusing on several factors that impact voting but are not considered in the simplifying assumptions that give us the Median Voter Theorem.

4.1.1. Factors that Impact Voting Dynamics

A simple one-dimensional, two-candidate model of elections ignores many complexities which have been addressed individually by political scientists since the appearance of the Median Voter Theorem. There may be more than two candidates, or a third option may threaten to enter the race depending on the ideological alignment of the two main candidates [223]; the ideological space that candidates are competing on may be multi-dimensional [86, 149]; voters may have probabilistic rather than deterministic voting rules, which can shift the point of candidate ideological convergence from the median to the center [27, 189]; candidates may not be purely concerned with winning, and gain more utility from winning with a specific ideological position [169].

Furthermore, many scholars have found that the polarization of politicians and other elites like political journalists, pundits, and party officials leads to the polarization of voters, both in terms of their ideological views [96] and their affective perceptions of the other party [25] and opposing political elites [252]. While these mechanisms of top-down social influence are important to consider, voters do not follow polarized elites unconditionally [198], and we expect that candidates also have an incentive to follow voters as the original Median Voter Theorem states. In our model we present a set of decision-making rules that could generate voter-driven elite polarization, rather than elite-driven voter polarization or voter-driven political convergence. Stated in another way, we show how polarization could emerge from elite preferences for winning, rather than elite commitments to predetermined ideologies. While voter-driven or elite-driven mechanisms may be sufficient to generate polarization, both are almost certainly at play in "the real world," and there is substantial

interplay between the two. Our modeling approach focuses on the voter-driven model, and presents a parameter space of voter behaviors and distributions. We assess which combinations of behaviors and distributions by themselves would be sufficient to generate polarization between strategically-minded candidates.

We focus on a set of three main complications that are present now in the United States, but have not yet been examined together. First, we consider the influence of ideologically motivated third-party candidates. While third-party voting has been on the decline in the United States [150], voting for non-competitive third-party candidates still occurs as an expression of cynicism or distrust with the the larger political system [234], often at levels that sway the results of major elections [9]. Our models will vary in terms of the intensity of third-party appeal to voters, but the political positions of ideologically-motivated candidates will remain fixed at the far ends of the spectrum in our primary model. This is primarily because our model aims to consider centrifugal mechanisms that drive candidates away from the position of the median voter. (Fixing third-party candidates also limits the dimensionality of the parameter space in a way that makes interpretation of findings more straightforward.) In practice, an ideologically motivated third-party candidate may appear between masses of voters on the left and the right, which we suspect may be an increasingly likely scenario in the United States if polarization increases or persists. (We briefly explore the case of an ideologically-driven candidate at the political center in Section 3.2.1, and find that under certain conditions this may both prevent the polarization of candidates or prevent their convergence to the center. Our main results and findings, however, will focus on the case of extreme third-parties.)

Second, elections in the United States typically feature large numbers of eligible voters who stay at home [115]. The reasons for voter abstention are multi-faceted, and draw on a range of perspectives from across the social sciences (for a comprehensive

review, see [274]). Spatial models of voter choice, where the only voter-level attribute considered is the ideological position of the voter, have considered abstention to be a function of a voter's proximity to the candidates [10, 87, 99, 147, 148, 220]. Empirical analysis of voting behavior in United States presidential elections [4, 238, 315] and United States Senate elections [236] support the notion that voters are not motivated to take on the cost of voting when they do not find any candidate appealing, or when they are indifferent between candidates. The idea that voters are more likely to show up in response to a higher perceived material or cognitive payoff from a more ideologically proximate candidate aligns with other cost-benefit based analyses of voting behavior, which find that adverse weather conditions lower voter turnout [129], while same-day voter registration decreases opportunity costs and increases turnout [108]. We also expect that candidates will disproportionately utilize resources to mobilize voters who are most ideologically proximate and likely to support them. Modern political campaigns that feature repeated targeted attempts at voter contact were found to increase voter turnout by up to 7-8 percentage points in targeted areas during the 2012 United States presidential election [100]. Voter turnout also varies as a function of individual characteristics such as sex, race, and age [12], as well as more mutable socio-demographic factors such as income and wealth [200], education [139], and health [40, 58], all of which are conceivably correlated with the relative costs and benefits of voting. Voter habit-forming and socialization [125] and social norms [16] also determine the likelihood of voter turnout. We expect that net of these social and psychological factors, the ideological positioning of candidates still shapes the motives and abstention level of voters. Furthermore we expect that vote-optimizing candidates will adjust their positions to reflect the ideological distribution of those who are expected to vote, not the ideological distribution of the entire-voting eligible population.

The combined threats of voter abstention and ideologically extreme candidates generates tension between winning over the center or appealing to the 'base' when determining what candidates or platforms to field for a general election [3]. On one hand extreme candidates might cede the center to the opponent (in line with the assumptions of the Median Voter Theorem), and on the other hand extremist voters may behave irrationally and stay at home rather than casting a vote for the candidate who is closest to them ideologically. An additional concern is that extremist candidates, while energizing their own base, may increase turnout among people who are extremely opposed to their agenda as well [137].

The final mechanism we consider in our model is *polarization*, where our one-dimensional population becomes more dispersed as voters become more extreme. Polarization has been examined extensively by political scientists [111], sociologists [23], and economists [92], and its empirical scope and potential causes have been the focus of impressive studies by information scientists [80] and computational social scientists [22], but it's implications for rational choice voting models and candidate competition are rarely considered [134, 307]. While there is some debate about whether the political divide in the United States is on the rise [23, 112], by many metrics polarization among American voters has increased over the last few decades [2, 45, 166], and the ideological differences between political elites has also grown [240, 241, 280]. Many causes of polarization have been identified, including homophily [119, 190], in-group identifiers [90, 91, 128], social media [29, 173, 283], and even media consumption [61, 245].

The median voter model can be seen as a "bottom-up" process that brings the political preferences of rational candidates in line with the more centrist preferences of the electorate. It is a model of anti-polarization [134], but its limitations have become apparent in the current political climate.

Given these three variables: voter tendencies towards third-party candidates, staying home, and polarized beliefs, we are primarily interested in whether specific combinations will motivate strategic candidates to pursue divergent ideological strategies. Given the growing polarization in the U.S. electorate [309], it is important to consider the conditions necessary for candidates to follow voters in their drift to extreme positions in the short term. While the use of formal modeling cannot discern the fundamental causality of any specific example of contemporary polarization in the United States or elsewhere, it can shed light on what combinations of voting behaviors could incentivize the drift of candidates away from the center. In practice, we believe this is a more complex phenomenon, likely explained by some combination of intersecting political, cultural, economic, and technological factors. However, we believe that reconciling this idea with one of the most straightforward, influential, and highly-cited models of candidate behavior (the spatial choice model of Hotelling (1929) and Downs (1957)) can be generative for future theoretical and empirical research on the intersection of elections and polarization.

Our approach builds on more parsimonious two-candidate models of voter choice, wherein we allow voters to either choose one of the two strategically-motivated main candidates, ideologically motivated third-party candidates, or to stay home altogether. We also consider the ideological distribution of the voter electorate as a proxy for political polarization. Following earlier advances in the voter choice literature, our approach treats voting as a stochastic rather than deterministic process [82]; the odds of a voter choosing a candidate increase with their relative ideological proximity but is almost never a certainty. This methodological decision is thought to lead to better models of voter uncertainty [57]. A stochastic voting model has shown that preferential skew does lead to non-median outcomes [79, 81, 146], but these models still have one unique equilibrium.

In our analysis, we systematically vary the ideological distribution of voters, the appeal of ideologically-motivated third-party candidates on the far ends of the political spectrum, and the appeal of staying home all together. Finally, we map the conditions under which rational political candidates fail to converge on the median ideological position, and also when candidates become more extreme than the electorate itself. We then analyze these dynamics with two empirically observed voter opinion distributions from the contemporary United States.

This chapter was published in [161].

Section 4.2

Methods and Model

4.2.1. A Model of Voter Selection and Population Polarization

Our model examines how a polarized population can influence the political positions of two strategically motivated candidates, who are purely interested in maximizing vote share. Building and integrating the spatial models of voter choice from Hotelling [155], Black [38,39], Poole and Rosenthal [237–241], we allow for the possibility that a voter may either select a ideologically motivated and extreme candidate instead of a major-party candidate, or that voters may vote for neither candidate if they find their choices unappealing. Third-party candidates in our model are ideologically fixed on either end of the spectrum, as it is presumed that they are motivated by representing a specific position at the far end of the political spectrum, rather than maximizing vote share. Unlike the two rational candidates, they do not shift their position in the model.

Our model considers how both the ideological distribution of the voters and voter tendencies to select one of the two major candidates should influence the political positioning of the two main candidates. These patterns change even when the median

and mean voter position is fixed at the center of the distribution. Before discussing the results of our approach, we first discuss the two main components of the model: the distribution of voter ideology and the function that is used to map voter ideology to voter choice and behavior.

4.2.2. Ideological Distribution of Voters

We assume a one-dimensional ideological distribution of voters, x, on a scale from 0 (left) to 1 (right). We assume that the population of voters is made up of two Gaussian sub-populations, consolidating around two "peaks" equidistant from the ideological center (0.5). The distance between the peaks is determined by α, and the variance in position around these two peaks is determined by σ^2.

The corresponding population density function, denoted $f(x)$, is the sum of two Gaussians centered at $0.5 \pm \alpha/2$:

$$f(x) = \frac{c}{\sigma\sqrt{2\pi}}\left[\exp\left(-\frac{(x-0.5-\alpha/2)^2}{2\sigma^2}\right) + \exp\left(-\frac{(x-0.5+\alpha/2)^2}{2\sigma^2}\right)\right] \quad (4.1)$$

where c is a normalizing constant to ensure that $\int_0^1 f(x)dx = 1$. The population is symmetric around $x = 0.5$, with the median voter always located at 0.5. Figure 4.1 illustrates this distribution. While we focus on our model on a hypothetical case in which there are two balanced left-leaning and right-leaning subpopulations, the ideological distribution of an actual population, which is not necessarily symmetric, can be determined using real voter data from any population of interest [56].

4.2.3. Voter Choice Function

When voters always select the most ideologically proximate candidate, both parties would still converge to the opinion of the median voter, which is fixed at 0.5. Varying the distribution of voters would have no effect on the strategic ideological positions

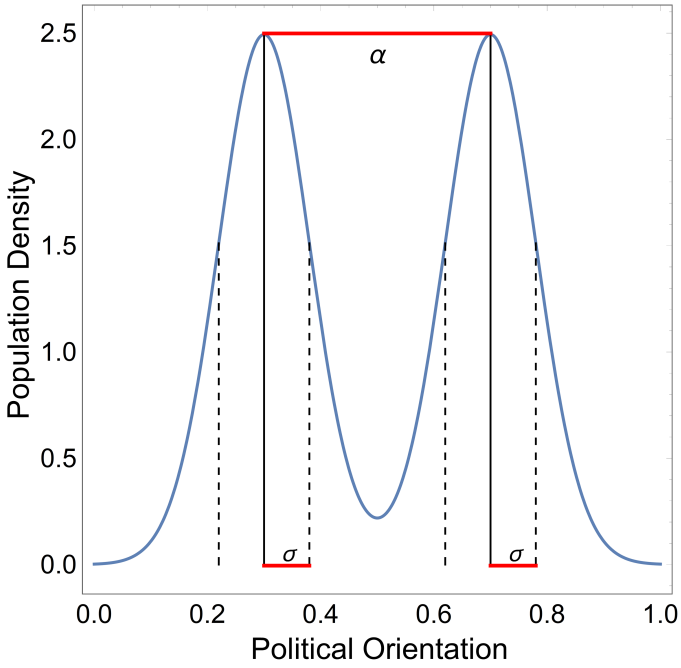

Figure 4.1: Ideological distribution of voters as a function of the two population parameters, α and σ. α is the distance between the two subpopulation centers, and σ is the variance around these subpopulation centers. As σ increases, the population distributions will become less pronounced and more diffuse.

of the candidates.

Yet in reality political candidates may have concerns about losing their "base" when trying to appeal to the "center". In the one-dimensional spatial model, the threat of losing the base only occurs when voters have the option of either abstaining or selecting a third-party candidate that adopts a position in accordance with their ideology as opposed to vote maximization. Conversely, voters in the center may abstain if both candidates assume positions that are too extreme for them.

In our model, there are three variables that control voting behavior: pragmatism (P), which can be thought of as the appeal of voting for a two-party candidate, relative cost of voting (Q) which adjusts the voter tendency for staying home, and rebelliousness (R), which determines the appeal of third-party candidates. P and R are similar, and balance the candidate's preferences towards an ideologically motivated third-party selection or a more practical two-party selection. When voters are more ideologically equidistant from candidates, they should be more likely to stay home altogether. Q is a multiplier for this, such that the utility a voter receives from not voting is a product of Q and the voter's ideological indifference between the two candidates.

The behavior of the voter is determined by behavioral utilities calculated from the three above parameters, the ideological position of both of the major parties, and the ideological position of the voter in question.

For an individual at v and major candidates at b and $r \in [0, 1]$, we create the following utilities:

$$\text{Vote Blue Utility} = u_B(b, v) = \frac{1}{|b - v|^P} \tag{4.2}$$

$$\text{Vote Red Utility} = u_R(r, v) = \frac{1}{|r - v|^P} \tag{4.3}$$

$$\text{Abstention Utility} = u_A(b, r, v) = (1 - |(|b - v| - |r - v|)|)Q \tag{4.4}$$

$$\text{Vote Third Party Utility} = u_T(v) = \frac{1}{(1-v)^R} + \frac{1}{v^R} \tag{4.5}$$

Each voter chooses from one of the four possible behaviors (vote for red, vote for blue, vote for third party, and abstain) with a probability proportional to each of their respective utilities. Figures 4.2 and 4.3 provide visual depictions of how voter behavior varies in the model as functions of voter and candidate ideology, respectively.

4.2.4. Voter Choice Dynamics

Figure 4.2 shows voter utilities and corresponding probabilities for a set of parameters. The candidates have ideological positions of 0.3 and 0.7, somewhere between being completely polarized and converging to the middle, which roughly reflects two-party elections in the contemporary United States. To illustrate the model, we select a set of parameters for the proposed voter utility functions that lead to an intuitively plausible relationship between voter ideology and voter behavior. The values $P = 2$, $Q = 30$, and $R = 1$ cause more "extreme" voters with an ideology closer to 0 or 1 to be more likely to select a third party candidate or stay home. Furthermore, the voters in the ideological valley between the two candidates are more likely to stay home, as they do not gain much of a relative benefit from either candidate. Obviously, different parameter values will result in different voter behavior. These values were chosen because they do a decent job of approximating real voters' behavior and have a strong degree of randomness when not closely aligned with one of the main candidates.

Figure 4.3 also uses this set of "common sense" decision parameters, but instead focused on the decision behavior of a single voter at a fixed ideological point, and examines how voter behavior corresponds to the ideological positions of the two main candidates. For a voter with an ideology of 0.5, a "median voter," they become more likely to choose a blue or red candidate when one of them adopts a platform that is ideologically moderate. They become more likely to abstain when both candidates

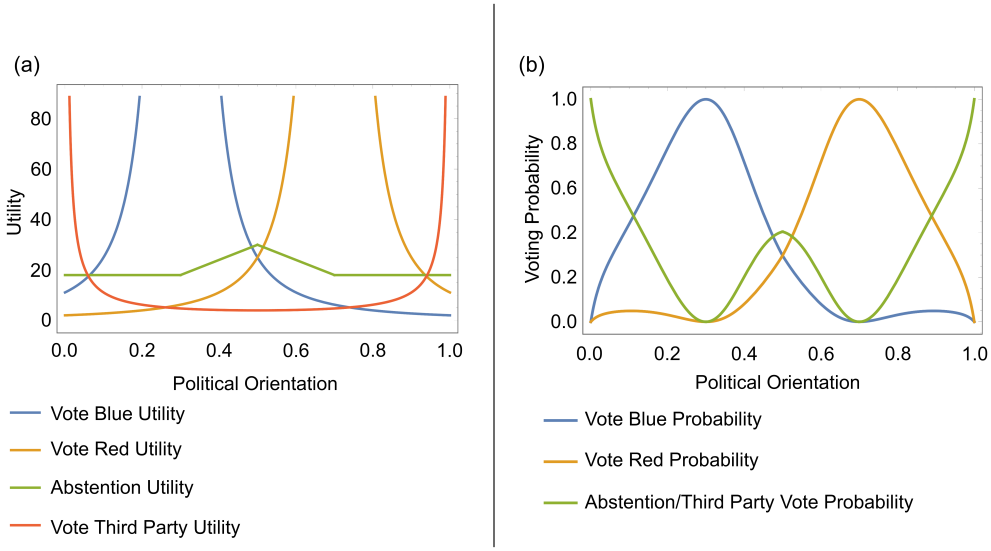

Figure 4.2: Voter utility and voting behavior of individuals across the entire political spectrum with two fixed political candidates. (a) shows the utility that a voter receives from different actions as a function of their position on the political spectrum, assuming candidate positions of 0.3 and 0.7 and a specific set of model parameters ($P = 2$, $Q = 30$, $R = 1$). (b) maps these utilities into one of three behaviors: voting for the "blue" (left-leaning) candidate, the "red" (right-leaning) candidate, or voting for neither (staying home or selecting an ideologically motivated third-party candidate).

choose more extreme candidate positions on either the same or opposing sides of the political spectrum.

4.2.5. Candidate Optimization

The behavior of each voter is stochastically determined as a function of their ideological position, the positions of the candidates, and the parameters of our model according to the following equations. We can determine the optimal ideological positions for two competing candidates who are motivated by maximizing vote share. For an ideological space that stretches from 0 (on the left) to 1 (on the right), the liberal and conservative candidates are each seeking an ideological position (called b

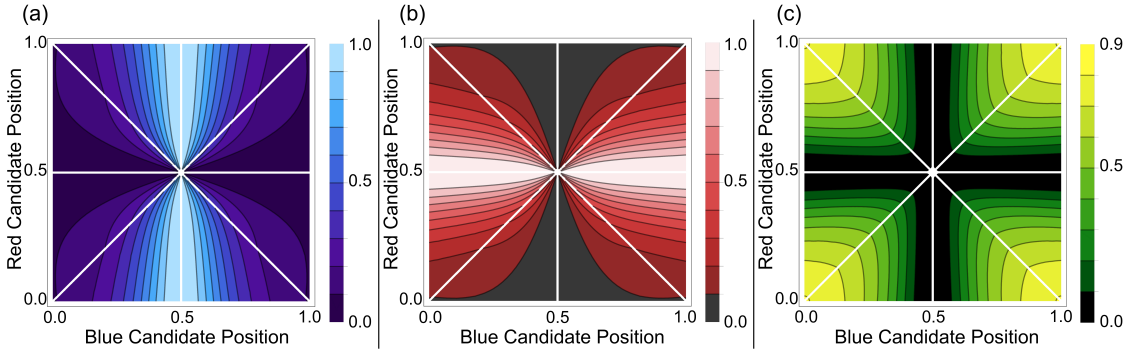

Figure 4.3: Varying the candidates' positions influences a fixed voter's behavior. Each panel shows how the likelihood of a given voter behavior (voting blue in (a), voting red in (b), or voting for neither in (c)) changes as a function of the two political candidates stated ideological position, $[0,1]^2$, assuming a voter ideology of 0.5, a specific set of model parameters ($P = 2, Q = 30, R = 1$). (The white lines crossing the space and point at the center reflect one-dimensional portions of the two-dimensional space where a candidate has the exact same preference as a voter ($r = 0.5$ or $b = 0.5$), where the two candidates are identical (r=b), or where the voter is equidistant between the two candidates ($|r - 0.5| = |b - 0.5|$). In these subsets of the space, the model loses practical interpretability for this single hypothetical voter.

and r respectively) that maximizes the value of one of the following integrals:

$$\text{Blue Votes} = v_B(b,r) = \int_0^1 f(v) \frac{u_B(b,v)}{u_B(b,v) + u_R(r,v) + u_A(b,r,v) + u_T(v)} dv \tag{4.6}$$

$$\text{Red Votes} = v_R(b,r) = \int_0^1 f(v) \frac{u_R(r,v)}{u_B(b,v) + u_R(r,v) + u_A(b,r,v) + u_T(v)} dv \tag{4.7}$$

Our model of the two major candidates' theoretical optimization process in response to voters' behavior can be described in terms of the so-called adaptive dynamics [152, 312]:

$$\begin{aligned} \frac{db}{dt} &= \frac{\partial v_B(b,r)}{\partial b}, \\ \frac{dr}{dt} &= \frac{\partial v_R(b,r)}{\partial r}. \end{aligned} \tag{4.8}$$

When voters choose the most ideologically proximate of the two competing candidates, both positions converge on the ideology of the median voter. Our model shows how this result does not necessarily hold when voters might choose to abstain or select a third party. In particular this can occur when the distribution of voter preferences is sufficiently bimodal with a large α and small σ. Figure 4.4 shows three different sample voter ideological distributions (d-e), and how two political candidates will adjust their ideological platform under a reasonable set of voter choice parameters for each (a-c).

With these three populations, candidate behavior varies from appealing to the median voter when competition is fierce in the high-density middle to being more polarized than the population as candidates work to protect their most extreme voters from a third party challenge.

Candidate Optimization Against Ideological Centrism. While our base model works under the assumption that ideologically-motivated candidates come from the far ends of the political spectrum, it is also possible that an ideologically fixed third-party could run from the center. We examine the effects this has on candidate optimization under the parameter setting $P = 5$, $Q = 0$, and $R = 5$ in Figure 4.5. Unsurprisingly, major party candidates in a bimodal population are pulled closer to the center than under conditions of extreme third-party candidates. However, in a unimodal population, major-party candidates are pushed towards the fringes by the centrist third-party, as they benefit from distinguishing themselves from the centrist. This suggests that under a certain set of voter behaviors, having an ideologically-fixed third party candidate may be a solution to both the problems of runaway extremism of polarization and the stasis of convergence to the center. While the remainder of our analysis continues to focus on the case of ideologically extreme third-parties, this insight may present a promising avenue for future empirical and theoretical research.

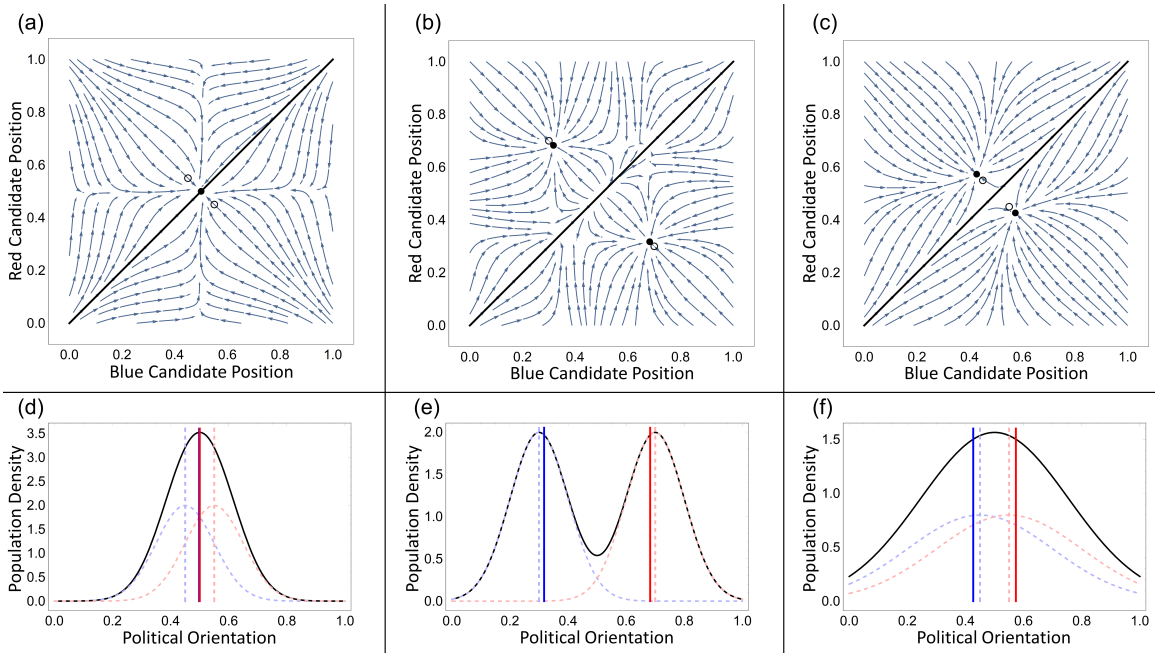

Figure 4.4: Optimal position of two competing major candidates. Using stream plots, subfigures (a)-(c) show how candidates will shift their position, with the black dot representing the candidates' equilibria, and the circles showing the subpopulation peaks. Subfigures (d)-(f) show the corresponding populations in black, the two subpopulations with dashed curves, the subpopulation centers represented by dashed vertical lines, and the candidate equilibrium positions represented by solid vertical lines. We see three types of behavior: candidates converging to the median voter (a,d), candidates less polarized than the population (b,e), and candidates more polarized than the population (c,f). All plots use a reasonable set of parameters $P = 2$, $Q = 30$, and $R = 1$.

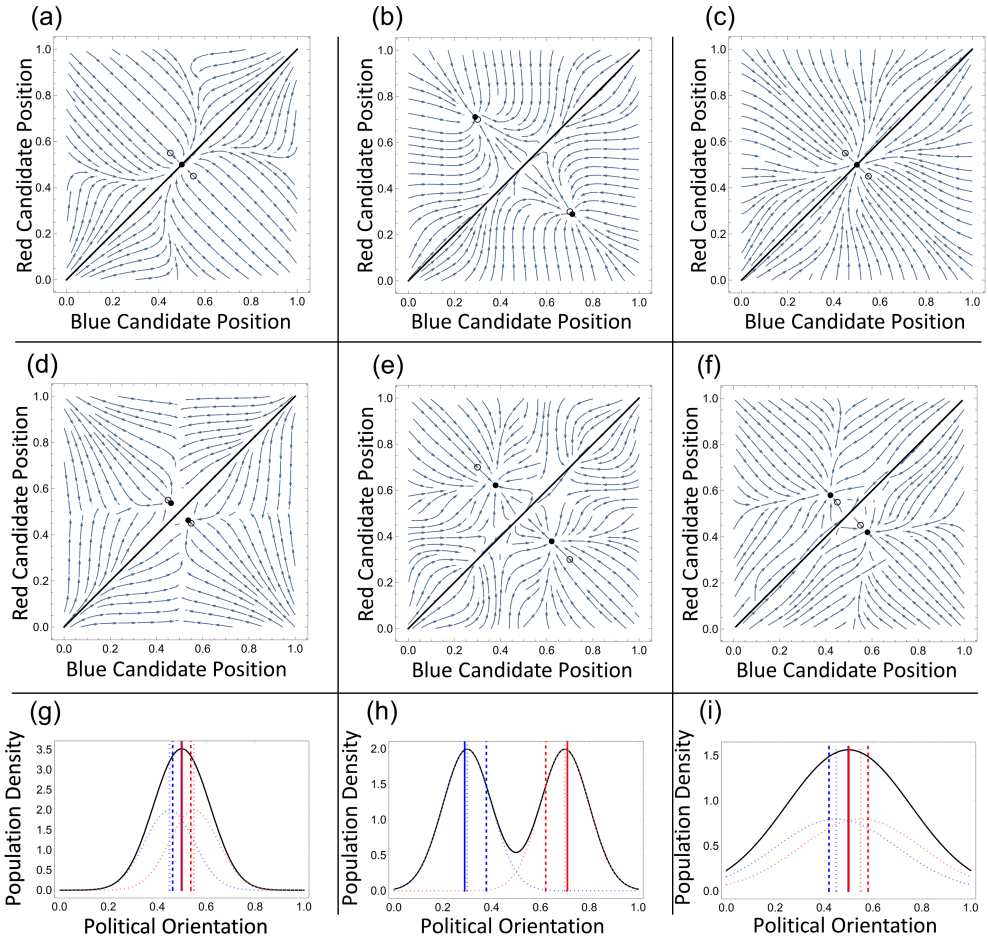

Figure 4.5: Optimal positions of two competing major candidates in conditions of two extreme third-parties (a)-(c) and one centrist third-party (d)-(f), with the black dot representing the candidates' equilibria, and the circles showing the subpopulation peaks. Subfigures (g)-(i) show the corresponding populations in black, the two subpopulations with dashed curves, the subpopulation centers represented by lightly-colored and dashed vertical lines, the candidate equilibrium positions in the extreme condition represented by darkly-colored and solid vertical lines, and candidate equilibrium positions in the centrist condition represented by darkly-colored and dashed vertical lines. When the population is bimodal (h), centrist parties "pull-in" main candidates from the ideological peaks (b) instead of matching the ideological peaks of the voters (e). However, when the population is more unimodal (g,i), centrist third-party candidates either prevent main parties from completely converging to the center (d) or push them away from the center (f), compared to extreme third-party candidates (a,c) who cause the main-parties to converge to the center. All plots use the set of parameters $P = 5$, $Q = 0$, and $R = 5$.

Section 4.3

Results

Depending on voter predisposition to extremist third party candidates, or their willingness to simply stay home in the absence of an appealing candidate, the rational positions taken by main candidates will vary. In our model, candidates do one of three things. They either (1) converge to the median, (2) deviate from the median but still select positions between the two peaks of public opinion, or (3) deviate from the median to a greater extent than the voting base. Two examples of how voter ideology distribution shapes candidate positions is shown in Figure 4.6. For each selected set of sample model parameters, each of the three candidate outcomes are possible.

While possibility (2) is interesting primarily because of its deviation from the results typically derived by the Median Voter Theorem, possibility (3) reveals a potential long-term mechanism for voter polarization. While our model assumes that voter preferences are static and the position of strategic candidates are dynamic, other models have considered the possibility that voter positions eventually come to resemble candidate positions [169]. If voter behavior and ideological distribution is one that motivates extremism among rational candidates, this may in turn create a larger spread among voters.

The model parameters have a strong influence on how the ideological structure shapes the strategic polarization of the two main candidates. In Figure 4.7, we see how changing P changes candidate position relative to ideological distribution of the population. For smaller values of P, voters are less likely to "settle" for a major candidate, and accordingly the candidates drift from one another to capture more extremal voters, although this split does not exceed the bimodal "peaks" in the

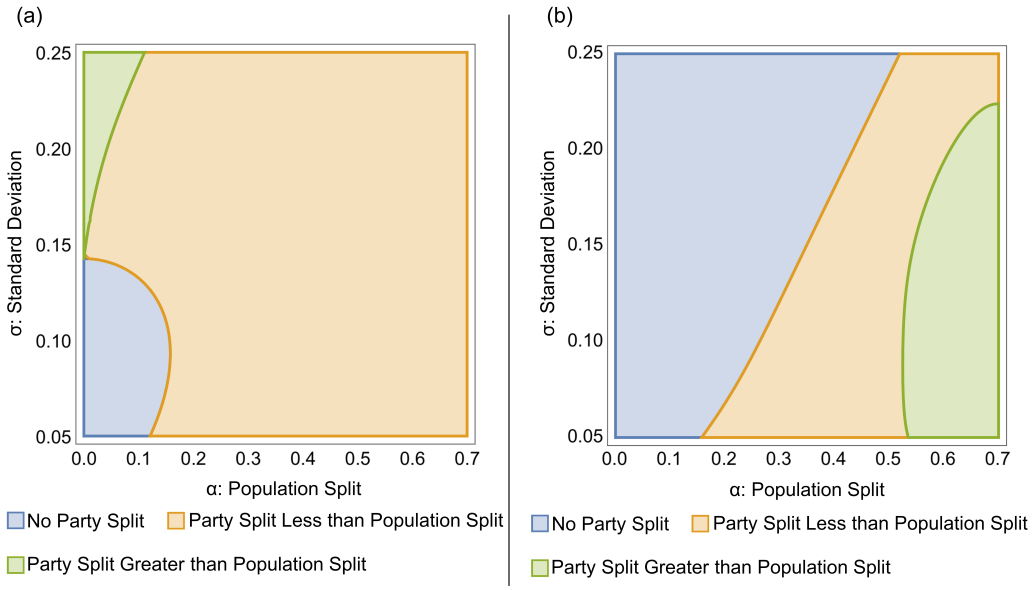

Figure 4.6: For two sets of model parameters, the type of equilibrium candidate positions is shown as a function of the distribution of voter ideology. In both plots, the x-axis is population split and the y-axis is the standard deviation of the two sub-populations. (a) uses parameters $P = 2$, $Q = 30$, $R = 1$, and (b) uses $P = 5$, $Q = 0$, and $R = 5$. The three regimes of interest are candidates converging to the same position (roughly the mean/Median Voter Theorem result), separating to a lesser extent than the population (the space between the two peaks), or separating to a greater extent than the population. Each space is shaded by adherence to one of these three regimes.

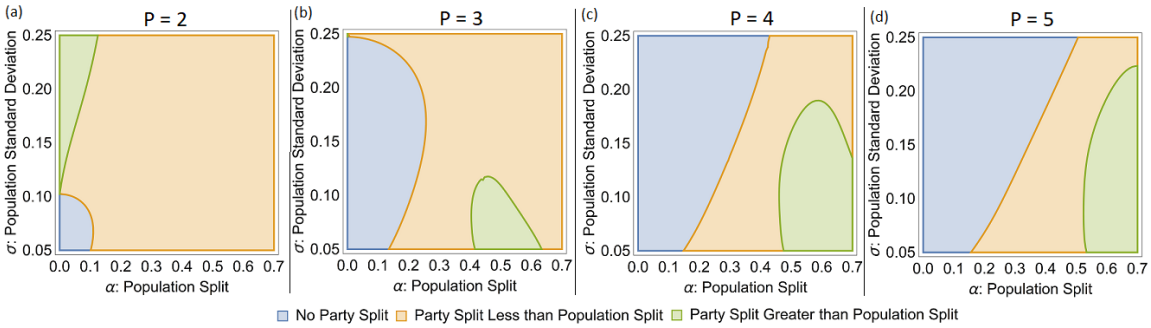

Figure 4.7: Sets of model parameters where $Q = 30$ and $R = 5$, and P takes on values from 2-5. As P increases, the voter choice model gives more weight to "pragmatic" main party opinions, and the mapping between ideological distribution and the relative political position of the two main parties changes. When P is smaller, political parties diverge under most ideological distributions, but only more than the public itself when there is large variance around two otherwise close ideological centers. When P is larger, parties converge to the middle when the population split is small, and are more polarized than the electorate when the population split is large, but the variance around the two ideological centers is small.

underlying population, unless variance around the peaks is large relative to variance between the peaks. Larger values of P mean that voters are more pragmatic and tend to favor voting for the two major parties. They are less likely to see the parties diverge from one another, but when distance between the ideological peaks is large, the parties may be on the extreme sides of the two population centers, and candidates then need to capture vote-share from extremist third parties.

In Figure 4.8, we explore the full range of values for σ and α, and both large and small values of P, Q, and R. This is shown in a series of eight two-dimensional plots, colored by whether the two strategic parties converge to the center, split but remain more centrist than the population centers, or split in a way that exceeds the ideological divergence of the population centers.

This visualization lends itself to broader insights. The two parties are more likely to converge when the ideological centers of the population structures are closer together (smaller α), although this is not the case when the cost of voting or the appeal of third parties is large. In general, when the appeal of third-party candidates is small

and the cost of voting is small (meaning most voters show up to the polls), assumptions that align more closely with earlier spatial voting models, convergence to the center is ubiquitous. Under more unimodal ideological distributions (with large ratios of spread around ideological centers compared to the split between the centers) the candidates converge to the center, and with more bimodal ideological distributions (small ratios of spread to split) the candidates still become more moderate than the population centers. The latter type of result deviates from the traditional median voter result because voters do not choose their most proximate candidate with certainty. Similar results hold when the cost of voting is large, but mainstream parties have stronger appeal.

Another outcome of interest is when the two parties diverge to a greater extent than the underlying population. This occurs under two combinations of conditions: (a) when the ideological distance between population centers (α) is small, the spread of voters around the population centers (σ) is large, and the cost of voting (Q) and/or third-party appeal (R) is large, or (b) when the distance between the population centers (α) is large, the spread around the population centers (σ) is small, the third-party appeal (R) is large, and the appeal of the main parties (P) is small. Condition (a) corresponds to a situation where there is a nearly unimodal but somewhat ideological diverse electorate that is reluctant to vote or drawn to third-parties, main parties must contend with apathy and third-party appeal by moving away from the center and distinguishing themselves from one another. Condition (2) corresponds to a situation where a highly and uniformly polarized electorate is drawn towards third-parties, where main parties will win few converts from the other side, and instead try to mobilize the extreme wings of their corresponding side of the ideological voter distribution.

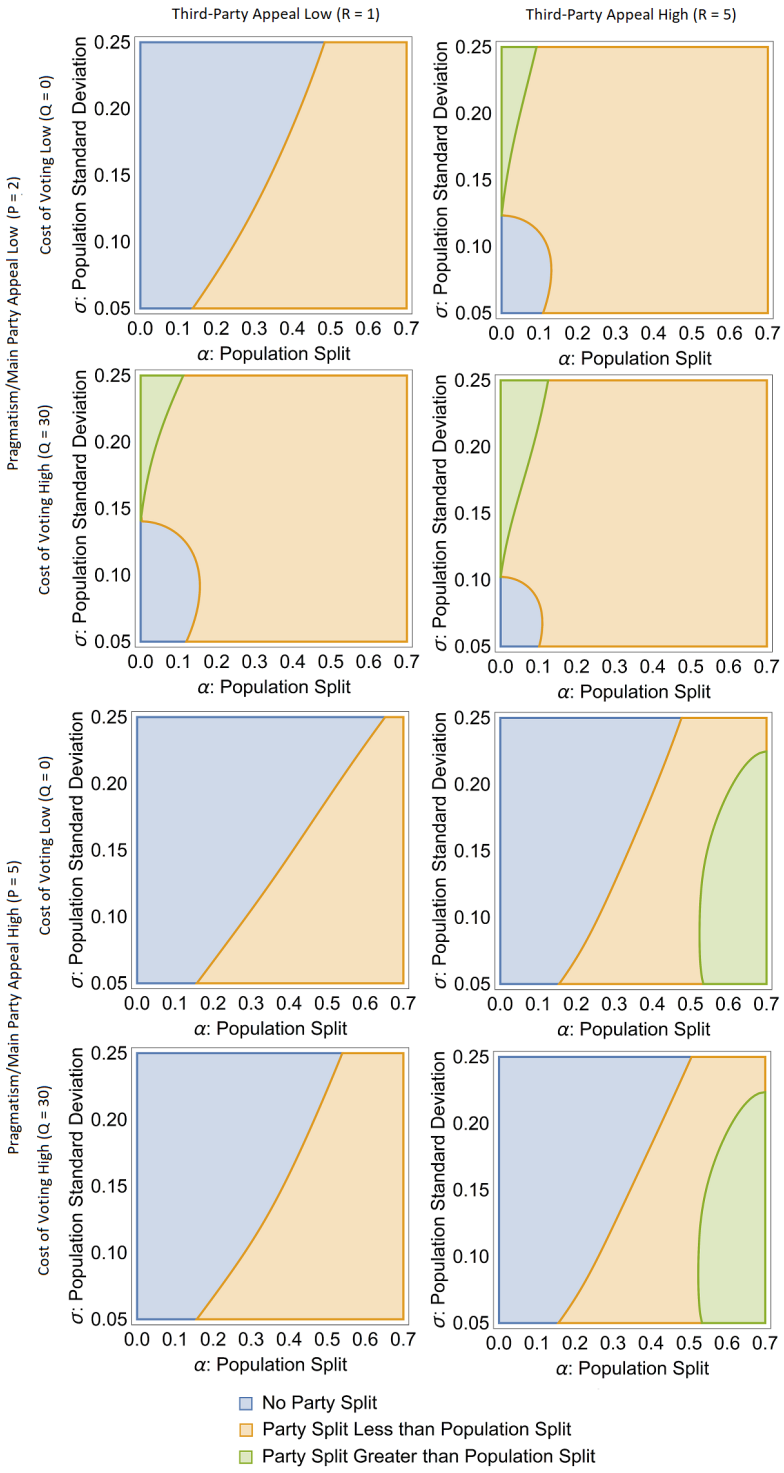

Figure 4.8: The positioning of political parties relative to ideological population centers for different parameters of the voter choice model (P, Q, R) and different population ideological distributions (σ, α).

4.3.1. Incorporating Empirically Observed Ideological Distributions

The results of our model show that under the very basic assumptions of voters being attracted to third party candidates or being prone to staying home, it may make sense for candidates to avoid the center depending on the distribution of voter ideology. We can incorporate observed empirical distributions of voter opinions into a set of model parameters ($P = 2$, $Q = 30$, and $R = 1$) to examine how this model of voter choice might function under contemporary ideological distributions in the United States. Our empirical voter distributions come from two sources. In Figure 4.9a, we see the first data set from [69]. As we can see, the population here is neither symmetric nor bimodal. However, there is still enough spread in the distribution of the voters to generate a separation between two candidates. The true median of the population ideology is roughly 0.42, but candidates converge to positions at about 0.25 and 0.51. Perhaps unsurprisingly, the asymmetric distribution of voter preferences leads to differing distances between the median position and each of the candidates.

Data from the estimated ideological positions of Twitter users provides a more polarized empirical distribution for examination, and was taken from Figure 3 of [69] by means of redigitialization. In Figure 4.9c we can see that there is a more roughly bimodal distribution, although it remains asymmetric. While we unfortunately cannot derive values of model parameters P, Q, and R without data on how individual opinions map into votes, we can look at how candidates should behave strategically under the observed empirical distributions and the previously used parameters of $P = 2$, $Q = 30$, and $R = 1$. Once again, political candidates converge on positions that deviate from the median of roughly 0.57, with the left-leaning candidate selecting a position all the way at 0.2, and the right-leaning candidate selecting a position at 0.65. Curiously, while there are more voters on the right than on the left here, the far left positions of the left half of the distribution bring the left-leaning candidate very

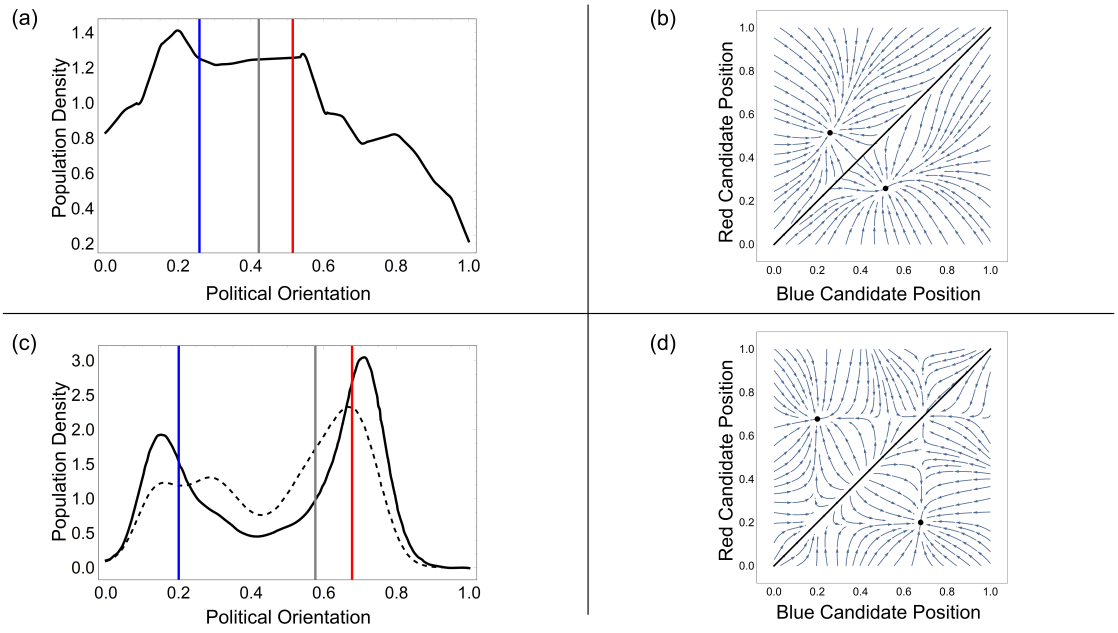

Figure 4.9: Candidate optimizations based on real-world ideological distributions. First, we show the results of testing our model on real-world data. First, we used the distribution of voter ideology according to the Summer 2017 Political Landscape Survey [69] (a), and the rational candidate responses to this landscape (b). We also used an analysis of Twitter users [197] (c), and the rational candidate responses to this landscape (d). In (a) and (c), vertical lines show the convergent position of each candidate (blue and red) and the position of the median voter (grey). (c) also shows the political ideology of political leaders on twitter in dashed grey, and we see that the vertical blue and red lines match nicely with peaks for this curve. Both models used $P = 2$, $Q = 30$, and $R = 1$, a set of parameters that approximate the credence that an average voter may give to not voting or voting for a third party. The population in (c) can be fit to a bimodal distribution. The best fit has the left and right subpopulations' peaks at 0.18 and 0.70, standard deviations at 0.09 and 0.07, and relative weights at 1 and 1.32, respectively. In (a), the population is less bimodal, and so the fit has less value. However, for completeness, we give the values here: positions 0.18 and 0.37, standard deviations 0.04 and 0.41, and relative weights of 1 and 0.81.

far from the median. If we suppose that in equilibrium the left choice would win half of the time and the right choice would win half the time, the average position of the winning candidate would be roughly 0.38, very far to the left of the median of 0.57. In this case, the willingness of voters to abstain or vote "irrationally" for third party candidates gives more weight to the side that entertains more extreme positions.

4.3.2. An Analytically Tractable Model of Voter Behavior

So far, our model considers a scenario where individuals decide between voting for a major party, staying home, or voting for a third party. Voters make their choices probabilistically rather than deterministically, with the ideological distance between voters and candidates impacting the weights of behavioral probabilities. This probabilistic decision process together with an abundance of choices make for a main model that is a good approximation of the thought process of the average voter, but is difficult to analyze mathematically.

Here, we examine the specific set of cases when $Q = 0$, and $P = R = \ell$ under the limit $\ell \to \infty$. This creates a simplified model of voting where each voter deterministically selects the candidate that is most proximate to them. In this case, everyone votes, and there is no bias in favor of strategic major party candidates at the expense of ideologically motivated third party candidates. This simplified model would most accurately reflect a population where the costs of voting are effectively zero, and voters are solely motivated by their ideological similarity to candidates.

Our simplified model presented here lends itself to more tractable functions for the total votes for a candidate, and with a few approximations, allows for closed form solutions. Once again, parties can attempt to maximize votes by making incremental changes to their platform. Now, however, the threat of third parties fixed at both ends of the ideological spectrum are greatly increased, and candidates have a much larger incentive to take polarized positions to motivate their more extreme bases. When the

ideological separation between the two centers of political opinion increases beyond a certain point, candidates begin to move away from the median. This is what we refer to as *the first phase shift*. When ideological separation reaches the *the second phase shift*, candidates begin to take on positions that are more extreme than the subpopulations' two ideological centers.

Voter Choice Functions and Candidate Equilibrium. In this deterministic model of voting, the share of votes each candidate gets is a simple integral of the population density function. A voter at v will voter for the closest of 0, 1, b, or r. Therefore, every voter between $\frac{b}{2}$ and $\frac{b+r}{2}$ votes for blue, and every voter between $\frac{b+r}{2}$ and $\frac{r+1}{2}$ votes for red. Up to a constant factor, these quantities can be expressed as integrals of f.

$$\text{Blue votes} = \int_{\frac{b}{2}}^{\frac{b+r}{2}} f(x)dx \tag{4.9}$$

$$\text{Red votes} = \int_{\frac{b+r}{2}}^{\frac{r+1}{2}} f(x)dx \tag{4.10}$$

With these simplified equations, we can find a necessary condition for admissible equilibrium. Consider the blue party, whose share of votes is given by (4.9). The share of votes is dependent on the party platform b, so the party can consider adjusting the platform to increase the number of votes according to adaptive dynamics, as shown in Equation (4.8). Blue party votes can be maximized by setting the derivative (4.9) to be zero, which leads to:

$$f\left(\frac{b+r}{2}\right) = f\left(\frac{b}{2}\right) \tag{4.11}$$

When this condition is satisfied, the blue party cannot further increase votes by changing their platform slightly. This equation can be used to find the admissible

evolutionary stable strategy (ESS) for blue party candidate positioning.

Similarly, the necessary condition for red party votes to be maximized is

$$f\left(\frac{b+r}{2}\right) = f\left(\frac{r+1}{2}\right) \tag{4.12}$$

Therefore, any given pair b and r that simultaneously ensures an admissible local maximum (ESS) of blue and red votes, respectively, should satisify the condition:

$$f\left(\frac{b}{2}\right) = f\left(\frac{b+r}{2}\right) = f\left(\frac{r+1}{2}\right) \tag{4.13}$$

If (4.13) is not satisfied, at least one party can shift their platform to increase the number of their votes.

Multiple equilibrium points can exist, both stable and unstable, depending on the ideological distribution of the population, but for the symmetric, bimodal populations we study in this chapter, there is a single stable equilibrium satisfying $0.5 - \frac{b}{2} = \frac{r+1}{2} - 0.5$, that is, we have $b + r = 1$ (this is because the population distribution $f(x)$ is symmetric with respect to $x = 0.5$). Just like in the stochastic model, depending on the split and variance of the population, the candidate equilibrium can approach the same position at median voter, have distinct positions that are bounded by the ideological peaks of the population, or have distinct positions that are more extreme than the ideological peaks of the population.

Analytical Results for Phase Changes. One result of central interest is when it is strategic for candidates to select positions that diverge from that of the median voter, and furthermore, when it benefits candidates to select ideological positions that are more distant than the two ideological peaks of the proposed bimodal population. We now focus on mathematically identifying the two phase changes between these

three possible qualitative outcomes.

First, observe that the behavior of the candidates is entirely determined by the shape of the population density function $f(x)$, which has two parameters, α and σ. Here we fix σ to be a constant, and consider how changing α affects the equilibrium positions of the candidates. The change in population shape for many values of α is shown in Figure 4.10.

When $\alpha = 0$, the population is unimodal, and both candidates will unsurprisingly converge on the median, where the density is highest. As α increases, eventually the population density at the median will be surpassed by the population density at $b = 0.25$ and $r = 0.75$, the points halfway between the median and the ideological location of the two third parties (shown in Figure 4.10c). At this point, appealing to the median voter at the expense of appealing to extreme voters is no longer optimal. Substituting to Equation (4.13), this condition can be written as

$$f(0.5) = f(0.25) \tag{4.14}$$

The second phase change will occur at the point where the population centers are equal to the optimal positions for the two opposing candidates, when $b = 0.5 - \alpha/2$ and $r = 0.5 + \alpha/2$ (Figure 4.10e). To determine when this scenario is an equilibrium, we substitute this into Equation (4.13), and observe that the second phase change occurs when

$$f(0.5) = f\left(\frac{0.5 - \frac{\alpha}{2}}{2}\right) \tag{4.15}$$

In the remainder of this section, we derive the phase changes as a function of both α and σ.

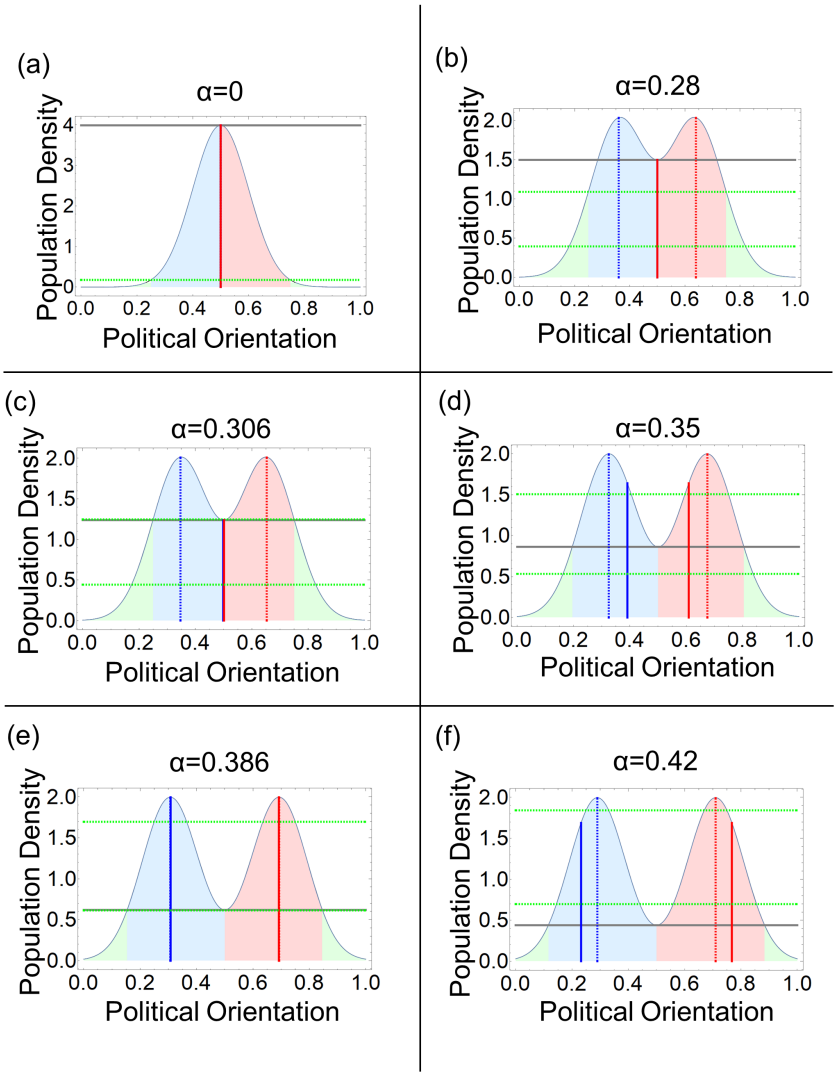

Figure 4.10: Optimal positions of candidates change as the population becomes more ideologically polarizes (as α increases). The red area reflects vote share for the red candidate, the blue area reflects vote share for the blue candidate, and the green area reflects votes for either third party. The solid blue and red vertical lines indicate the positions taken by the political parties, and the dashed red and blue lines indicate the ideological peaks of the voting population. In (a) and (b), the median has a high voter density and candidates compete for the median voter. In (c) the first phase change is shown: α has grown so that a candidate at 0.5 can take an infinitesimally small step away from the median and not lose any votes. The loss of the median voter is exactly offset by the votes gained at 0.25 or 0.75. In (d), candidates are polarized but not as much as the population. In (e), the second phase change has been reached, and the optimal candidate position is the same as the population centers. In (f), the population spread is wide enough that the candidates' optimal positions are outside the ideological modes of the population. Model parameters: $\sigma = 0.1$.

Threshold of Population Split α for First Phase Change. We now turn to identifying phase changes with regards to both α and σ. In order to solve Equation (4.14) for α in terms of fixed σ, we rewrite Equation (4.1) as

$$f(x) = c\left[g\left(x - 0.5 - \frac{\alpha}{2}\right) + g\left(x - 0.5 + \frac{\alpha}{2}\right)\right] \tag{4.16}$$

where $g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$ is the probability density function for the standard normal distribution with variance σ^2.

At $x = 0.5$, we have

$$f(0.5) = c\frac{2}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\frac{\alpha}{2})^2}{2\sigma^2}\right) \tag{4.17}$$

At $x = 0.25$, we have

$$f(0.25) = \frac{c}{\sigma\sqrt{2\pi}}[\exp\left(-\frac{\left(-0.25 - \frac{\alpha}{2}\right)^2}{2\sigma^2}\right) + \exp\left(-\frac{\left(-0.25 + \frac{\alpha}{2}\right)^2}{2\sigma^2}\right)] \tag{4.18}$$

Therefore, the threshold of population split α for the first phase change satisfies

$$2\exp\left(-\frac{\left(\frac{\alpha}{2}\right)^2}{2\sigma^2}\right) = \exp\left(-\frac{\left(0.25 + \frac{\alpha}{2}\right)^2}{2\sigma^2}\right) + \exp\left(-\frac{\left(-0.25 + \frac{\alpha}{2}\right)^2}{2\sigma^2}\right) \tag{4.19}$$

The above equation can be further simplified to be:

$$\exp\left(-\frac{1 + 4\alpha}{32\sigma^2}\right) + \exp\left(-\frac{1 - 4\alpha}{32\sigma^2}\right) = 2. \tag{4.20}$$

An approximation of α_1 can be obtained since the term $\exp\left(-\frac{1+4\alpha}{32\sigma^2}\right)$ is close to

zero, and by solving $\exp\left(-\frac{1-4\alpha}{32\sigma^2}\right) \approx 2$, we get

$$\alpha_1 \approx \frac{1}{4} + 8\sigma^2 \ln 2. \tag{4.21}$$

Threshold of Population Split α for Second Phase Change. We proceed in a similar fashion to identify the threshold α for the second phase shift, by solving (4.15) for α in terms of fixed σ.

By equating $f(0.5) = f\left(0.25 - \frac{\alpha}{4}\right)$, we obtain

$$2\exp\left(-\frac{\left(\frac{\alpha}{2}\right)^2}{2\sigma^2}\right) = \exp\left(-\frac{\left(0.25+\frac{3\alpha}{4}\right)^2}{2\sigma^2}\right) + \exp\left(-\frac{\left(-0.25+\frac{\alpha}{4}\right)^2}{2\sigma^2}\right). \tag{4.22}$$

An approximation of α_2 can be obtained by observing that the term $\exp\left(-\frac{\left(0.25+\frac{3\alpha}{4}\right)^2}{2\sigma^2}\right)$ is close to zero and solving $2\exp\left(-\frac{\left(\frac{\alpha}{2}\right)^2}{2\sigma^2}\right) = \exp\left(-\frac{\left(-0.25+\frac{\alpha}{4}\right)^2}{2\sigma^2}\right)$. We get

$$\alpha_2 \approx -\frac{1}{3} + \frac{2}{3}\sqrt{1 + 24\sigma^2 \ln 2}. \tag{4.23}$$

Population Parameters and Candidate Behavior. Now that we have full equations for both of our phase changes, we can examine how the shape of the population density function affects the location of the candidate position equilibrium. Here we compare the deterministic model presented in this section where voters choose the most proximate candidate, and the model presented in the main text where $P = R = 5$ and $Q = 0$. In this case voters never stay home, and they give third-parties the same consideration as major parties, but they do not always choose the most ideologically proximate candidate. We visually compare the results between these two models in Figure 4.11.

In Fig. 4.11c, we see that for a fixed standard deviation, as the population split

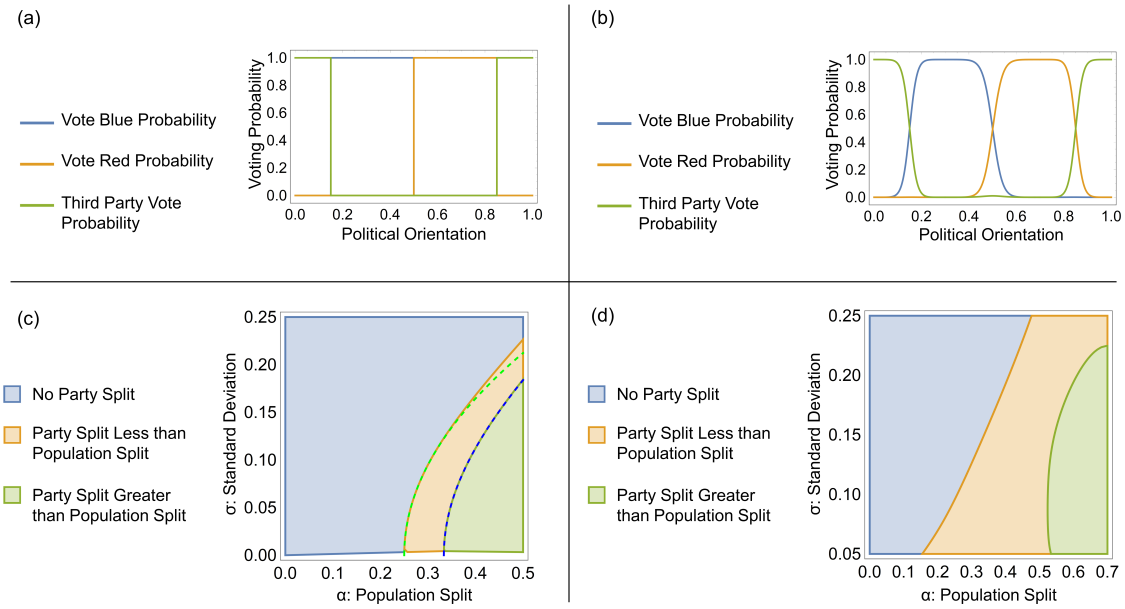

Figure 4.11: Voting probabilities and corresponding regions of behavior for the simplified model presented here and the main model with parameters $P = R = 5$ and $Q = 0$. In (a) and (b), we see how $P = R = 5$ and $Q = 0$ gives a rough approximation of the simplified model's voting behaviors with candidates at 0.3 and 0.7. In (c), we see how population structure affects the equilibrium position. The dashed curves show the analytic approximations of the phase changes (equations (4.21) and (4.23)), which match the actual phase changes very well. As the standard deviation increases, the sub-populations become more diffuse and the approximation becomes less accurate. (d) shows similar behavior as (c), with differences between the two being explained by the roughness of approximation demonstrated in (a) and (b).

increases, the density at the median goes down and the candidate split increases from 0 to greater than the population split. As the standard deviation for each of the two underlying distributions (σ) increases, the population becomes more diffuse, and the density around the median voter remains large, encouraging candidates to compete for the middle and allowing more extreme voters to choose an extreme third-party. While the two region plots have quantitative differences, they are qualitatively similar. In both Fig. 4.11c and Fig. 4.11d there is a narrow, diagonal band in which the party split is non-zero but less than the population split. Interestingly, this comparison also suggests that stochastic voter decision making widens the range of population ideology distributions that lead to outcomes in the "middle" phase space. That is, voter stochasticity may incentivize candidates to adopt differing positions, but not positions that are more extreme than the bimodal centers of the electorate.

Section 4.4

Discussion

The assumption that voters are purely rational is a strong one. There is strong evidence that collective opinion dynamics are shaped by processes of social influence [63, 113, 210], and voting behavior is no exception [251]. Here, we assume that voter irrationality is captured by their opinion formation process, and we instead consider how candidates should rationally respond to different levels of voter polarization and indifference. There is no shortage of proposed mechanisms that explain why voter attitudes have become more polarized over the years. Attitude polarization can result from the twin-mechanisms of homophily, a phenomenon that spans the social and biological sciences [120,190], and social influence, or the diffusion of pairs of associated beliefs [90, 91, 128]. The programming decisions of large media outlets [61, 245, 283], and the recommendation algorithms of social media sites can send people into wildly

different information landscapes [29, 173]. The influence these social processes may have on political candidates, however, is less examined. Our model shows that there are very realistic conditions under which rationally behaving major-party candidates will benefit from reinforcing polarization rather than by pivoting to the center.

This approach, like any model, is limited by the complexity that it omits. Regarding the specifics of voting in the United States, it omits details on the primary process and how candidates may be bound by verbal commitments they made to a primary electorate while running a general campaign. It omits the possibility of a serious third-party entering the race with strategic rather than ideological motives. It omits the draw that candidate personality may have on the behavior of voters. And it assumes a linear single-dimensional model of ideological positions rather than a multi-dimensional space [312]. It also omits the institutional and geographic complexities of voting induced by district or state-based electoral systems combined with the tendency for voters to self-sort geographically [184], and strategic attempts to manipulate this process such as gerrymandering [279]. Furthermore, our model assumes that voters have independent, static, and rational voting preferences. The study of collective opinion dynamics often focuses on the role of social influence, and voting behavior is no exception. Our model assumes a pre-polarized and static distribution of voter preferences, a simplifying assumption that allows us to focus on the rational behavior of candidates who are trying to win over a polarized electorate.

Yet, the minimal number of realistic assumptions necessary to obtain this result makes it all the more compelling and concerning. Stochastic voting behavior with a bimodal ideological distribution and the option to not vote for a major candidate may incentivize more extreme political parties. If we are to believe that voters follow candidates and parties just as candidates and parties follow voters, then a distributional tipping point may exist where voters and candidates chase each other to ideological

extremes. The solutions to this problem may be found in practices not explored in this model. For example, as we briefly explored in Section 4.2.5, ideologically motivated candidates running from the center may effectively "pull in" extreme but strategic candidates, in the same way that ideologically motivated extremist candidates can pull strategic candidates away from the center. The polarized political climate in the United States (and elsewhere) remains a serious problem, and continued reconsideration of rational choice voting models with more contemporary assumptions may provide the theoretical material necessary to develop pragmatic solutions for ending what is being referred to by some as a "cold civil war" [166].

Chapter 5

Spatial Games of Fake News

Section 5.1

Introduction

In the last decade, the study of misinformation has grown rapidly. In [263], Shin, Jian, Driscoll & Bar traced the lifecycle of 17 popular political rumors that circulated on Twitter over 13 months during the 2012 U.S. presidential election; they found that misinformation tends to come back multiple times after the initial publication, while facts do not. In [305], using massive Twitter datasets, Vosoughi, Roy, & Aral reported that the spread of true and false news follow distinctive patterns: falsehoods diffused significantly faster, deeper, and more broadly than truths in all categories of information [305]. Predictably, being surrounded by such a constant stream of fake news only makes it more difficult for consumers to differentiate real and fake news in the future [228].

Unfortunately, there are no simple solutions to "solving" the problem of fake news. Labeling articles as containing false information may backfire by giving consumers a false sense of security [227], and fact-checking on users posts may actually have the reverse effect of causing them to dig in and share even more false and toxic mate-

rial [195]. Even when individuals acknowledge that the information was false, their opinions about the spreaders may be unchanged [286]. Making matters worse, because almost all news media is advertiser-driven, content publishers may be incentivized to spread false information to increase engagement from consumers [278]. Because of these confounding factors for centralized fact-checking, we will focus on fact-checking at the level of the individual consumer.

Misinformation is able to thrive in social networks by dominating insular communities with little outside moderation known as "echo chambers." In [106], Evans & Fu investigated opinion formation on dynamic social networks through the lens of coevolutionary games [233], and using the voting records of the United States House of Representatives over a timespan of decades, the work presented and validated the conditions for the emergence of partisan echo chambers [65, 67, 260]. Integrating publicly available Twitter data with an agent-based model of opinion formation driven by socio-cognitive biases, in [307], Wang, Sirianni, Tang, Zheng, & Fu recently found that open-mindedness of individuals is a key determinant of forming echo chambers under dueling campaign influence.

A recent study suggested using an online crowdsourced fact-checking approach as one possible intervention to reduce misinformation [230]. Inspired by this empirical work, here we use spatial games to study the spread of fake news by including a model of distributed fact-checking efforts like "peer policing" to reduce the perceived payoff to share or disseminate false information (fake news) while rewarding the spread of trustworthy information (real news). Fact-checkers are placed into the population to model the effect of peer policing efforts. Our agent-based model, which has individuals sharing real or fake news depending on the behavior and success of their neighbors, is studied with simulations as well as a rigorous mathematical analysis. We find that the presence of echo chambers impede crowdsourced fact-checking, thereby requiring

a much higher critical distribution threshold of fact-checkers across the population.

5.1.1. Network Formation Models

In this study, we will run our game theoretic model of fake news on a variety of networks to test the effect of network structure on the spread of fake news. Here we describe some of the most common families of networks and their relevant properties.

In the simplest cases, the placement of individuals and the edges between them follow a very simple repeating pattern called a *lattice*. A few simple examples, shown in Figure 5.1, include the (potentially infinite) square lattice with degree four (connected to the vertices above, below, left, and right, also called the von Neumann neighborhood) or eight (also connected to the four diagonally nearby vertices, called the Moore neighborhood), the hexagonal lattice, and the ring lattice where each vertex is connected to the m nearest neighbors.

Lattices have found great applicability in statistical physics [31, 218, 257] as well in modeling human interaction [214, 216, 232, 288, 289]. Lattices are especially suited to statistical physics because each particle has a location in space and interacts with other nearby particles. While this may not be quite as true for social systems where friends often live very far apart, physical proximity still drives many social connections and lattices can provide an analytically clean starting point for investigations into the effects of network structure on behavior.

Lattices of a fixed size and interior degree are completely determined. However, there are other models of network formation that use randomness to create large families of distinct networks that still have one or more important features in common. When using these randomized network formation models, it is importnat to consider the underlying probability distribution on the space of all possible graphs.

The quintessential "random graph" model is the *Erdős-Rényi random graph*, and the model comes in two flavors. The first version, introduced by Paul Erdős and Alfréd

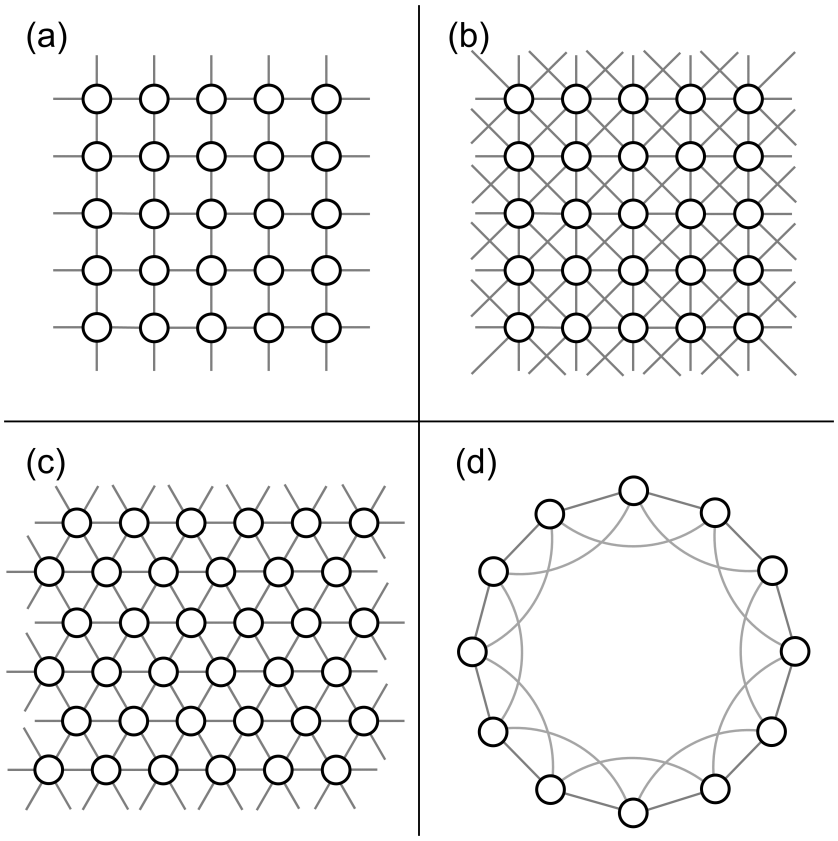

Figure 5.1: The four types of lattice described above: square lattice with degree 4 (a), square lattice with degree 8 (b), hexagonal lattice (c), and a ring lattice with 12 vertices and degree 4 (d).

Rényi, is known as $G(n, m)$ and is a uniform sample of all graphs on n vertices with m edges [102]. The second, independently developed by Edgar Gilbert, is known as $G(n, p)$ and adds every possible edge to the network with probability p [126]. Both models are popular for their simplicity and elegance, and often easily obtained analytical results. However, the Erdős-Rényi random graph model does not typically produce networks resembling those found in real life [5,281]. Therefore, other network formation models have been developed to create more "realistic" networks.

The first failure of Erdős-Rényi random graphs is in their degree distribution. Degree for the vertices of $G(n, p)$ is binomially distributed. On the other hand,

it has been well-documented that real-life networks have very different distributions [51, 208, 242]. For example, many naturally occurring networks have *scale-free degree distributions* in which most vertices have a relatively low degree, but a few have extremely high numbers of neighbors. Price introduced one such model in 1976 for directed networks [243], and over 20 years later, Barabási-Albert independently developed a model of "preferential attachment" for undirected networks [28, 209]. Both models consider how networks are formed and have a temporal aspect to their creation: after beginning with a small network, vertices are added to the network one at a time, choosing a predetermined number of individuals to be neighbors. Critically, in the preferential attachment model, neighbors are chosen with probability proportional to current degree, so more popular individuals are more likely to be chosen, demonstrating the "rich get richer" phenomenon that appears in various areas of social science [269]. The networks created by this model tend to have a degree distribution which follows a power law, meaning the probability of a having degree k is proportional to $k^{-\alpha}$. It turns out that this can have important implications for dynamics on a network [6, 225, 258]. We see this in Figure 5.5 when the Twitter network is vulnerable to targeted removal of high-degree individuals.

The last model of network formation that we will make use of is the *Watts-Strogatz small-world model*, which attempts to simultaneously capture two interesting characteristics of social networks: large *clustering coefficient* and small *average path lengths* [308]. The clustering coefficient is the probability that two individuals with a mutual neighbor are connected by an edge. In social networks, where people often meet through mutual friends, this can be quite high [207, 208, 308]. At the same time, thanks to Stanley Milgram's famous "six degrees of separation" experiment [193], we know that path lengths in social networks are deceptively small, meaning it takes a relatively small number of steps to reach a large number of people [19, 93]. Watts-

Strogatz small-world networks are formed by taking a ring lattice with a naturally high clustering coefficient and adding a small number of random shortcuts that drastically reduce path lengths without significantly altering the average clustering coefficient.

We will use several of these models in this chapter, and all have found use in the study of social systems.

The work presented in this chapter is currently being prepared for publication.

Section 5.2

Methods & Model

The foundation of our model is inspired by the virtual interactions that occur repeatedly on online social media sites. To begin, an individual posts a news story (either true or false) for all of their friends or followers to see. Those who believe the story is true can react positively to the post by liking or sharing, while those who disagree may simply ignore the post or even attempt to debunk a false story by pointing out flaws or sharing a link to a fact-checking website, potentially causing embarrassment and inflicting a penalty on those who share fake news.

A key focus of this chapter is peer fact-checking as opposed to institutional fact-checking. Recent research shows that "inoculation" with exposure to a weakened version of misleading arguments (similar to vaccination ideas) is effective at reducing susceptibility to misleading persuasion and thus confers psychological resistance to fake news [24, 188, 253, 254]. By training some subset of the population to identify and respond to fake news, we can create a decentralized fact-checking system where innoculated individuals will be positioned to apply pressure to their social neighbors that share fake news while also supporting their real news sharing neighbors. Inspired by "zealot models" from the field of opinion dynamics [306], we assume that these fact-checkers will never change belief because of the behavior of those around them,

having been successfully inoculated against fake news,

Consider a network with a two-layer structure: the *spreader layer* describes the information sharing dynamics among spreaders of real news vs. fake news. Unfortunately, fake news tends to spread more effectively than real news on social media [305], so our model gives a higher payoff to individuals sharing fake news in simulations (our analytic results, on the other hand, work for any set of prescribed payoffs). If a spreader of real news (labelled A) interacts with another A, they both receive a moderate payoff; if a spreader of fake news (labelled B) interacts with another B, they both receive a slightly larger payoff; if an A interacts with a B, both receive a very small (or possibly negative) payoff. To contain the spread of fake news, the natural advantage given to B players will have to be counterbalanced by a penalty inflicted by fact-checkers. The second layer of the social network describes the enforcement infrastructure where these designated fact-checkers, denoted by C, perform distributed fact-checking to their spreader neighbors: they provide a reward to A and a harsh penalty to B. We assume the proportion of fact-checkers p_C is prescribed and static, representing the level of innoculation effort. The payoff to fact-checkers is irrelevant as the fact-checker population is static, so for simplicity we arbitrarily set it to zero. A selection strength parameter controls how much impact an individual's payoff has on her reproductive success in the update step. The payoffs and selection strength can take arbitrary numerical values, but for the rest of this paper, unless otherwise noted, we will use a selection strength of $\beta = 0.5$ and the payoff matrix for this symmetric, two-player game will be:

$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & -4 \\ 0 & 0 & 0 \end{pmatrix} \end{array} \tag{5.1}$$

Notice that in Equation (5.1), the payoff for fake news is twice the payoff for real news, but fact-checkers also inflict a stiff punishment.

Over time, if individuals see that only certain types of stories receiving positive feedback, they may be convinced of the accuracy of those (potentially false) narratives [228] and begin sharing those same stories themselves (Figure 5.2c). We will use a death-birth process for the evolutionary strategy update [216] to capture this social imitation phenomenon. After computing the expected payoff π_i for every individual i, a focal individual imitates the strategy of one of its neighbors, chosen with probability proportional to their fitness $f_i = \exp(\beta\pi_i)$. Thus, individuals with high payoff are likely to be selected, but even individuals with a low payoff due to repeated fact-checks or social isolation could be chosen to reproduce occasionally.

In our investigation, we use two flavors of this update rule: synchronous and asynchronous. In the synchronous update, used in our simulations, every individual simultaneously updates their strategy, while in the asynchronous update, which lends itself to easier mathematical analysis, a single individual is chosen uniformly at random to update. These two update rules will lead to very similar outcomes, and the minor differences between them are manifested only in edge cases that occur rarely. Keeping this in mind, we will treat them as the qualitatively same process operating on different time scales.

The basic operating procedure for our model is shown in Figure 5.2. First, individuals play the fake news game with neighbors by broadcasting a real or fake

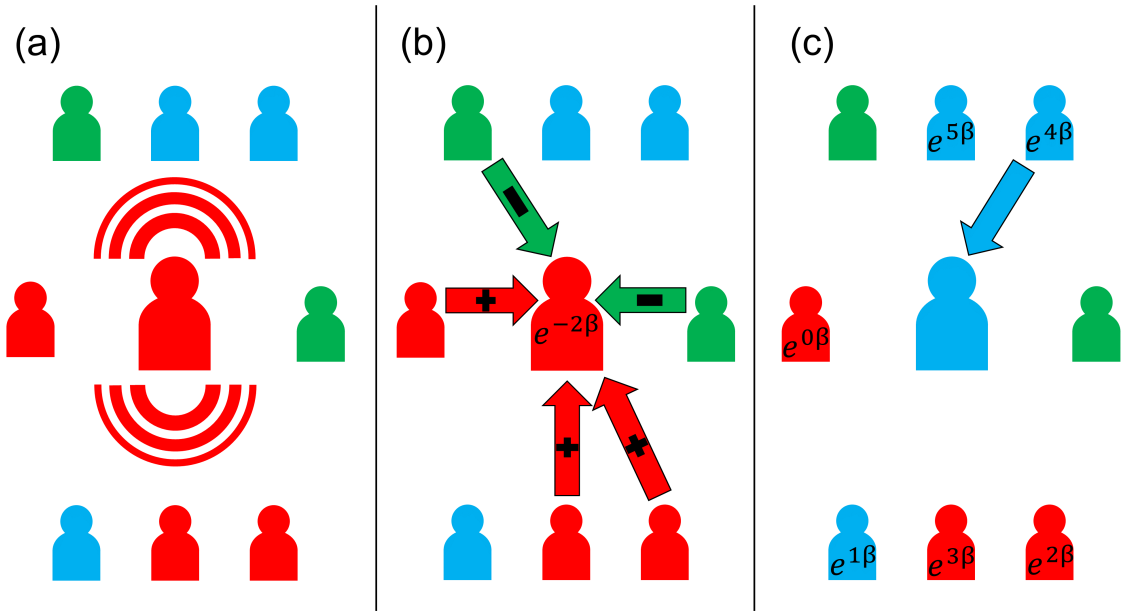

Figure 5.2: Model schematic. We model information sharing and fact-checking through the lens of spatial games. First, every individual shares news that is either true (blue) or false (red), shown in (a). In (b), we see a focal individual receiving positive or negative feedback from her neighbors depending on their relative beliefs. These information sharing dynamics are modulated by the presence of crowdsourced fact-checkers (green), characterized by the effect of their policing (positive or negative) and their static nature. Finally, in (c), individuals in the spreading layer updates their strategy by copying a neighbor proportional to fitness.

post. This post generates positive and negative feedback, which is converted into a fitness. Figure 5.2c demonstrates the asynchronous update, where only a single focal individual updates strategy by considering the fitness of all neighbors.

Our study of the spread of fake news is focused on three types of networks: a 30×30 square lattice [214], the family of small-world networks [308] (also with $N = 900$), and a portion of the Twitter follower network [255] ($N = 404719$). Our small-world networks are calibrated to have high clustering coefficients and short path lengths. To accomplish this, we use the following parameters: base degree is 8 and the rewiring probability is 0.03, giving us approximately 200 shortcuts. The Twitter network is interesting for its size but also its natural clustering and the gatekeeping individuals that control the flow of information through the network. Although edges in the network were originally directed, we symmetrized the network to match the bidirectional flow of information in the original model.

To initialize the system, we assign some fraction p_C of the individuals as fact-checkers, and the rest we set to be A or B players with probability $\frac{1}{2}$. After initializing the system, we allow it to evolve using the one of the updating processes described above until all possible players are sharing the same type of news or a predetermined number of time steps is reached. At the end of the simulation, the type of news with more sharers is said to be dominant, and if there are no individuals sharing one type of news, we say that that strategy has gone extinct and the other strategy has fixated.

Section 5.3

Results

5.3.1. Echo Chambers and Critical Fact-checker Density

When there are very few fact-checkers, the natural advantage that fake news sharers have allows them to drive the real news sharing strategy to extinction. Similarly,

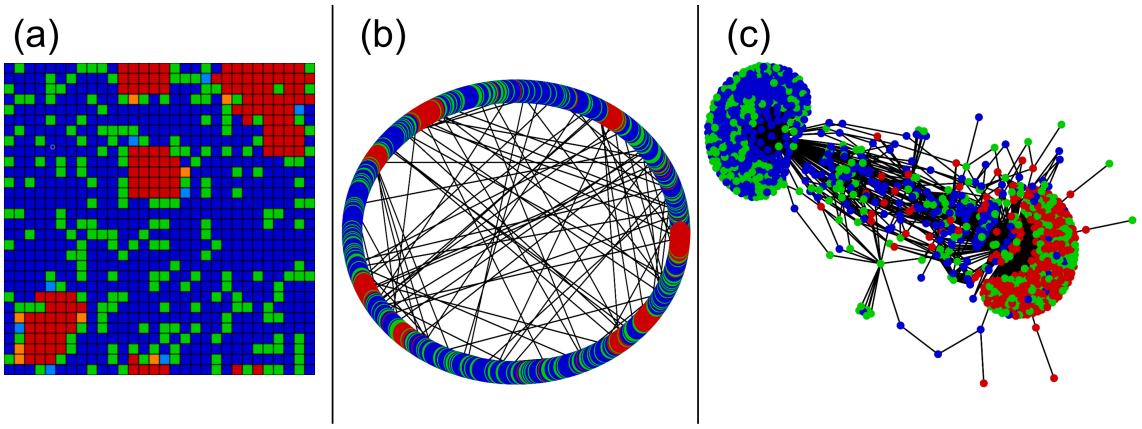

Figure 5.3: Echo chambers of fake news spreaders in a majority real news-spreading population that are isolated from the rest of the population. In (a), the lightly-colored individuals are those that have changed strategy recently. The network in (c) is a small breadth-first subgraph of the Twitter network of approximately 1,000 vertices, but the simulation was run on the entire $\approx 400,000$ vertex network (see Methods & models).

when there is a sufficient fact-checker presence, the risk of punishment for spreading misinformation is too great and the entire population shares real news. However, there is a wide range of fact-checker densities where we see the spontaneous formation of echo chambers in our simulations. We define echo chambers by their longevity, as either real or fake news goes extinct unless the minority strategy manages to form small, highly connected communities that are secured from invasion by the majority strategy. For additional discussion about the longevity of these pseudo-steady states, see Section A.1. Figure 5.3 shows examples of these echo chambers on the three different network topologies we studied.

Once they form, these echo chambers are incredibly resistant to invasion, resulting in a *pseudo-steady state* that cannot last forever, but will take an extremely long time to break down. Observe in Figure 5.3a that the only individuals changing strategy are on the borders of the echo chambers in the system. It is very unlikely that a

small perturbation on the border will result in any change to the interior of the echo chamber, because while individuals on the periphery of the echo chamber may be exposed to both views, those in the interior are surrounded by like-minded individuals and have high fitness to support the more exposed group members on the border.

Unsurprisingly, the density of fact-checkers determines which type of news sharing is the majority and which is the minority, trapped in small and isolated communities. Figure 5.4 shows how the long-term behavior of the system changes as the density of fact-checkers grows. When a critical number of fact-checkers is reached, the probability of success for real news sharing increases dramatically. This critical density is different for different network types: $p_c \approx 0.235$ on the square lattice, $p_C \approx 0.2$ for small-worlds, and $p_C \approx 0.275$ for the Twitter network. These results come from simulating 50 populations at 20 different, evenly spaced fact-checker densities. At 5,000 time steps, a pseudo-steady state was declared, except the Twitter simulations which ended at 500 time steps for computational reasons.

We can compare these results to the simple case of an infinite, well-mixed population evolving according to replicator dynamics [153]. When initialized with some fraction p_C of fact-checkers and the rest of the population evenly divided between real and fake news, the system will evolve so that real news grows and fake news is driven to extinction as long as $p_C > \frac{1}{11} \approx 0.091$. We conclude that the network structure of the spatial game makes containing fake news significantly more challenging. In fact, between two to three times as many fact-checkers are needed to contain the sharing of fake news in small, isolated echo chambers, and even more fact-checkers are needed to have a good chance of driving fake news sharing behavior to total extinction.

Our comprehensive simulations using the payoff values in Equation (5.1) and the synchronous update rule confirm that the formation of echo chambers occurs across a wide range of payoff values, selection strengths, and network structures.

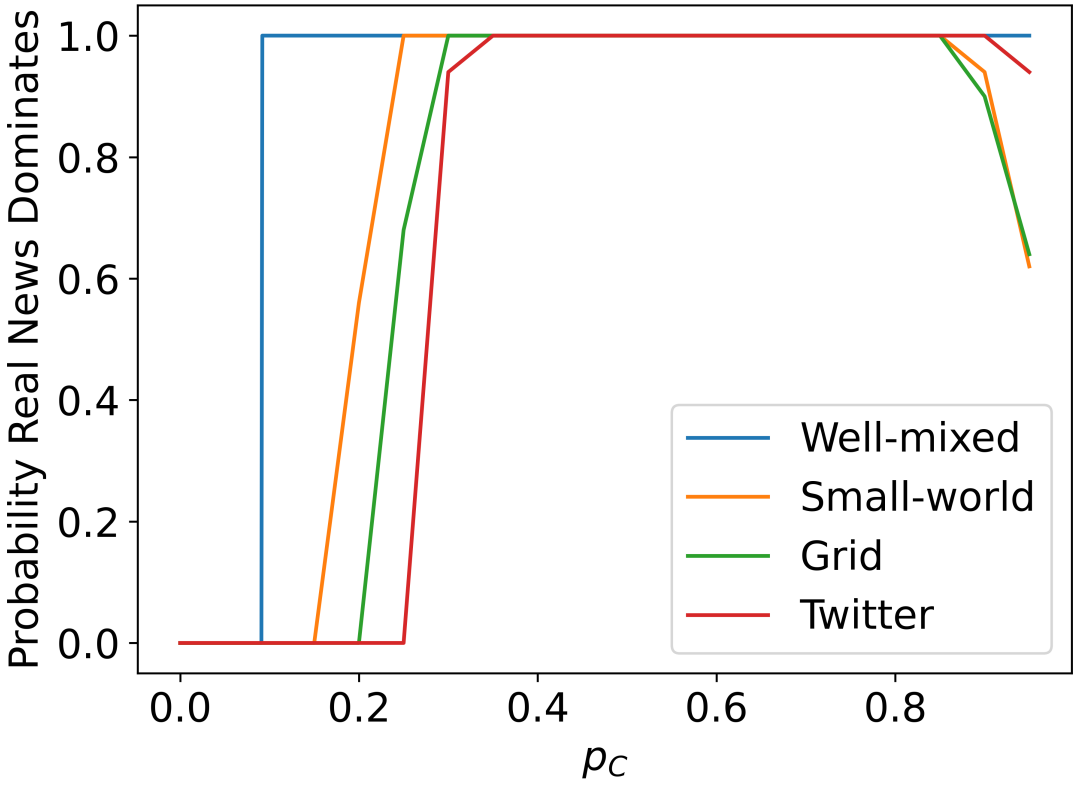

Figure 5.4: The probability that over half the viable population ends up sharing real news as a function of fact-checker density for different network topologies.

Local variation in fact-checker density means in some areas there are no fact-checkers (leaving room for a fake news echo chamber) and in others they make a fact-checking wall which becomes more and more difficult for fake news sharers to penetrate as selection strength grows.

5.3.2. Targeted Fact-checking

So far, we have only considered populations where fact-checkers are placed randomly. However, in almost all networks, some vertices are more centrally located than others, and this effect is particularly pronounced in naturally formed social networks. To improve the effectiveness of crowdsourced fact-checking with limited resources, it is vitally important to study targeted intervention algorithms by selecting the most central vertices. Once again, we run simulations with 50 iterations, 20 density values, and a limit of 5,000 (or 500) time steps. Our results, shown in Figure 5.5, focus on two measures of network centrality, degree and betweenness [30], but there are many more centrality measures and the problem of selecting individuals for optimal fact-checking remains an open problem. Since all vertices in an infinite square lattice have the same centrality, our work here is restricted to small-world networks and the Twitter network.

Figure 5.5 has several interesting features. First, we see that in small-worlds, using the degree and betweenness centralities have virtually the same performance. This is unsurprising as the additional shortcut edges are what create short path lengths and therefore give those individuals a high betweenness value, so the two centralities are highly correlated. More surprising is the fact that targeted fact-checking is only marginally more successful than random fact-checker placement, which can be seen by comparing Figures 5.4 and 5.5. This may be due to the relatively uniform nature of small-world networks, where there is very little variation in degree from vertex to vertex.

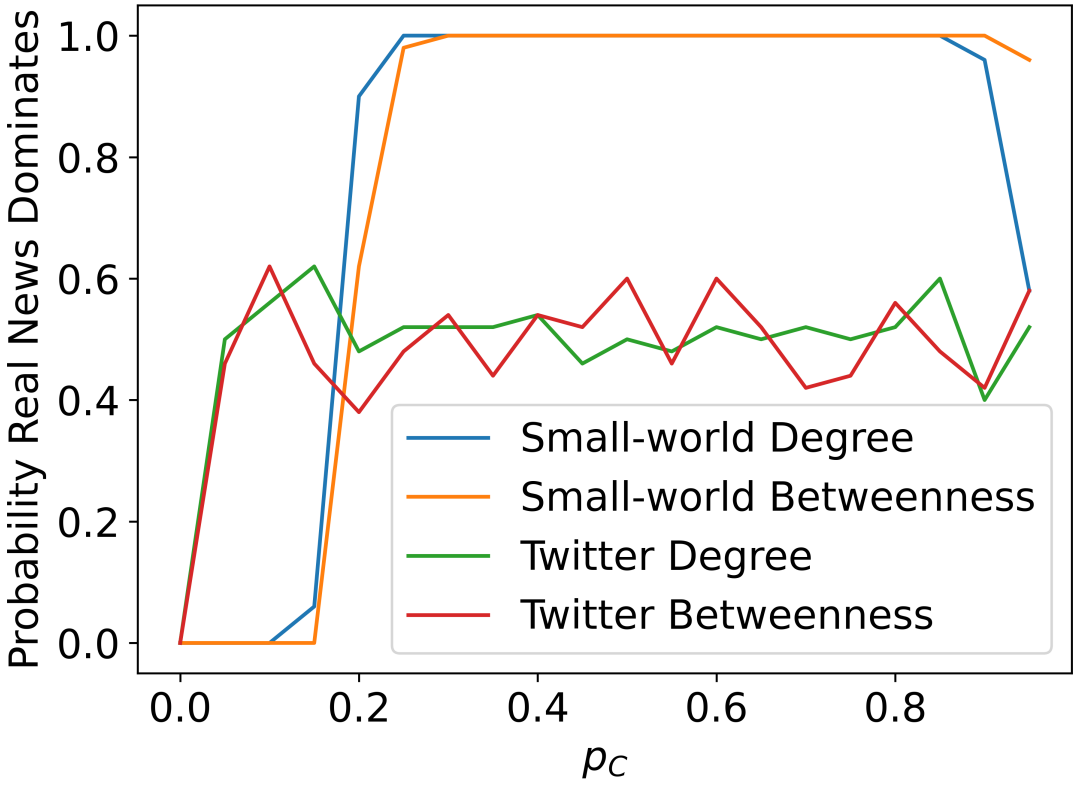

Figure 5.5: The probability of real news dominating on small-world networks and the Twitter network using the degree and betweenness centralities to place fact-checkers.

However, the Twitter network has much more diversity in its degree distribution and here we see a large change between random and targeted fact-checking. By targeting high degree or betweenness centrality individuals to be fact-checkers, we quickly separate the spreader layer of the network into small disconnected network components, as these types of networks are very vulnerable to targeted percolation of the most central vertices [6, 297]. After removing the high-centrality vertices from the spreader layer, what remains is thousands of disconnected and extremely small networks in which there is no diffusion of any strategy regardless of payoffs because there are no connections along which strategies can spread. These isolated singletons and pairs are completely restrained by their initial conditions and it is about equally likely that the initial random distribution will have more fake or real news sharers, so the probability that real news "dominates" by having over half the viable population hovers around 0.5 for almost all values of fact-checker density. We observed a similar effect in Figure 5.4 for very high fact-checker densities, where the probability of real news dominating actually decreases when $p_C > 0.85$, but this level of fact-checking is clearly unrealistic.

This suggests that in real world networks, a *targeted* crowdsourced fact-checking effort where fact-checkers are also encouraged to share real news with their neighbors could be highly effective with relatively little collective effort, as the network structure will actually benefit real news instead of fake news by amplifying fact-checking efforts. Enhancing our model by allowing fact-checkers to "pass along" real news between neighbors is one way to more effectively study targeted fact-checking algorithms.

5.3.3. Analytic Results under Weak Selection

The selection strength β determines the effect payoff from the fake news game has on reproductive success. As β approaches zero [215, 216], the evolution of the system comes to resemble *neutral drift*, in which individuals choose strategy with no regard

for payoff. In this domain, the pseudo-steady state with its echo chambers becomes transient and short-lived. In the following section, we derive analytical results in this limit of weak selection.

Assuming a k-regular network structure like the square lattice, we will use an extended pair approximation method [168] to study the emergence and spread of honest behavior. For example, one quantity of interest is the probability that a population with some initial condition evolves so that the entire viable population evolves to play A, called the fixation probability of A. Our aim here is to study the effects of changing the payoffs for real news, fake news, and fact-checkers, so we will begin with a general payoff matrix:

$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} a & b & \alpha \\ c & d & \gamma \\ 0 & 0 & 0 \end{pmatrix} \end{array} \tag{5.2}$$

In the limit of weak selection $\beta \ll 1$, we will obtain closed-form analytical conditions for the fixation probabilities of A and B as functions of these payoff values.

When we suppose the network system begins with a proportion $p_A(0) = p$ of A individuals, we can calculate the expected value $m_A(p)$ and variance $v_A(p)$ of the change in abundance of A during the asynchronous update step where a single random individual considers changing strategy. The fixation probability of A for an initial fraction p of A players, denoted $\rho_A(p)$, satisfies the diffusion approximation equation for large populations (see [216] for details):

$$m_A(p)\frac{d}{dp}\rho_A(p) + \left(\frac{v_A(p)}{2}\right)\frac{d^2}{dp^2}\rho_A(p) = 0 \tag{5.3}$$

with the boundary condition $\rho_A(0) = 0$ and $\rho_A(1) = 1$. This equation has closed-form

solution, and thus we can obtain an exact formula for ρ_A.

Our derivation of the following explicit expressions for the fixation probabilities in terms of the payoff values, lattice degree k, and fact-checker density p_c, is detailed in Section A.3. For small values of p:

$$\rho_A(p) \approx p + \frac{\beta N p(1-p)}{6k}(-u_1 - 3u_2) \tag{5.4}$$

$$\rho_B(p) \approx p + \frac{\beta N p(1-p)}{6k}(-w_1 - 3w_2) \tag{5.5}$$

where $u_1 = (a-b-c+d)\Big(1 - k^2 - \frac{1+k}{(p_C-1)(1-p_C)}\Big)$, $u_2 = -a+b+c-d-ak+bk-bk^2+dk^2+(k-1)\Big(c+(b-\alpha+\gamma)k-d(1+k)\Big)p_C$, $w_1 = u_1$, and $w_2 = -(u_1+u_2)$.

In particular, we are interested in the emergence of new behavior in a previously homogeneous population. We calculate the fixation probability ρ_A of a single initial A player, called the invasion probability, and derive the conditions for truthful behavior to be favored, that is, when $\rho_A > 1/N$ where N is the size of the population. We also repeat the process for a single B player. Using Equations (5.4) and (5.5), we examine the effect p_C and γ, the punishment defectors suffer from fact-checkers, have on the invasion probabilities of real and fake news.

This allows us to determine the conditions under which fact-checking will be effective at stemming misinformation and quantify how steep the penalty γ needs to be for a given proportion of fact-checkers, p_C, in the system. In Figure 5.6a, we see that for strong penalties, $\gamma < -4$, only a fifth of the population or less needs to be fact-checkers for selection to favor real news. However, as γ gets closer to zero, the number of fact-checkers need goes up to about half the population. The green region of the $p_C - \gamma$ plane shows where selection favors fake news; this only happens when there are very few fact-checkers. Notice that there is a wide region in orange

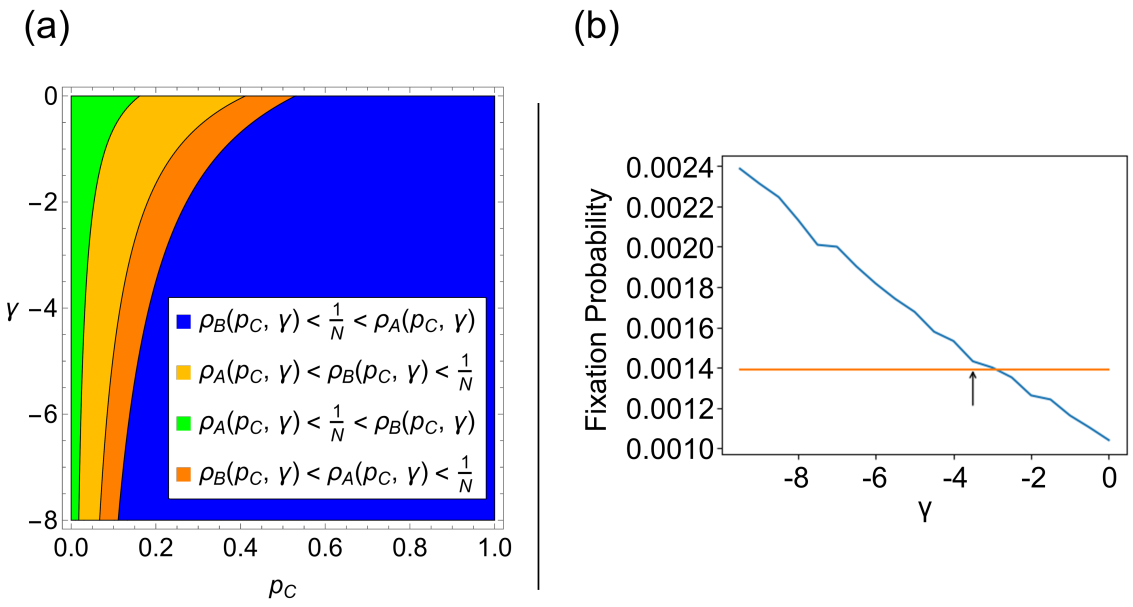

Figure 5.6: The invasion probabilities of real and fake news spreaders in the limit of weak selection using payoff values from (5.1), except for γ which varies from 0 to -8. In (a), we see what regions of the $p_C - \gamma$ plane give true stories an advantage (blue region), false news an advantage (green region), or neither (orange regions). In (b), we see an approximation of the invasion probability for a single real news sharer from simulations, when $p_C = 0.2$ and $\beta = 0.0001$. These simulation results intersect the threshold line $\frac{1}{N} \approx 0.0014$ close to where it was predicted by the analytic results, indicated by the arrow.

where selection does not favor invasion by real or fake news. This is because the fake news game is a coordination game that tends to put minorities (like a single invading mutant) at a disadvantage. These analytic approximations closely match extensive simulations, as shown in Figure 5.6b.

Section 5.4

Discussion and Conclusion

This work adds to the growing body of research surrounding fake news, echo chambers, and fact-checking and we believe that this work has immediate implications for the study of misinformation. We have shown that the spatial structure of social networks tends to favor the spread of fake news, but by carefully selecting fact-checkers, that same structure can be used to combat misinformation by amplifying the effects of fact-checking.

Our analytic results allow us to easily test potential combinations of reward and punishment and use both "carrots and sticks" to encourage real news and dampen fake news. Like previous work studying public goods games, we see that a strong punishment of defectors is effective at stopping bad behavior [145, 267, 268].

Future work combining potential experimental behavior data [228] with our present model will help incorporate relevant social network and psychological factors in our research. In particular, the constants in the payoff matrix and the selection strength were chosen fairly arbitrarily. Analyzing real-world data may allow us better estimates of some of these values, which in turn can give better actionable advice about how to actually control the spread of fake news. We would also like to analyze preexisting data sets or create new empirical studies to confirm our predictions regarding the effects that the rewards and punishments of sharing real and fake news have on the ability of fake news to spread through a population. As an example, perhaps

placing fact-checking comments at the top of any fake news threads would sufficiently increase the punishment suffered by fake news's sharers to prevent its spread.

Recent theoretical research has demonstrated that partisan bias [164] and information cascades [295] are two possible explanations for the formation of echo chambers. Our work here shows that the spatial distribution of fact-checkers can be another force behind echo chamber creation. However, these echo chambers require certain conditions. One requirement is a reasonable selection strength value, but there is also a lot of work to do studying the impact that network topology has on the spread of fake news and the formation of echo chambers. An obvious application of this work is to look at real-world social networks and determine what structural changes can be made to discourage the spread of fake news. Preliminary results show that the formation of resilient echo chambers is dependent on the type of network used. While social media sites do resemble lattices or small worlds in some respects, there are other properties of social networks that may be more or less conducive to echo chamber formation.

Extensions of our present work on targeted fact-checking efforts will likely lead to useful insights for optimizing field deployment of crowdsourcing fact-checking. There will be a good deal of further work to do, for example, on using other network topologies and other targeting centralities. In addition, the use of larger network data sets will give us more realistic behavior as there may be large-scale social network features essential to the development of echo chambers that are not captured in any of the network models we used.

Last but not least, our present work will help stimulate future work extending targeting algorithms to multiplex networks that take into account the fact that the interconnected ecosystems of social media platforms enable multi-channel communication and spillover from one platform to the other. In doing so, we hope to develop

mechanistic models that allow us to explore realistic extensions incorporating social psychological factors such as heterogeneity of social influence, repeated exposure, and pre-existing beliefs.

Chapter 6

Conclusion

Section 6.1

A Brief Review

Mathematics provides a vast range of tools to study real-world phenomena. Models have been developed to take real, physical scenarios and convert them into equations, algorithms, and other abstract ideas, at which point they can be mathematically analyzed. Historically, in the study of social systems, game theory and network science have been particularly versatile and useful in a wide range of applications, and this thesis continues that tradition by applying these ideas to a range of different collective action problems. Stochastic elements are often used to include the effects from other factors that are not accounted for in the model, similar to the effect of white noise in other applications, and stochastic behavior can help groups reach an optimal outcome that may have been unreachable with greedy decision-making.

We find graph colorings to be a useful analog of coordination games that require players that interact to make different choices, like university registrars scheduling exams so students do not have two exams at the same time or radio stations choosing different broadcast frequencies so there is no interference. This is different from

the traditional study of graph colorings in which a single decision-maker has access to all the information in the network. In our model, individuals can only see the small subgraph that consists of them and their neighbors. When each individual has such limited information, this becomes a non-trivial problem that is often unsolvable without the use of random, non-greedy decisions by the players. We find that random behavior is useful only in moderation, as too much randomness will keep the system from ever settling into the global optimum. Also, the exact way in which randomness is applied matters; depending on the size of the network and the number of individuals equipped to make random decisions, it may be better for random individuals to behave in a more restrained way.

By phrasing the random behavior of the players in terms of Markov chains, we are able to show that the anti-coordination game and the coordination game are equally difficult when there are only two choices. This is a surprising result, because these are fundamentally different problems for a single omniscient observer. Solving the coordination game and finding a uniform coloring is always trivial with global information but solving the anti-coordination game and finding a graph coloring can be extremely difficult when there are many choices. This work may be a guide for drawing connections between other kinds of collective action problems in the future.

Voting as a social system is obviously critical to any democratic society and has therefore been extensively modeled and studied by generations of political scientists. We use a classic one-dimensional model of voting to study three competing factors that influence modern US elections: voter abstention, radical third parties, and growing political polarization. We find that these factors working in concert can cause vote-maximizing politicians to become more polarized than the electorate, in defiance of the Median Voter Theorem. Troublingly, there is a "tipping point" where the political elites actually become more polarized than their voters, as they seek to

defend themselves to their supporters against extremist challengers. We also use real data from surveys and from Twitter as a first attempt to estimate some of the parameter values that determine voters' pragmatism, their incentive to not vote, and their interest in third-parties.

Related to rising polarization, one of the most troubling social phenomena of the last few years has been the rapid rise of misinformation and fake news spreading across social media. Drawing on a large body of recent empirical work, we develop a model of fake news spread that simulates the spread of false stories on a networked population. Our model allows us to consider how the benefits of sharing false stories can be counterbalanced by the possibility of being publicly shamed for spreading disinformation. We test the effectiveness of a crowd-sourced peer policing strategy that relies on citizen fact-checkers, and see how small variation in fact-checker density facilitates the formation of echo chambers where individuals who hold a minority opinion can believe they are in the vast majority. Using various social network models, we are also able to provide evidence that a complex social structure gives fake news a significant benefit as it is able to occupy small, isolated cliques that are hidden away from the mainstream. Two to three times as many fact-checkers are required to contain misinformation on a social network as compared to a well-mixed population where everyone knows everyone.

These are powerful results, but it is important to keep in mind that they follow from models of human behavior, and are therefore constrained by the fidelity of the models to reality. We make many assumptions about individuals' behavior that make the model work but do not perfectly represent actual behavior.

In the distributed graph coloring problem, we assign individuals one of several extremely oversimplified algorithms to guide the update process, but a preliminary analysis of real human behavior data in the distributed coloring problem from Shirado

and Christakis [264] shows that very few individuals pick a given strategy and stick with it for the duration of the experiment. Despite this gap between model and reality, our results can could still help improve human behavior by priming individuals to be prepared to make more random choices or consider the total size of the network before making a random choice.

The assumptions we make about voting behavior and polarization are discussed in great detail in Chapter 4. The number of assumptions we needed to make to build a functioning model of polarization means that our results have limited applicability in highly sensitive activities like election forecasting. Instead, we see value in this work as an addition to the literature examining the effects of voter abstention, third-party candidates, and polarization on candidate behavior. Much like the Median Voter Theorem is considered to be a broad, suggestive idea instead of a hard and fast rule, our model can be useful to examine the interactive effects of several different factors in voter behavior without claiming to describe every aspect individuals consider when casting their votes.

Our study of fake news also simplifies individual behavior drastically by supposing that all possible incentives are contained in one small payoff matrix. There are many empirically observed effects related to fake news that are simply approximated in the relative payoff values. For example, the possibility of fake news sharers increasing their fake news sharing behavior after being fact-checked [195] is ignored in this model, and we hope that any effect this might have is encompassed in the advantage fake news sharers have over real news sharers in the absence of fact-checkers. This is obviously an oversimplification, and a future model could be designed specifically to study this and other aspects of online misinformation spread.

In each of these models, simplifications are made because it is impossible in practice (and probably in theory) to design a model that considers every factor in human

behavior. The mark of a good model is having enough sophistication to derive meaningful results while also being simple enough to be understandable to researchers in the field and perhaps allow for some analytic, as opposed to simulation based, results. We believe that the models presented in this thesis do a good job of existing in the middle of this spectrum, with analytic results in the simple cases while also using computer simulations to explore the full range of behavior possible in the model.

Section 6.2

Future Work

Each of these separate mathematical analyses of social systems also has their own natural path for additional study.

6.2.1. The Distributed Graph Coloring Problem

Our current work using graph colorings has focused on the simplest case: using two colors to solve the simplest of anti-coordination games. The obvious next step is to examine what happens when the game becomes more complex and there are more roles or colors to choose from. It is not clear how the likelihood of gridlock changes as the number of colors, the size of the network, and the number of edges in the network grow. Understanding these relationships could be instrumental in understanding coordination problems for larger group sizes.

For two colors, we know that coordination and anti-coordination are equally difficult problems. This equivalence will not hold for more than two colors, but it is an open question to think about how much more difficult the anti-coordination game is when there are more options. Like the question of gridlock, this is complicated by the fact that the number of colors needed is intricately linked with network size and structure.

6.2.2. Voting

There are at least two avenues for extending our study of voting and polarization. In the first, we take a more sophisticated look at voter abstention, and think about how the positions of both major parties changes the behavior of a single voter. In the second, we consider how cognitive dissonance and holding seemingly contradictory beliefs is necessary for social cohesion in a pluralistic society. We have also begun some promising work using our one-dimensional voting model to study the impact the system of primary elections has on which candidates win their parties nominations and which win in the general election.

6.2.3. Fake News

One immediate extension of our work with fake news is to examine how different network topologies impacts the spread of misinformation and the formation of echo chambers. Much of our work is done on square lattices, but preliminary results show that certain network types lend themselves to the formation of these isolated communities while others are simply too connected to allow small groups to wall themselves off from the rest of the population. Studying a wider variety of network types also means we can go deeper in our investigation regarding identifying the most important individuals to be fact-checkers. A targeted fact-checking effort can reduce fake news with minimal community engagement if the fact-checkers are carefully selected. Determining which individuals are most important for the containment of fake-news has obvious implications for tackling the very real problem of misinformation today.

Appendix A

Derivation of Analytic Results for Fake News Invasion Probabilities

This supplementary information contains some additional exploration of the fake news spatial game described in the main paper, as well as the derivation of the invasion probabilities in the limit of weak selection.

Section A.1
Echo Chamber Longevity and the Pseudo-steady State

In this section, we expand our investigation into the role fact-checkers play in containing the spread of fake news. The density of static fact-checkers has a significant effect on the formation of echo chambers and which strategy "dominates" by controlling over half the viable population. Fig A.1a and A.1b show two examples of this on the square lattice. Different strategies dominate, dependent on the fact-checker density.

In the main paper, we focused on a critical value of p_C at which point selection favors real news instead of fake. However, there is an additional point to consider.

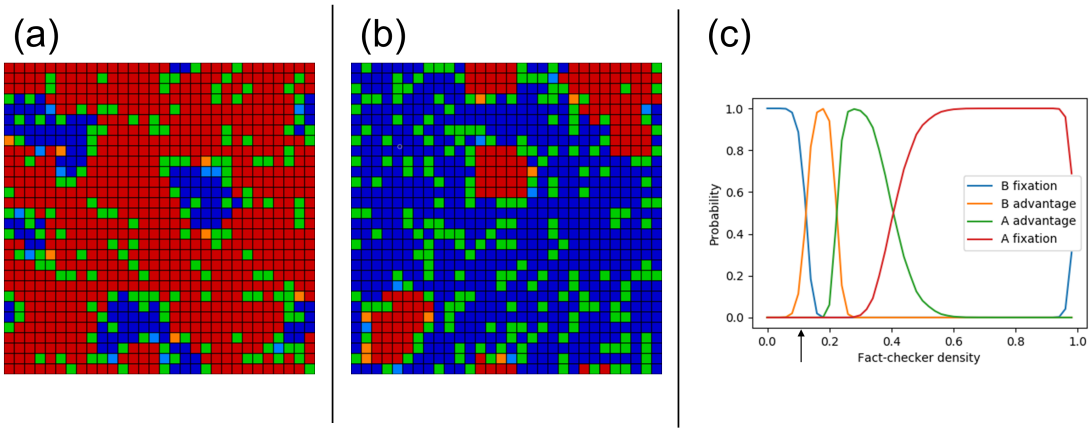

Figure A.1: Panels a and b show echo chambers of real news (blue) or fake news (red) sharers that are isolated from the rest of the population by a barrier of fact-checkers (green). Lightly-colored individuals are those that have changed strategy in the last time step. The plot in (c) used simulations to show how the long-term behavior changes as the fact-checker density varies, with the arrow indicating the fact-checker density at which real news has an advantage in a well-mixed population, $p_C = 1/11$. As the number of fact-checkers increases, the population moves towards more real news and less false news stories being shared.

Instead of simply containing fake news to isolated echo chambers, we may want to select enough fact-checkers to completely eradicate fake news. On the other hand, for a sufficiently small number of fact-checkers, it is extremely likely that eventually the entire population will be sharing fake news. Therefore, there are actually four different regions of behavior: fake news (B) fixates and real news (A) goes extinct, fake news has the advantage in the population with small real news echo chambers, real news has the advantage with small fake news echo chambers, and real news fixates while fake news goes extinct. This sequence of behaviors and their probabilities are shown in Fig A.1c.

We can see the formation of echo chambers for a wide range of fact-checker densi-

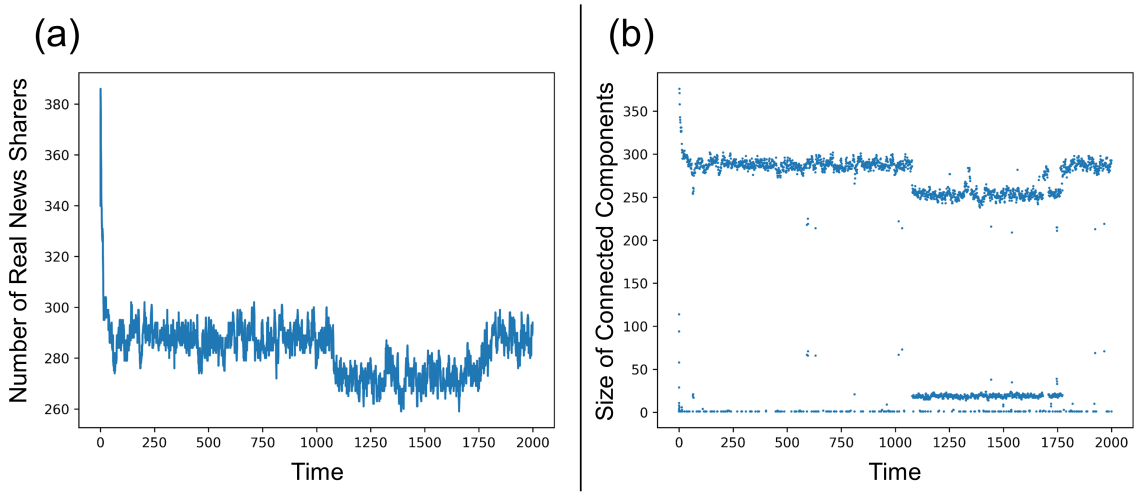

Figure A.2: The characteristic evolution of a 900 individual population with $p_C = 0.2$ over the course of 2000 time steps. In (a), we can see that after a short chaotic period, the system reaches a pseudo-steady state and the number of true news sharers is fairly constant except for short bursts of disruption when clusters of individuals all shift strategy together. In (b), we get a more detailed look at what happened in the same system by looking at the size of individual connected components. Around $t = 1100$, the single large component of real news sharers splits into two separate components. Then at about $t = 1800$, the two components are joined together as a small cluster between them changes back to sharing real news.

ties, approximately 0.15 to 0.5 in the case of the square lattice with selection strength $\beta = 0.5$. We call behavior in this region the pseudo-steady state because these echo chambers are highly resistant to invasion and thus can persist for millions of time steps. However, it is not a true steady state because with an infinite amount of time, eventually the echo chambers will break down and one strategy will go extinct.

We can see the resilience of these echo chambers by looking at the number of real news sharers as a function of time. Figure A.2a shows the prevalence of real news in a single representative simulation. The number of cooperators drops swiftly at first before stabilizing at around 290 cooperators. There are small shifts at $t \approx 1100$ and

$t \approx 1800$, but otherwise the population is unchanging except for minor perturbations on the border of echo chambers. Fig A.2b gives more detail, showing the size of each path-connected component of real news sharers. By comparing Fig A.2a and b, we see that the changes in cooperator population size corresponds to the large 290-individual echo chamber breaking into two smaller components, one with ≈ 250 individuals and the other with ≈ 20, and then fusing back together.

On the square lattice, the formation of echo chambers and the pseudo-steady state seems to occur across a wide range of fact-checker densities. As shown in the main paper, we also observe echo chamber formation on small-world networks and the twitter network. However, this is not a uniform property of all networks. Preliminary results show that the formation of echo chambers and the critical p_C value are dependent on network topology; lattices and small-worlds are fertile ground for echo chambers, but Erdös-Renyí random graphs and scale-free networks are not. This leads us to hypothesize that a relatively high clustering coefficient is essential for the formation of echo chambers. This intuitively makes sense, as echo chambers are dependent on the feedback loops possible in cliquish, highly connected communities.

Section A.2

Fact-checker Inaccuracy

In reality, fact-checking is subject to human errors. Some fake news occasionally goes unnoticed and endorsed, and some real news is temporally labelled to be fake by well-meaning fact-checkers. When relying on citizen fact-checkers instead of professional journalists for peer policing purposes, the accuracy of fact-checking will inevitably go down as laymen are less prepared to accurately assess fake news. Suppose that fact-checkers have an accuracy in their policing of $\lambda \in [0, 1]$. With probability λ, they correctly assess a post's accuracy and reward benefit α to true news spreaders and

penalty γ to fake news spreaders. With probability $1-\lambda$, an error occurs, leading to the opposite payoff assignments. Using the same method we use to calculate the analytic fixation probabilities, we will quantify the precision threshold required for fact-checkers to ensure fair and transparent policing of wrongdoers while in favor of real news spreaders. For the exact expressions, see the end of the section on analytic derivations below. Figure A.3 shows the relationship between invasion probabilities on the $p_C - \lambda$ plane when using the following payoff matrix:

$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & -4 \\ 0 & 0 & 0 \end{pmatrix} \end{array} \tag{A.1}$$

In Fig A.3a, we see that when $\lambda < 0.5$, selection always favors fake news. This is unsurprising, as it means that the supposed fact-checkers are actually giving more benefit to fake news spreaders than real news spreaders. However, there is a clear buffer in which fact-checkers can be accurate only about 80% of the time without necessitating a drastic increase in the critical fact-checker density for selection to favor real news.

Fig A.3b shows an interesting phenomenon. When fact-checker accuracy is very close to 1/2 and the number of fact-checkers is extremely high, selection actually favors invasion by both real *and* fake news. This is surprising because this real vs fake news game is a coordination game which tends to oppose invading mutants. While this set of parameters is unrealistic and would never appear in any real population, it still demonstrates an interesting property of the dynamics of coordination games in the presence of zealots or extreme environmental conditions.

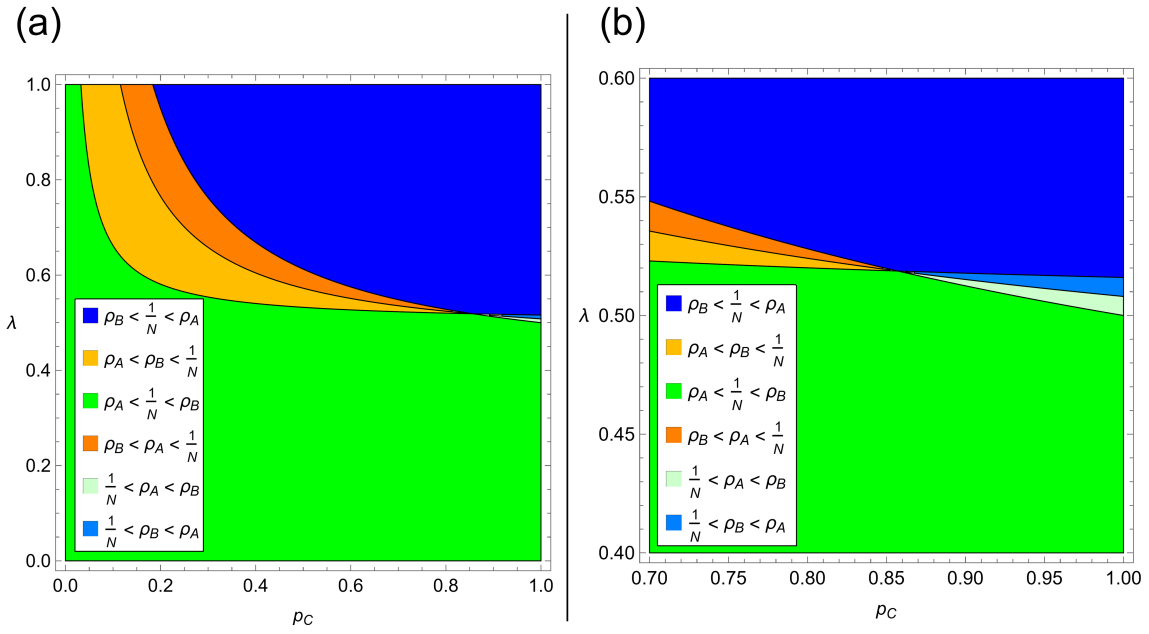

Figure A.3: The results of varying the accuracy of fact-checkers. In (a), we see the where in the $p_C - \lambda$ plane selection favors true news (blue), false news (green), or neither (orange). However, when the density of fact-checkers is very high and fact-checkers are not very accurate, selection can actually favor invasion by true or false news, as shown in (b). This is surprising because this is a coordination game and it is rare for selection to favor invasion by both strategies. However, this combination of parameter values is highly unrealistic and would never occur in real life.

Section A.3

Derivation of Analytic Results

In this section, we derive the invasion probabilities of single cooperators and defectors in the limit of weak selection. We begin by introducing the necessary notation. We have N individuals on a network, each with k neighbors, and they play a game with a general payoff matrix

$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} a & b & \alpha \\ c & d & \gamma \\ 0 & 0 & 0 \end{pmatrix} \end{array} \tag{A.2}$$

p_A, p_B, and p_C are the proportions of A, B, and C players. Similarly, $p_{S_1S_2}$ is the proportion of edges leading from an individual playing S_1 to an individual playing S_2, where S_1 and S_2 can be A, B, or C. We will also be interested in the conditional probability of finding an individual playing S_2 by following a random edge that starts at an individual playing S_1, which will be denoted $q_{S_2|S_1}$. By basic probability, $q_{S_2|S_1} = \frac{p_{S_1S_2}}{p_{S_1}}$.

For an individual playing S_i, π_{S_i} is the total payoff, or the sum of the payoffs from each interaction with a neighbor. The payoff of any A or B individual is dependent on the neighbors' strategies, but we are interested in the expected payoff which only depends on the quantities already listed. With selection strength β, $f_{S_i} = e^{\beta\pi_{S_i}}$ is the fitness of an individual playing S_i.

We have two normalization conditions that ensure that all our probabilities sum to 1:

$$p_A + p_B + p_C = 1 \tag{A.3}$$

$$p_{AA} + p_{AB} + p_{AC} + p_{BA} + p_{BB} + p_{BC} + p_{CA} + p_{CB} + p_{CC} = 1 \tag{A.4}$$

Additionally, there are three symmetry conditions. These need not be true in general, but because the network we are using is undirected, an edge from S_1 to S_2 is also an edge from S_2 to S_1. Therefore:

$$p_{AB} = p_{BA} \tag{A.5}$$

$$p_{AC} = p_{CA} \tag{A.6}$$

$$p_{BC} = p_{CB} \tag{A.7}$$

Finally, we have three consistency conditions:

$$p_A = p_{AA} + p_{AB} + p_{AC} \tag{A.8}$$

$$p_B = p_{BA} + p_{BB} + p_{BC} \tag{A.9}$$

$$p_C = p_{CA} + p_{CB} + p_{CC} \tag{A.10}$$

With all these conditions, we can simplify the system until there are only five independent variables: p_A, p_B, p_{AA}, p_{BB}, p_{CC}. The other four variables can be solved in terms of these five:

$$p_C = 1 - p_A - p_B \tag{A.11}$$

$$p_{AB} = p_{BA} = 1/2\Big[(p_A - p_{AA}) + (p_B - p_{BB}) - (p_C - p_{CC})\Big] \tag{A.12}$$

$$p_{AC} = p_{CA} = 1/2\Big[(p_C - p_{CC}) - (p_B - p_{BB}) + (p_A - p_{AA})\Big] \tag{A.13}$$

$$p_{BC} = p_{CB} = 1/2\Big[(p_B - p_{BB}) + (p_C - p_{CC}) - (p_A - p_{AA})\Big] \tag{A.14}$$

Now we are ready to derive differential equations for the systems evolution in time.

A.3.1. Pair Approximation

The game between real and fake news is a coordination game, and because of this, individuals will tend to form clusters of like-minded individuals, as observed in simulations. However, because of this, the probabilities along two successive edges are not independent. That is to say, if $p_{S_1S_2S_3}$ is the probability of starting at an S_1 player, following a random edge to an S_2 player, and then following another random edge to an S_3 player, we do **not** get that

$$p_{S_1S_2S_3} = \frac{p_{S_1S_2}p_{S_2S_3}}{p_{S_2}} \tag{A.15}$$

However, this makes studying the system untenable. Pair approximation alleviates this problem by making the simplifying assumptions that edges are independent and therefore Equation (A.15) holds.

In the death-birth process, an individual is chosen to "die" and a neighbor is chosen to replicate and take the deceased individuals place. However, if the two individuals are playing the same strategy, nothing in the population will have changed. The only way the system changes is if an A individual takes the place of a B individual or vice versa, so we focus on the frequency of these two events to study the system.

We use the modified update step where only one individual is replaced per time step. This slows down the system's evolution by a factor of $\frac{1}{N}$, but it has very little effect on the behavior of the system, and it makes the system much easier to approach analytically. With a discrete time step $\Delta t = \frac{1}{N}$ so that one individual is replaced per time step, the differential equations for p_A and p_{AA} are:

$$\dot{p_A} = \frac{1}{N}\frac{E(\Delta n_A)}{\Delta t} = E(\Delta n_A) \tag{A.16}$$

$$\dot{p_{AA}} = \frac{2}{kN}\frac{E(\Delta n_{AA})}{\Delta t} = \frac{2}{k}E(\Delta n_{AA}) \tag{A.17}$$

We first focus on computing $E(\Delta n_A)$. Because only one individual updates at a time, $E(\Delta n_A) = P(\Delta n_A = 1) - P(\Delta n_A = -1)$. n_A increases by one when a B player is replaced by an A player, and n_A decreases by one when an A player is replaced by a B player. We now derive the probability of an A player replacing a B player. The probability of B invading A follows by symmetry.

The B player that is being replaced has k neighbors, each of which can be an A, B, or C player. Specifically, the focal B player has k_B^A A neighbors, k_B^B B neighbors, and k_B^C C neighbors with probability

$$\frac{k!}{k_B^A! k_B^B! k_B^C!} q_{A|B}^{k_B^A} q_{B|B}^{k_B^B} q_{C|B}^{k_B^C} \tag{A.18}$$

and there is always the restriction that $k_B^A + k_B^B + k_B^C = k$.

Each of these neighbors has $k-1$ neighbors (not including the focal B player) that are also multinomially distributed. An A-playing neighbor will have k'^A_A A neighbors, k'^B_A B neighbors, and k'^C_A C neighbors with probability

$$\frac{(k-1)!}{k'^A_A! k'^B_A! k'^C_A!} q_{A|A}^{k'^A_A} q_{B|A}^{k'^B_A} q_{C|A}^{k'^C_A} \tag{A.19}$$

Here we used pair approximation, because we ignore the higher-order terms that might arise knowing that the A player already has a B neighbor.

Likewise, the B and C players neighboring the focal B player have neighbors whose strategies are multinomially distributed. To determine the strategy the focal B player will choose to imitate, we need to know the payoffs of all of the neighbors.

An A neighbor of the focal B player who has k'^A_A A neighbors, k'^B_A B neighbors (not including the focal B player), and k'^C_A C neighbors has payoff

$$\pi_A = k'^A_A a + (k'^B_A + 1)b + k'^C_A \alpha \tag{A.20}$$

and fitness

$$f_A(k'^A_A, k'^B_A, k'^C_A) = e^{\beta\pi_A} \tag{A.21}$$

The same quantities for the B and C neighbors work the same way.

$$\pi_B = k'^A_B c + (k'^B_B + 1)d + k'^C_B \gamma \tag{A.22}$$

$$f_B(k'^A_B, k'^B_B, k'^C_B) = e^{\beta\pi_B} \tag{A.23}$$

$$\pi_C = k'^A_C 0 + (k'^B_C + 1)0 + k'^C_C 0 = 0 \tag{A.24}$$

$$f_C(k'^A_C, k'^B_C, k'^C_C) = e^{\beta\pi_C} = 1 \tag{A.25}$$

We are interested in the focal B player being replaced by an A player. Because individuals choose who to copy proportional to fitness, the probability of the B player selecting one of its A neighbors is

$$\frac{k^A_B f_A}{k^A_B f_A + k^B_B f_B + k^C_B f_C} \tag{A.26}$$

All that remains is to sum over all possible configurations of the B player's neighbors and their neighbors and multiply by p_B (the probability that a B player is selected to update) to get the final probability W_{AB} that a B player is replaced by an A player:

$$
\begin{aligned}
W_{AB} = p_B \cdot & \sum_{k_B^A+k_B^B+k_B^C=k} \frac{k!}{k_B^A!k_B^B!k_B^C!} q_{A|B}^{k_B^A} q_{B|B}^{k_B^B} q_{C|B}^{k_B^C} \\
& \cdot \sum_{k'^A_A+k'^B_A+k'^C_A=k-1} \frac{(k-1)!}{k'^A_A!k'^B_A!k'^C_A!} q_{A|A}^{k'^A_A} q_{B|A}^{k'^B_A} q_{C|A}^{k'^C_A} \\
& \cdot \sum_{k'^A_B+k'^B_B+k'^C_B=k-1} \frac{(k-1)!}{k'^A_B!k'^B_B!k'^C_B!} q_{A|B}^{k'^A_B} q_{B|B}^{k'^B_B} q_{C|B}^{k'^C_B} \\
& \cdot \sum_{k'^A_C+k'^B_C+k'^C_C=k-1} \frac{(k-1)!}{k'^A_C!k'^B_C!k'^C_C!} q_{A|C}^{k'^A_C} q_{B|C}^{k'^B_C} q_{C|C}^{k'^C_C} \\
& \cdot \frac{k_B^A f_A(k'^A_A, k'^B_A+1, k'^C_A)}{k_B^A f_A(k'^A_A, k'^B_A+1, k'^C_A) + k_B^B f_B(k'^A_B, k'^B_B+1, k'^C_B) + k_B^C f_C(k'^A_C, k'^B_C+1, k'^C_C)}
\end{aligned}
\tag{A.27}
$$

(Though it is difficult to typeset within the margins, note that this is a nested sum and not the product of four separate sums.) Likewise, W_{BA}, the probability of B invading A, is

$$
\begin{aligned}
W_{BA} = p_A \cdot & \sum_{k_A^A+k_A^B+k_A^C=k} \frac{k!}{k_A^A!k_A^B!k_A^C!} q_{A|A}^{k_A^A} q_{B|A}^{k_A^B} q_{C|A}^{k_A^C} \\
& \cdot \sum_{k'^A_A+k'^B_A+k'^C_A=k-1} \frac{(k-1)!}{k'^A_A!k'^B_A!k'^C_A!} q_{A|A}^{k'^A_A} q_{B|A}^{k'^B_A} q_{C|A}^{k'^C_A} \\
& \cdot \sum_{k'^A_B+k'^B_B+k'^C_B=k-1} \frac{(k-1)!}{k'^A_B!k'^B_B!k'^C_B!} q_{A|B}^{k'^A_B} q_{B|B}^{k'^B_B} q_{C|B}^{k'^C_B} \\
& \cdot \sum_{k'^A_C+k'^B_C+k'^C_C=k-1} \frac{(k-1)!}{k'^A_C!k'^B_C!k'^C_C!} q_{A|C}^{k'^A_C} q_{B|C}^{k'^B_C} q_{C|C}^{k'^C_C} \\
& \cdot \frac{k_A^B f_B(k'^A_B+1, k'^B_B, k'^C_B)}{k_A^A f_A(k'^A_A+1, k'^B_A, k'^C_A) + k_A^B f_B(k'^A_B+1, k'^B_B, k'^C_B) + k_A^C f_C(k'^A_C+1, k'^B_C, k'^C_C)}
\end{aligned}
\tag{A.28}
$$

Furthermore, when B is invaded by A it increases the number of $A-A$ pairs by k_B^A, so we can define ϕ_{AB}^A to be the expected value for the change in $A-A$ edges due

to a B player being invaded by an A player. (The subscript describes the direction of invasion and the superscript determines which pair it corresponds to, so ϕ^A_{AB} means an A player is replacing a B player, and this term tells us about the change in $A-A$ pairs.) Like in (A.27), we have

$$\begin{aligned}\phi^A_{AB} = p_B \cdot & \sum_{k^A_B+k^B_B+k^C_B=k} k^A_B \frac{k!}{k^A_B!k^B_B!k^C_B!} q_{A|B}^{k^A_B} q_{B|B}^{k^B_B} q_{C|B}^{k^C_B} \\ \cdot & \sum_{k'^A_A+k'^B_A+k'^C_A=k-1} \frac{(k-1)!}{k'^A_A!k'^B_A!k'^C_A!} q_{A|A}^{k'^A_A} q_{B|A}^{k'^B_A} q_{C|A}^{k'^C_A} \\ \cdot & \sum_{k'^A_B+k'^B_B+k'^C_B=k-1} \frac{(k-1)!}{k'^A_B!k'^B_B!k'^C_B!} q_{A|B}^{k'^A_B} q_{B|B}^{k'^B_B} q_{C|B}^{k'^C_B} \\ \cdot & \sum_{k'^A_C+k'^B_C+k'^C_C=k-1} \frac{(k-1)!}{k'^A_C!k'^B_C!k'^C_C!} q_{A|C}^{k'^A_C} q_{B|C}^{k'^B_C} q_{C|C}^{k'^C_C} \\ \cdot & \frac{k^A_B f_A(k'^A_A, k'^B_A+1, k'^C_A)}{k^A_B f_A(k'^A_A, k'^B_A+1, k'^C_A) + k^B_B f_B(k'^A_B, k'^B_B+1, k'^C_B) + k^C_B f_C(k'^A_C, k'^B_C+1, k'^C_C)}\end{aligned} \tag{A.29}$$

Note that (A.29) only differs from (A.27) in a single k^A_B term in the first line, which is there because we are interested in the expected value of the change in $A-A$ edges, and there are k^A_B new $A-A$ edges being formed. Similarly, we can write down:

$$\begin{aligned}\phi^A_{BA} = p_A \cdot & \sum_{k^A_A+k^B_A+k^C_A=k} k^A_A \frac{k!}{k^A_A!k^B_A!k^C_A!} q^{k^A_A}_{A|A} q^{k^B_A}_{B|A} q^{k^C_A}_{C|A} \\ \cdot & \sum_{k'^A_A+k'^B_A+k'^C_A=k-1} \frac{(k-1)!}{k'^A_A!k'^B_A!k'^C_A!} q^{k'^A_A}_{A|A} q^{k'^B_A}_{B|A} q^{k'^C_A}_{C|A} \\ \cdot & \sum_{k'^A_B+k'^B_B+k'^C_B=k-1} \frac{(k-1)!}{k'^A_B!k'^B_B!k'^C_B!} q^{k'^A_B}_{A|B} q^{k'^B_B}_{B|B} q^{k'^C_B}_{C|B} \\ \cdot & \sum_{k'^A_C+k'^B_C+k'^C_C=k-1} \frac{(k-1)!}{k'^A_C!k'^B_C!k'^C_C!} q^{k'^A_C}_{A|C} q^{k'^B_C}_{B|C} q^{k'^C_C}_{C|C} \\ \cdot & \frac{k^B_A f_B(k'^A_B+1, k'^B_B, k'^C_B)}{k^A_A f_A(k'^A_A+1, k'^B_A, k'^C_A) + k^B_A f_B(k'^A_B+1, k'^B_B, k'^C_B) + k^C_A f_C(k'^A_C+1, k'^B_C, k'^C_C)}\end{aligned} \tag{A.30}$$

$$\begin{aligned}\phi^B_{AB} = p_B \cdot & \sum_{k^A_B+k^B_B+k^C_B=k} k^B_B \frac{k!}{k^A_B!k^B_B!k^C_B!} q^{k^A_B}_{A|B} q^{k^B_B}_{B|B} q^{k^C_B}_{C|B} \\ \cdot & \sum_{k'^A_A+k'^B_A+k'^C_A=k-1} \frac{(k-1)!}{k'^A_A!k'^B_A!k'^C_A!} q^{k'^A_A}_{A|A} q^{k'^B_A}_{B|A} q^{k'^C_A}_{C|A} \\ \cdot & \sum_{k'^A_B+k'^B_B+k'^C_B=k-1} \frac{(k-1)!}{k'^A_B!k'^B_B!k'^C_B!} q^{k'^A_B}_{A|B} q^{k'^B_B}_{B|B} q^{k'^C_B}_{C|B} \\ \cdot & \sum_{k'^A_C+k'^B_C+k'^C_C=k-1} \frac{(k-1)!}{k'^A_C!k'^B_C!k'^C_C!} q^{k'^A_C}_{A|C} q^{k'^B_C}_{B|C} q^{k'^C_C}_{C|C} \\ \cdot & \frac{k^A_B f_A(k'^A_A, k'^B_A+1, k'^C_A)}{k^A_B f_A(k'^A_A, k'^B_A+1, k'^C_A) + k^B_B f_B(k'^A_B, k'^B_B+1, k'^C_B) + k^C_B f_C(k'^A_C, k'^B_C+1, k'^C_C)}\end{aligned} \tag{A.31}$$

$$\begin{aligned}\phi_{BA}^{B} = p_A \cdot & \sum_{k_A^A+k_A^B+k_A^C=k} k_A^B \frac{k!}{k_A^A!k_A^B!k_A^C!} q_{A|A}^{k_A^A} q_{B|A}^{k_A^B} q_{C|A}^{k_A^C} \\ \cdot & \sum_{k'^A_A+k'^B_A+k'^C_A=k-1} \frac{(k-1)!}{k'^A_A!k'^B_A!k'^C_A!} q_{A|A}^{k'^A_A} q_{B|A}^{k'^B_A} q_{C|A}^{k'^C_A} \\ \cdot & \sum_{k'^A_B+k'^B_B+k'^C_B=k-1} \frac{(k-1)!}{k'^A_B!k'^B_B!k'^C_B!} q_{A|B}^{k'^A_B} q_{B|B}^{k'^B_B} q_{C|B}^{k'^C_B} \\ \cdot & \sum_{k'^A_C+k'^B_C+k'^C_C=k-1} \frac{(k-1)!}{k'^A_C!k'^B_C!k'^C_C!} q_{A|C}^{k'^A_C} q_{B|C}^{k'^B_C} q_{C|C}^{k'^C_C} \\ \cdot & \frac{k_A^B f_B(k'^A_B+1, k'^B_B, k'^C_B)}{k_A^A f_A(k'^A_A+1, k'^B_A, k'^C_A) + k_A^B f_B(k'^A_B+1, k'^B_B, k'^C_B) + k_A^C f_C(k'^A_C+1, k'^B_C, k'^C_C)}\end{aligned} \tag{A.32}$$

Once we have these quantities (Equations (A.27) - (A.32)), we have expressions for all of our independent variables.

$$\dot{p_{CC}} = 0 \tag{A.33}$$

$$\dot{p_A} = -\dot{p_B} = W_{AB} - W_{BA} \tag{A.34}$$

$$\dot{p_{AA}} = \frac{2}{k}(\phi_{AB}^{A} - \phi_{BA}^{A}) \tag{A.35}$$

$$\dot{p_{BB}} = \frac{2}{k}(\phi_{BA}^{B} - \phi_{AB}^{B}) \tag{A.36}$$

A.3.2. Weak Selection

Even with the substantial simplification from pair approximation, the previous results are too complicated and unwieldy to be useful by themselves. Because of compounding sums, directly calculating the derivatives requires adding millions of terms if $k = 8$. Furthermore, the pair approximation means that we lose the information critical to clustering, and therefore the analytic results here will fail to capture the

pseudo-steady states that we observe when β is much larger than zero.

We can sidestep both these issues by working in the limit of weak selection. In weak selection, the success or failure of an individual in the fake news game is only one small factor in the individual's success, and fitnesses are much more uniform across the population. When β is close to zero, we can throw out higher order terms which simplifies the expression, and when β is close to zero, the pseudo-steady states cannot exist anyways because the system behaves approximately like neutral drift. Taking the Taylor expansion of the exponential in equations (A.21) and (A.23) with respect to β and only keeping the low order terms, what is left is mathematically tractable. We have expressions for each of $W_{AB}, W_{BA}, \phi^A_{AB}, \phi^A_{BA}, \phi^B_{BA}, \phi^B_{AB}$. We manipulate each separately and bring them back together at the end.

A.3.3. W_{AB} and W_{BA}:

Equation (A.27) gives us an expression for W_{AB}. The fact-checkers playing C have constant fitness, $f_C = 1$, and no other terms in the last line of (A.27) depend on the neighbors of C players, so we can pull it all through the final sum which collapses to 1 because it is the sum of the probabilities of all possible configurations of neighbors, which must be 1. Therefore,

$$\begin{aligned} W_{AB} = p_B \cdot & \sum_{k_B^A + k_B^B + k_B^C = k} \frac{k!}{k_B^A! k_B^B! k_B^C!} q_{A|B}^{k_B^A} q_{B|B}^{k_B^B} q_{C|B}^{k_B^C} \\ & \cdot \sum_{k_A^{A'} + k_A^{B'} + k_A^{C'} = k-1} \frac{(k-1)!}{k_A^{A'}! k_A^{B'}! k_A^{C'}!} q_{A|A}^{k_A^{A'}} q_{B|A}^{k_A^{B'}} q_{C|A}^{k_A^{C'}} \\ & \cdot \sum_{k_B^{A'} + k_B^{B'} + k_B^{C'} = k-1} \frac{(k-1)!}{k_B^{A'}! k_B^{B'}! k_B^{C'}!} q_{A|B}^{k_B^{A'}} q_{B|B}^{k_B^{B'}} q_{C|B}^{k_B^{C'}} \\ & \cdot \frac{k_B^A f_A(k_A^{A'}, k_A^{B'} + 1, k_A^{C'})}{k_B^A f_A(k_A^{A'}, k_A^{B'} + 1, k_A^{C'}) + k_B^B f_B(k_B^{A'}, k_B^{B'} + 1, k_B^{C'}) + k_B^C} \end{aligned} \tag{A.37}$$

Then, using the Taylor expansion for the exponentials in f_A and f_B but only keeping the low order terms of β, we have

$$\begin{aligned}
&\frac{k_B^A f_A(k_A^{A'}, k_A^{B'}+1, k_A^{C'})}{k_B^A f_A(k_A^{A'}, k_A^{B'}+1, k_A^{C'}) + k_B^B f_B(k_B^{A'}, k_B^{B'}+1, k_B^{C'}) + k_B^C} \\
&\approx \frac{k_B^A\big(1+\beta(ak_A^{A'}+b(k_A^{B'}+1)+ck_A^{C'})\big)}{k_B^A\big(1+\beta(ak_A^{A'}+b(k_A^{B'}+1)+\alpha k_A^{C'})\big) \\
&\qquad + k_B^B\big(1+\beta(ck_B^{A'}+d(k_B^{B'}+1)+\gamma k_B^{C'})\big) + k_B^C} \\
&= \frac{k_B^A(1+\beta u_1)}{k+\beta(k_B^A u_1 + k_B^B u_2)} \\
&\approx k_B^A(1+\beta u_1)[\frac{1}{k} - \frac{k_B^A u_1 + k_B^B u_2}{k^2}\beta] \\
&\approx \frac{k_B^A}{k} + \beta[\frac{k_B^A u_1}{k} - k_B^A\frac{k_B^A u_1 + k_B^B u_2}{k^2}]
\end{aligned} \tag{A.38}$$

where $u_1 = ak_A^{A'} + b(k_A^{B'}+1) + \alpha k_A^{C'}$ and $u_2 = ck_B^{A'} + d(k_B^{B'}+1) + \gamma k_B^{C'}$. By carefully pulling terms through the sums, we have the following identities:

$$\begin{aligned}
&\sum_{k_A^{A'}+k_A^{B'}+k_A^{C'}=k-1} \frac{(k-1)!}{k_A^{A'}!k_A^{B'}!k_A^{C'}!} q_{A|A}^{k_A^{A'}} q_{B|A}^{k_A^{B'}} q_{C|A}^{k_A^{C'}} u_1 \\
&\qquad = a(k-1)q_{A|A} + b\big((k-1)q_{B|A}+1\big) + \alpha(k-1)q_{C|A} \\
&\qquad = E_A + b
\end{aligned} \tag{A.39}$$

$$\begin{aligned}
&\sum_{k_B^{A'}+k_B^{B'}+k_B^{C'}=k-1} \frac{(k-1)!}{k_B^{A'}!k_B^{B'}!k_B^{C'}!} q_{A|B}^{k_B^{A'}} q_{B|B}^{k_B^{B'}} q_{C|B}^{k_B^{C'}} u_2 \\
&\qquad = c(k-1)q_{A|B} + d\big((k-1)q_{B|B}+1\big) + \gamma(k-1)q_{C|B} \\
&\qquad = E_B + d
\end{aligned} \tag{A.40}$$

Notice that E_A and E_B are the expected payoffs for A and B players from $k-1$

neighbors. Using these identities on our equation for W_{AB}, we get that

$$\begin{aligned} W_{AB} = p_B \cdot \sum_{k_B^A+k_B^B+k_B^C=k} \frac{k!}{k_B^A!k_B^B!k_B^C!} q_{A|B}^{k_B^A} q_{B|B}^{k_B^B} q_{C|B}^{k_B^C} \\ \cdot \left[\frac{k_B^A}{k} - \beta \frac{k_B^B k_B^A}{k^2}(E_B+d) + \beta \frac{k_B^A}{k}(E_A+b) - \beta \frac{{k_B^A}^2}{k^2}(E_A+b) \right] \end{aligned} \tag{A.41}$$

Each of these four terms in the brackets can be dealt with separately in similar fashion:

$$\sum_{k_B^A+k_B^B+k_B^C=k} \frac{k!}{k_B^A!k_B^B!k_B^C!} q_{A|B}^{k_B^A} q_{B|B}^{k_B^B} q_{C|B}^{k_B^C} \left[\frac{k_B^A}{k} \right] = q_{A|B} \tag{A.42}$$

$$\begin{aligned} \sum_{k_B^A+k_B^B+k_B^C=k} \frac{k!}{k_B^A!k_B^B!k_B^C!} q_{A|B}^{k_B^A} q_{B|B}^{k_B^B} q_{C|B}^{k_B^C} \left[-\beta \frac{k_B^B k_B^A}{k^2}(E_B+d) \right] \\ = -\beta \frac{(E_B+d)}{k^2} k(k-1) q_{A|B} q_{B|B} \end{aligned} \tag{A.43}$$

$$\sum_{k_B^A+k_B^B+k_B^C=k} \frac{k!}{k_B^A!k_B^B!k_B^C!} q_{A|B}^{k_B^A} q_{B|B}^{k_B^B} q_{C|B}^{k_B^C} \left[\beta \frac{k_B^A}{k}(E_A+b) \right] = \beta(E_A+b) q_{A|B} \tag{A.44}$$

$$\begin{aligned} \sum_{k_B^A+k_B^B+k_B^C=k} \frac{k!}{k_B^A!k_B^B!k_B^C!} q_{A|B}^{k_B^A} q_{B|B}^{k_B^B} q_{C|B}^{k_B^C} \left[-\beta \frac{{k_B^A}^2}{k^2}(E_A+b) \right] \\ = -\beta \frac{(E_A+b)}{k^2} k q_{A|B} [(k-1) q_{A|B} + 1] \end{aligned} \tag{A.45}$$

Therefore,

$$\begin{aligned} W_{AB} = p_B \Bigg[q_{A|B} + \beta \Big((E_A+b) q_{A|B} - \frac{E_A+b}{k} q_{A|B} \\ - \frac{k-1}{k} q_{A|B} \big[(E_B+d) q_{B|B} + (E_A+b) q_{A|B} \big] \Big) \Bigg] + \mathcal{O}(\beta^2) \end{aligned} \tag{A.46}$$

Using the same techniques, we can simplify our expression for W_{BA}:

$$W_{BA} = p_A\Bigg[q_{B|A} + \beta\Big((E_B + c)q_{B|A} - \frac{E_B + c}{k}q_{B|A} \\ - \frac{k-1}{k}q_{B|A}\big[(E_B + c)q_{B|A} + (E_A + a)q_{A|A}\big]\Big)\Bigg] + \mathcal{O}(\beta^2) \tag{A.47}$$

Note immediately that since $p_B q_{A|B} = p_A q_{B|A}$, the zero-th order terms of W_{AB} and W_{BA} are equal.

A.3.4. The ϕs:

The pair derivatives are non-zero, even when $\beta = 0$, so we will focus only on the zeroth order terms, because these will dominate the first-order terms when β is small.

$$\begin{aligned}\phi^A_{AB} = p_B \cdot & \sum_{k^A_B + k^B_B + k^C_B = k} k^A_B \frac{k!}{k^A_B! k^B_B! k^C_B!} q_{A|B}^{k^A_B} q_{B|B}^{k^B_B} q_{C|B}^{k^C_B} \\ \cdot & \sum_{k^{A'}_A + k^{B'}_A + k^{C'}_A = k-1} \frac{(k-1)!}{k^{A'}_A! k^{B'}_A! k^{C'}_A!} q_{A|A}^{k^{A'}_A} q_{B|A}^{k^{B'}_A} q_{C|A}^{k^{C'}_A} \\ \cdot & \sum_{k^{A'}_B + k^{B'}_B + k^{C'}_B = k-1} \frac{(k-1)!}{k^{A'}_B! k^{B'}_B! k^{C'}_B!} q_{A|B}^{k^{A'}_B} q_{B|B}^{k^{B'}_B} q_{C|B}^{k^{C'}_B} \\ \cdot & \sum_{k^{A'}_C + k^{B'}_C + k^{C'}_C = k-1} \frac{(k-1)!}{k^{A'}_C! k^{B'}_C! k^{C'}_C!} q_{A|C}^{k^{A'}_C} q_{B|C}^{k^{B'}_C} q_{C|C}^{k^{C'}_C} \\ \cdot & \frac{k^A_B f_A(k^{A'}_A, k^{B'}_A + 1, k^{C'}_A)}{k^A_B f_A(k^{A'}_A, k^{B'}_A + 1, k^{C'}_A) + k^B_B f_B(k^{A'}_B, k^{B'}_B + 1, k^{C'}_B) + k^C_B f_C(k^{A'}_C, k^{B'}_C + 1, k^{C'}_C)}\end{aligned} \tag{A.48}$$

The zeroth order terms are what is left when $\beta = 0$, or when we have neutral drift. In that case, $f_A = f_B = f_C = 1$, and most of the sums collapse to 1. We quickly get that

$$\phi^A_{AB} = \frac{p_B}{k} \sum_{k^A_B + k^B_B + k^C_B = k} \frac{k!}{k^A_B! k^B_B! k^C_B!} q_{A|B}^{k^A_B} q_{B|B}^{k^B_B} q_{C|B}^{k^C_B} {k^A_B}^2 \tag{A.49}$$

We relabel for notational convenience and readability when evaluating this sum. Let $X = k_B^A, Y = k_B^B, Z = k_B^C$. Then the sum is

$$\begin{aligned}
&\sum_{X+Y+Z=k} \frac{k!}{X!Y!Z!} q_{A|B}^X q_{B|B}^Y q_{C|B}^Z X^2 \\
=&kq_{A|B} \sum_{(X-1)+Y+Z=k-1} \frac{(k-1)!}{(X-1)!Y!Z!} q_{A|B}^{X-1} q_{B|B}^Y q_{C|B}^Z (X) \\
=&kq_{A|B} \sum_{(X-1)+Y+Z=k-1} \frac{(k-1)!}{(X-1)!Y!Z!} q_{A|B}^{X-1} q_{B|B}^Y q_{C|B}^Z (X-1) \\
&+ kq_{A|B} \sum_{(X-1)+Y+Z=k-1} \frac{(k-1)!}{(X-1)!Y!Z!} q_{A|B}^{X-1} q_{B|B}^Y q_{C|B}^Z \\
=&kq_{A|B}\Big((k-1)q_{A|B} \sum_{(X-2)+Y+Z=k-2} \frac{(k-2)!}{(X-2)!Y!Z!} q_{A|B}^{X-2} q_{B|B}^Y q_{C|B}^Z + 1\Big) \\
=&kq_{A|B}\Big((k-1)q_{A|B} + 1\Big)
\end{aligned} \tag{A.50}$$

Immediately, we get,

$$\phi_{AB}^A = p_B q_{A|B}\Big((k-1)q_{A|B} + 1\Big) + \mathcal{O}(\beta) \tag{A.51}$$

The other ϕ terms are calculated in the same way. They are:

$$\phi_{BA}^A = p_A(k-1)q_{A|A}q_{B|A} + \mathcal{O}(\beta) \tag{A.52}$$

$$\phi_{BA}^B = p_A q_{B|A}\Big((k-1)q_{B|A} + 1\Big) + \mathcal{O}(\beta) \tag{A.53}$$

$$\phi_{AB}^B = p_B(k-1)q_{B|B}q_{A|B} + \mathcal{O}(\beta) \tag{A.54}$$

A.3.5. The Slow Manifold

With these simplified equations, we can solve the system. Consider the zero-th order terms, setting $\beta = 0$. $W_{AB} = W_{BA}$, so $\dot{p_A} = \dot{p_B} = \dot{p_C}$. Now we address $\dot{p_{AA}}$, $\dot{p_{BB}}$, and $\dot{p_{AB}}$:

With the above derivatives and (A.12), we get that

$$\dot{p_{AB}} = -\frac{1}{2}(\dot{p_{AA}} + \dot{p_{AA}}) \tag{A.55}$$

By substituting (A.51) and (A.52) into (A.35):

$$\begin{aligned}
\dot{p_{AA}} &= \frac{2}{k}\Big[\phi^A_{AB} - \phi^A_{BA}\Big] \\
&= \frac{2}{k}\Big[p_B q_{A|B}\Big((k-1)q_{A|B}+1\Big) - p_A(k-1)q_{A|A}q_{B|A}\Big] \\
&= \frac{2}{k}\Big[p_B q_{A|B}q_{A|B}(k-1) - p_A q_{A|A}q_{B|A}(k-1) + p_B q_{B|A}\Big] \\
&= \frac{2}{k}\Big[\frac{p^2_{AB}}{p_B}(k-1) - \frac{p_{AA}p_{AB}}{p_A}(k-1) + p_{AB}\Big]
\end{aligned} \tag{A.56}$$

Similarly, with (A.53) and (A.54) in (A.36):

$$\begin{aligned}
\dot{p_{BB}} &= \frac{2}{k}\Big[\phi^B_{BA} - \phi^B_{AB}\Big] \\
&= \frac{2}{k}\Big[p_A q_{B|A}\Big((k-1)q_{B|A}+1\Big) - p_B(k-1)q_{B|B}q_{A|B}\Big] \\
&= \frac{2}{k}\Big[p_A q_{B|A}q_{B|A}(k-1) - p_B q_{B|B}q_{A|B}(k-1) + p_A q_{B|A}\Big] \\
&= \frac{2}{k}\Big[\frac{p^2_{AB}}{p_A}(k-1) - \frac{p_{AB}p_{BB}}{p_B}(k-1) + p_{AB}\Big]
\end{aligned} \tag{A.57}$$

Now subtract (A.57) from (A.56):

$$\begin{aligned}
\dot{p_{AA}} - \dot{p_{BB}} =& \frac{2}{k}\Big[\frac{p^2_{AB}}{p_B}(k-1) - \frac{p_{AA}p_{AB}}{p_A}(k-1) + p_{AB}\Big] \\
&- \frac{2}{k}\Big[\frac{p^2_{AB}}{p_A}(k-1) - \frac{p_{AB}p_{BB}}{p_B}(k-1) + p_{AB}\Big] \\
=& \frac{2(k-1)}{k}\Big[\frac{p^2_{AB}}{p_B} - \frac{p_{AA}p_{AB}}{p_A} - \frac{p^2_{AB}}{p_A} + \frac{p_{AB}p_{BB}}{p_B}\Big]
\end{aligned} \tag{A.58}$$

When the system is initialized at $t = 0$, it is well-mixed and $p_{S_1S_2}(0) = p_{S_1}(0)p_{S_2}(0)$ for all strategies S_1 and S_2. Thus, at $t = 0$, by equation (A.58), $\dot{p_{AA}} - \dot{p_{BB}} = 0$. And

together with (A.55), we have

$$\dot{p_{AA}} = \dot{p_{BB}} = -\dot{p_{AB}} \tag{A.59}$$

In fact, this will hold for all time steps, because as long as it holds, it will continue to hold. A sketch of a formal proos is as follows: solve the system with Euler's method and take the limit as the discrete time step goes to zero. By the convergence of Euler's method, (A.59) holds for all t.

From this, (A.13) and (A.14) show that $\dot{p_{AC}} = \dot{p_{BC}} = 0$. Then,

$$\dot{q_{C|A}} = \frac{d}{dt}\frac{p_{AC}}{p_A} = \frac{\dot{p_{AC}}p_A - p_{AC}\dot{p_A}}{p_A^2} = 0 \tag{A.60}$$

Similarly, $\dot{q_{C|B}} = 0$. These results are expected because in neutral drift, the fact-checkers do not give either strategy an advantage, so fact-checkers will not naturally attract A players or repel B players.

Because β is very small, the zero-th order terms in $\dot{p}_{AA}$ and $\dot{p}_{BB}$ will go to zero much quicker than the first order terms in $\dot{p}_A$ and $\dot{p}_B$. Set $\dot{p}_{AA} = 0$:

$$\dot{p}_{AA} = \frac{2}{k}\Big[p_B q_{A|B}\Big((k-1)q_{A|B} + 1\Big) - p_A(k-1)q_{A|A}q_{B|A}\Big] = 0 \tag{A.61}$$

Rearranging and dividing by $\frac{2p_{AB}}{k}$ gives

$$(k-1)q_{A|B} + 1 = (k-1)q_{A|A} \tag{A.62}$$

Now use the identities $q_{A|B} = \frac{p_A}{p_B}q_{B|A}$ and $q_{B|A} = 1 - q_{A|A} - p_C$ and rearrange to get

$$q_{A|A} = p_A + \frac{p_B}{(k-1)(1-p_C)} \tag{A.63}$$

A similar procedure with $\dot{p}_{BB} = 0$ yields

$$q_{B|B} = p_B + \frac{p_A}{(k-1)(1-p_C)} \tag{A.64}$$

These conditions define the slow manifold, where the system changes slowly due to β being close to zero. The system may start as a well-mixed population, but it will very quickly approach a state where the above conditions hold, at least approximately. Notice that the slow manifold is one-dimensional; everything can be expressed in terms of p_A, because $p_B = 1 - p_C - p_A$, and p_C will be a constant.

A.3.6. Fixation Probabilities

Consider a system starting with $p_A(0) = p$ and a small time step Δt in which we assume one death-birth occurs. Renormalize with $p_{A_{\text{new}}} = p_{A_{\text{old}}}/(1-p_C)$ and $p_{B_{\text{new}}} = p_{B_{\text{old}}}/(1-p_C)$ so that p_A and p_B are between 0 and 1. Now p_A and p_B represent the proportion of individuals playing A or B out of all the individuals that are capable of changing their strategy (the A and B players). There is a mean $m_A(p)$ and variance $v_A(p)$ of Δp_A for a single time step. We have

$$m_A(p) = E(\Delta p_A) = \frac{1}{N}[W_{AB} - W_{BA}] = [W_{AB} - W_{BA}]\Delta t \tag{A.65}$$

$$\begin{aligned} v_A(p) &= E(\Delta p_A^2) - E(\Delta p_A)^2 = E(\Delta p_A^2) + \mathcal{O}(\beta^2) \\ &\approx \frac{1}{N^2}[W_{AB} + W_{BA}] = \frac{1}{N}[W_{AB} + W_{BA}]\Delta t \end{aligned} \tag{A.66}$$

The relevant value will be $-\frac{2m_A(p)}{v_A(p)}$, which can be obtained by substituting in the constraints of the slow manifold: Equations (A.63) and (A.64). After substituting in the expressions for W_{AB} and W_{BA}, simplifying gets us:

$$-\frac{2m_A(p)}{v_A(p)} = \frac{\beta N}{k}(u_1 p + u_2) \tag{A.67}$$

where

$$u_1 = (a - b - c + d)(1 - k^2 - \frac{1+k}{p_C - 1})(1 - p_C) \tag{A.68}$$

$$u_2 = -a+b+c-d-ak+bk-bk^2+dk^2+(k-1)\big(c+(b-\alpha+\gamma)k-d(1+k)\big)p_C \tag{A.69}$$

According to diffusion theory, the fixation probability of A beginning with $p_A(0) = p$, denoted $\rho_A(p)$, satisfies the equation

$$m_A(p)\frac{d\rho_A(p)}{dp} + \frac{v_A(p)}{2}\frac{d^2\rho_A(p)}{dp^2} = 0 \tag{A.70}$$

This equation is separable and first order with respect to $\frac{d\rho_A(p)}{dp}$.

$$\ln \frac{d\rho_A(p)}{dp} = \int -\frac{2m_A(p)}{v_A(p)}dp \tag{A.71}$$

The low order terms are

$$\frac{d\rho_A(p)}{dp} = 1 + \frac{\beta N}{k}\Big(\frac{u_1}{2}p^2 + u_2 p\Big) + c_1 \tag{A.72}$$

c_1 is a constant of integration. Integrating once more gives

$$\rho_A(p) = p + \frac{\beta N}{k}\Big(\frac{u_1}{6}p^3 + \frac{u_2}{2}p^2\Big) + c_1 p + c_2 \tag{A.73}$$

Using the boundary conditions $\rho_A(0) = 0$ and $\rho_A(1) = 1$ to solve for the constants of

integration, we get

$$\begin{aligned}\rho_A(p) =& p + \frac{\beta N}{k}\Big(\frac{u_1}{6}p^3 + \frac{u_2}{2}p^2 - (\frac{u_1}{6} + \frac{u_2}{2})p\Big) \\ =& p + \frac{\beta N p(1-p)}{6k}\Big(- 3u_2 - u_1(1+p)\Big)\end{aligned} \tag{A.74}$$

When $p \ll 1$, such as when $p = 1/N$ for invasion probabilities, (A.74) becomes

$$\rho_A(p) \approx p + \frac{\beta N p(1-p)}{6k}\Big(- 3u_2 - u_1\Big) \tag{A.75}$$

We can use this work to calculate the fixation probability for the B strategy, as well. For a given p_A and p_B with $p_A + p_B = 1$, $m_B(p) = -m_A(1-p)$ and $v_B(p) = v_A(1-p)$. Therefore

$$\frac{-2m_B(p)}{v_B(p)} = \frac{2m_A(1-p)}{v_A(1-p)} = -\frac{\beta N}{k}\Big(u_1(1-p) + u_2\Big) = \frac{\beta N}{k}\Big(u_1 p - (u_1 + u_2)\Big) \tag{A.76}$$

From this, as in (A.75),

$$\rho_B(p) = p + \frac{\beta N}{k}\Big(\frac{w_1}{6}p^3 + \frac{w_2}{2}p^2 - (\frac{w_1}{6} + \frac{w_2}{2})p\Big) \approx p + \frac{\beta N}{k}\Big(- w_1 - 3w_2\Big) \tag{A.77}$$

with $w_1 = u_1$ and $w_2 = -(u_2 + u_1)$.

A.3.7. Fact-checker Accuracy

The adjustment to include a parameter λ to take into account inaccurate fact-checkers is very simple. Recall that a fact-checker with accuracy $\lambda \in [0, 1]$ gives benefit α to an A player and penalty γ to a B player with probability λ, and gives the opposite payoffs with probability $1 - \lambda$. Therefore, the expected payoff an A player receives from a C player is $\lambda\alpha + (1-\lambda)\gamma$ and the expected payoff for a B player is $\lambda\gamma + (1-\lambda)\alpha$.

All the previous work with pair approximation, weak selection, and the diffusion

approximation still hold, but we can replace the old expected payoffs of α and γ with the new expected payoffs $\lambda\alpha+(1-\lambda)\gamma$ and $\lambda\gamma+(1-\lambda)\alpha$, respectively. Conveniently, the substitution can be done at the very end, where α and γ appear as coefficients in Equation (A.69).

Bibliography

[1] Robert P. Abelson, *Mathematical models in social psychology*, Advances in Experimental Social Psychology (1967), 1–54.

[2] Alan I. Abramowitz and Kyle L. Saunders, *Is polarization a myth?*, The Journal of Politics **70** (2008), no. 2, 542–555.

[3] Paul R. Abramson, John H. Aldrich, Phil Paolino, and David W. Rohde, *"Sophisticated" voting in the 1988 presidential primaries*, American Political Science Review **86** (1992), no. 1, 55–69.

[4] James Adams, Jay Dow, and Samuel Merrill, *The political consequences of alienation-based and indifference-based voter abstention: Applications to presidential elections*, Political Behavior **28** (2006), no. 1, 65–86.

[5] Réka Albert and Albert-László Barabási, *Statistical mechanics of complex networks*, Reviews of Modern Physics **74** (2002), no. 1, 47–97.

[6] Réka Albert, Hawoong Jeong, and Albert-László Barabási, *Error and attack tolerance of complex networks*, Nature **406** (2000), no. 6794, 378–382.

[7] Gerald L. Alexanderson, *About the cover: Euler and Königsberg's bridges: A historical view*, Bulletin of the American Mathematical Society **43** (2006), no. 04, 567–574.

[8] Benjamin Allen, Gabor Lippner, and Martin A. Nowak, *Evolutionary games on isothermal graphs*, Nature Communications **10** (2019), no. 1, 5107.

[9] Neal Allen and Brian J. Brox, *The roots of third party voting: the 2000 Nader campaign in historical perspective*, Party Politics **11** (2005), no. 5, 623–637.

[10] Simon P. Anderson and Gerhard Glomm, *Alienation, indifference and the choice of ideological position*, Social Choice and Welfare **9** (1992), no. 1, 17–31.

[11] Clio Andris, David Lee, Marcus J. Hamilton, Mauro Martino, Christian E. Gunning, and John Armistead Selden, *The rise of partisanship and super-cooperators in the U.S. house of representatives*, PLOS ONE **10** (2015), no. 4, 1–14.

[12] Stephen Ansolabehere and Eitan Hersh, *Gender, race, age and voting: A research note*, Politics and Governance **1** (2013), no. 2, 132–137.

[13] Chris G. Antonopoulos and Yilun Shang, *Opinion formation in multiplex networks with general initial distributions*, Scientific Reports **8** (2018), no. 1, 2852.

[14] Kenneth Appel and Wolfgang Haken, *Every planar map is four colorable*, vol. 98, American Math. Soc, 1989.

[15] Krzysztof R. Apt, Mona Rahn, Guido Schäfer, and Sunil Simon, *Coordination games on graphs (extended abstract)*, Web and Internet Economics (2014), 441–446.

[16] Kevin Arceneaux and David W. Nickerson, *Who is mobilized to vote? A re-analysis of 11 field experiments*, American Journal of Political Science **53** (2009), no. 1, 1–16.

[17] Kenneth J. Arrow, *Social choice and individual values*, Yale University Press, 2012.

[18] Robert Axelrod and William Donald Hamilton, *The evolution of cooperation*, Science **211** (1981), no. 4489, 1390–1396.

[19] Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna, *Four degrees of separation*, Proceedings of the 4th Annual ACM Web Science Conference (New York, NY, USA), WebSci '12, Association for Computing Machinery, 2012, p. 33–42.

[20] Remo Badii and Antonio Politi, *Complexity: Hierarchical structures and scaling in physics*, Cambridge University Press, 2003.

[21] Franco Bagnoli and Raúl Rechtman, *Bifurcations in models of a society of reasonable contrarians and conformists*, Physical Review E **92** (2015), no. 4, 042913.

[22] Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, Alexander Volfovsky, and et al., *Exposure to opposing views on social media can increase political polarization*, Proceedings of the National Academy of Sciences **115** (2018), no. 37, 9216–9221.

[23] Delia Baldassarri and Peter Bearman, *Dynamics of political polarization*, American Sociological Review **72** (2007), no. 5, 784–811.

[24] John A. Banas and Stephen A. Rains, *A meta-analysis of research on inoculation theory*, Communication Monographs **77** (2010), no. 3, 281–311.

[25] Kevin K. Banda and John Cluverius, *Elite polarization, party extremity, and affective polarization*, Electoral Studies **56** (2018), 90–101.

[26] Abhijit V. Banerjee, *A simple model of herd behavior*, The Quarterly Journal of Economics **107** (1992), no. 3, 797–817.

[27] Jeffrey S. Banks and John Duggan, *Probabilistic voting in the spatial model of elections: The theory of office-motivated candidates*, Social Choice and Strategic Decisions (2005), 15–56.

[28] Albert-László Barabási and Réka Albert, *Emergence of scaling in random networks*, Science **286** (1999), no. 5439, 509–512.

[29] Pablo Barberá, *Social media, echo chambers, and political polarization*, Social Media and Democracy (2020), 34–55.

[30] Marc Barthelemy, *Betweenness centrality in large complex networks*, The European Physical Journal B - Condensed Matter **38** (2004), no. 2, 163–168.

[31] R. J. Baxter and F. Y. Wu, *Exact solution of an ising model with three-spin interactions on a triangular lattice*, Physical Review Letters **31** (1973), no. 21, 1294–1297.

[32] Peter S. Bearman, James Moody, and Katherine Stovel, *Chains of affection: The structure of adolescent romantic and sexual networks*, American Journal of Sociology **110** (2004), no. 1, 44–91.

[33] G. S. Becker and K. M. Murphy, *The division of labor, coordination costs, and knowledge*, The Quarterly Journal of Economics **107** (1992), no. 4, 1137–1160.

[34] Marianna Belloc, Ennio Bilancini, Leonardo Boncinelli, and Simone D'Alessandro, *Intuition and deliberation in the stag hunt game*, Scientific Reports **9** (2019), no. 1, 14833.

[35] Abram Bergson, *A reformulation of certain aspects of welfare economics*, The Quarterly Journal of Economics **52** (1938), no. 2, 310–334.

[36] G. D. Birkhoff and D. C. Lewis, *Chromatic polynomials*, Transactions of the American Mathematical Society **60** (1946), 355–451.

[37] Andreas Björklund, Thore Husfeldt, and Mikko Koivisto, *Set partitioning via inclusion-exclusion*, SIAM Journal on Computing **39** (2009), no. 2, 546–563.

[38] Duncan Black, *On the rationale of group decision-making*, Journal of Political Economy **56** (1948), no. 1, 23–34.

[39] ———, *The theory of committees and elections*, Cambridge University Press, 1958.

[40] Tony A. Blakely, Bruce P. Kennedy, and Ichiro Kawachi, *Socioeconomic inequality in voting participation and self-rated health*, American Journal of Public Health **91** (2001), no. 1, 99–104.

[41] Phillip Bonacich, *Power and centrality: A family of measures*, American Journal of Sociology **92** (1987), no. 5, 1170–1182.

[42] Ernesto Bonomi and Jean-Luc Lutton, *The n-city travelling salesman problem: Statistical mechanics and the metropolis algorithm*, SIAM Review **26** (1984), no. 4, 551–568.

[43] Thomas E. Borcherding and Robert T. Deacon, *The demand for the services of non-federal governments*, The American Economic Review **62** (1972), no. 5, 891–901.

[44] Gary Bornstein, Uri Gneezy, and Rosmarie Nagel, *The effect of intergroup competition on group coordination: An experimental study*, Games and Economic Behavior **41** (2002), no. 1, 1–25.

[45] Levi Boxell, Matthew Gentzkow, and Jesse M Shapiro, *Is the internet causing political polarization? Evidence from demographics*, Working Paper 23258, National Bureau of Economic Research, March 2017.

[46] Yann Bramoullé, *Anti-coordination and social interactions*, Games and Economic Behavior **58** (2007), no. 1, 30–49.

[47] H. Brandt, C. Hauert, and K. Sigmund, *Punishment and reputation in spatial public goods games*, Proceedings of the Royal Society of London. Series B: Biological Sciences **270** (2003), no. 1519, 1099–1104.

[48] Daniel Brélaz, *New methods to color the vertices of a graph*, Communications of the ACM **22** (1979), no. 4, 251–256.

[49] Daniel Breslau and Yuval Yonay, *Beyond metaphor: Mathematical models in economics as empirical research*, Science in Context **12** (1999), no. 2, 317–332.

[50] Sergey Brin and Lawrence Page, *The anatomy of a large-scale hypertextual web search engine*, Computer Networks and ISDN Systems **30** (1998), no. 1-7, 107–117.

[51] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener, *Graph structure in the web*, The Structure and Dynamics of Networks (2011), 183–194.

[52] Joris Broere, Vincent Buskens, Henk Stoof, and Angel Sánchez, *An experimental study of network effects on coordination in asymmetric games*, Scientific Reports **9** (2019), no. 1, 6482.

[53] Joris Broere, Vincent Buskens, Jeroen Weesie, and Henk Stoof, *Network effects on coordination in asymmetric games*, Scientific Reports **7** (2017), no. 1, 17016.

[54] Heather Z. Brooks and Mason A. Porter, *A model for the influence of media on the ideology of content in online social networks*, Physical Review Research **2** (2020), no. 2, 023041.

[55] R. L. Brooks, *On colouring the nodes of a network*, Mathematical Proceedings of the Cambridge Philosophical Society **37** (1941), no. 2, 194–197.

[56] Jacob R. Brown and Ryan D. Enos, *The measurement of partisan sorting for 180 million voters*, Nature Human Behaviour **5** (2021), no. 8, 998–1008.

[57] Barry C. Burden, *Deterministic and probabilistic voting models*, American Journal of Political Science **41** (1997), no. 4, 1150–1169.

[58] Barry C. Burden, Jason M. Fletcher, Pamela Herd, Bradley M. Jones, and Donald P. Moynihan, *How different forms of health matter to political participation*, The Journal of Politics **79** (2017), no. 1, 166–178.

[59] Leonardo Bursztyn, Aakaash Rao, Christopher Roth, and David Yanagizawa-Drott, *Misinformation during a pandemic*, Working Paper 27417, National Bureau of Economic Research, 2020.

[60] Fuxi Cai, Suhas Kumar, Thomas Van Vaerenbergh, Xia Sheng, Rui Liu, Can Li, Zhan Liu, Martin Foltin, Shimeng Yu, Qiangfei Xia, and et al., *Power-efficient combinatorial optimization using intrinsic noise in memristor Hopfield neural networks*, Nature Electronics **3** (2020), no. 7, 409–418.

[61] Filipe R. Campante and Daniel A. Hojman, *Media and polarization*, Journal of Public Economics **100** (2013), 79–92.

[62] Marcos Cardinot, Josephine Griffith, Colm O'Riordan, and Matjaž Perc, *Co-operation in the spatial prisoner's dilemma game with probabilistic abstention*, Scientific Reports **8** (2018), no. 1, 14531.

[63] Claudio Castellano, Santo Fortunato, and Vittorio Loreto, *Statistical physics of social dynamics*, Reviews of Modern Physics **81** (2009), no. 2, 591–646.

[64] Claudio Castellano and Romualdo Pastor-Satorras, *Thresholds for epidemic spreading in networks*, Physical Review Letters **105** (2010), no. 21, 218701.

[65] Pew Research Center, *Political polarization in the American public*, Tech. report, Pew Research Center, June 2014.

[66] ———, *Few Clinton or Trump supporters have close friends in the other camp*, Tech. report, Pew Research Center, August 2016.

[67] ———, *Partisanship and political animosity in 2016*, Tech. report, Pew Research Center, June 2016.

[68] ———, *Partisan conflict and congressional outreach*, Tech. report, Pew Research Center, February 2017.

[69] ———, *The partisan divide on political values grows even wider*, Tech. report, Pew Research Center, October 2017.

[70] G. J. Chaitin, *Register allocation & spilling via graph coloring*, ACM SIGPLAN Notices **17** (1982), no. 6, 98–101.

[71] Kamalika Chaudhuri, Fan Chung Graham, and Mohammad Shoaib Jamall, *A network coloring game*, Internet and Network Economics (Berlin, Heidelberg), Springer Berlin Heidelberg, 2008, p. 522–530.

[72] Eric Chea and Dennis R Livesay, *How accurate and statistically robust are catalytic site predictions based on closeness centrality?*, BMC Bioinformatics **8** (2007), no. 1, 153.

[73] Xiaojie Chen, Attila Szolnoki, and Matjaž Perc, *Probabilistic sharing solves the problem of costly punishment*, New Journal of Physics **16** (2014), no. 8, 083016.

[74] Syngjoo Choi, Douglas Gale, Shachar Kariv, and Thomas Palfrey, *Network architecture, salience and coordination*, Games and Economic Behavior **73** (2011), no. 1, 76–90.

[75] Fan R. K. Chung, *Spectral graph theory*, vol. 92, American Mathematical Soc., 1997.

[76] Joel E. Cohen, Kemperman Johannes Henricus Bernardus, and Zbăganu Gheorghe, *Comparisons of stochastic matrices with applications in information theory, statistics, economics, and population sciences*, Birkhäuser, 1998.

[77] James Coleman, Elihu Katz, and Herbert Menzel, *The diffusion of an innovation among physicians*, Sociometry **20** (1957), no. 4, 253–270.

[78] James Samuel Coleman, *Introduction to mathematical sociology*, Free Pr, 1964.

[79] William S. Comanor, *The median voter rule and the theory of political choice*, Journal of Public Economics **5** (1976), no. 1-2, 169–177.

[80] Michael D. Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini, *Political polarization on Twitter*, Fifth international AAAI conference on weblogs and social media, 2011.

[81] Peter J. Coughlin, *Unidimensional median voter results in probabilistic voting models*, Economics Letters **14** (1984), no. 1, 9–15.

[82] ———, *Probabilistic voting theory*, Cambridge Univ. Press, 1992.

[83] Iain D. Couzin, Christos C. Ioannou, Güven Demirel, Thilo Gross, Colin J. Torney, Andrew Hartnett, Larissa Conradt, Simon A. Levin, and Naomi E. Leonard, *Uninformed individuals promote democratic consensus in animal groups*, Science **334** (2011), no. 6062, 1578–1580.

[84] Iain D. Couzin, Jens Krause, Nigel R. Franks, and Simon A. Levin, *Effective leadership and decision-making in animal groups on the move*, Nature **433** (2005), no. 7025, 513–516.

[85] Gerald F. Davis, Mina Yoo, and Wayne E. Baker, *The small world of the American corporate elite, 1982-2001*, Strategic Organization **1** (2003), no. 3, 301–326.

[86] Otto A. Davis and Melvin J. Hinich, *A mathematical model of policy formation in a democratic society*, Graduate School of Industrial Administration, Carnegie Institute of Technology, 1965.

[87] Otto A. Davis, Melvin J. Hinich, and Peter C. Ordeshook, *An expository development of a mathematical model of the electoral process*, American Political Science Review **64** (1970), no. 2, 426–448.

[88] D. de Werra, *An introduction to timetabling*, European Journal of Operational Research **19** (1985), no. 2, 151–162.

[89] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi, *The spreading of misinformation online*, Proceedings of the National Academy of Sciences **113** (2016), no. 3, 554–559.

[90] Daniel DellaPosta, *Pluralistic collapse: The "oil spill" model of mass opinion polarization*, American Sociological Review **85** (2020), no. 3, 507–536.

[91] Daniel DellaPosta, Yongren Shi, and Michael Macy, *Why do liberals drink lattes?*, American Journal of Sociology **120** (2015), no. 5, 1473–1511.

[92] Avinash K. Dixit and Jörgen W. Weibull, *Political polarization*, Proceedings of the National Academy of Sciences **104** (2007), no. 18, 7351–7356.

[93] Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts, *An experimental study of search in global social networks*, Science **301** (2003), no. 5634, 827–829.

[94] Michael Doebeli and Christoph Hauert, *Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game*, Ecology Letters **8** (2005), no. 7, 748–766.

[95] Anthony Downs, *An economic theory of democracy*, Harper, 1957.

[96] James N. Druckman, Erik Peterson, and Rune Slothuus, *How elite partisan polarization affects public opinion formation*, American Political Science Review **107** (2013), no. 1, 57–79.

[97] R. Durrett and S. Levin, *The importance of being discrete (and spatial)*, Theoretical Population Biology **46** (1994), no. 3, 363–394.

[98] Richard Durrett, *Essentials of stochastic processes*, Springer, 2018.

[99] James M. Enelow and Melvin J. Hinich, *Probabilistic voting and the importance of centrist ideologies in democratic elections*, The Journal of Politics **46** (1984), no. 2, 459–478.

[100] Ryan D. Enos and Anthony Fowler, *Aggregate effects of large-scale campaigns on voter turnout*, Political Science Research and Methods **6** (2018), no. 4, 733–751.

[101] Ziv Epstein, Gordon Pennycook, and David Gertler Rand, *Will the crowd game the algorithm? Using layperson judgments to combat misinformation on social*

media by downranking distrusted sources, p. 1–11, Association for Computing Machinery, New York, NY, USA, 2020.

[102] P. Erdős and A. Rényi, *On the evolution of random graphs*, The Structure and Dynamics of Networks (2011), 38–82.

[103] Güler Ergün, *Human sexual contact network as a bipartite graph*, Physica A: Statistical Mechanics and its Applications **308** (2002), no. 1-4, 483–488.

[104] Emily Erikson and Hirokazu Shirado, *Networks, property, and the division of labor*, American Sociological Review **86** (2021), no. 4, 759–786.

[105] Leonhard Euler, *Solutio problematis ad geometriam situs pertinentis*, Commentarii academiae scientiarum Petropolitanae **8** (1741), 128–140.

[106] Tucker Evans and Feng Fu, *Opinion formation on dynamic networks: Identifying conditions for the emergence of partisan echo chambers*, Royal Society Open Science **5** (2018), no. 10, 181122.

[107] D. Fang, S. O. Kimbrough, S. Pace, A. Valluri, and Z. Zheng, *On adaptive emergence of trust behavior in the game of stag hunt*, Group Decision and Negotiation **11** (2002), no. 6, 449–467.

[108] Mark J. Fenster, *The impact of allowing day of registration voting on turnout in U.S. elections from 1960 to 1992*, American Politics Quarterly **22** (1994), no. 1, 74–87.

[109] Miroslav Fiedler, *Algebraic connectivity of graphs*, Czechoslovak Mathematical Journal **23** (1973), no. 2, 298–305.

[110] Irene Finocchi, Alessandro Panconesi, and Riccardo Silvestri, *An experimental analysis of simple, distributed vertex coloring algorithms*, Algorithmica **41** (2004), no. 1, 1–23.

[111] Morris P. Fiorina and Samuel J. Abrams, *Political polarization in the American public*, Annual Review of Political Science **11** (2008), no. 1, 563–588.

[112] Morris P. Fiorina, Samuel J. Abrams, and Jeremy C. Pope, *Culture war?: The myth of a polarized America*, Longman, 2011.

[113] Andreas Flache, Michael Mäs, Thomas Feliciani, Edmund Chattoe-Brown, Guillaume Deffuant, Sylvie Huet, and Jan Lorenz, *Models of social influence: Towards the next frontiers*, Journal of Artificial Societies and Social Simulation **20** (2017), no. 4, 2.

[114] L. R. Ford and D. R. Fulkerson, *Maximal flow through a network*, Canadian Journal of Mathematics **8** (1956), 399–404.

[115] Mark N. Franklin, Cees van der Eijk, Diana Evans, Michael A. Fotos, Wolfgang Hirczy, Michael Marsh, and Bernhard Wessels, *Voter turnout and the dynamics of electoral competition in established democracies since 1945*, Cambridge University Press, 2004.

[116] Robert Franzese, *Career achievement award of the society for political methodology: Keith Poole*, 2016.

[117] Noah E. Friedkin, Anton V. Proskurnikov, Roberto Tempo, and Sergey E. Parsegov, *Network science on belief system dynamics under logic constraints*, Science **354** (2016), no. 6310, 321–326.

[118] Feng Fu and Xingru Chen, *Leveraging statistical physics to improve understanding of cooperation in multiplex networks*, New Journal of Physics **19** (2017), no. 7, 071002.

[119] Feng Fu, Christoph Hauert, Martin A. Nowak, and Long Wang, *Reputation-based partner choice promotes cooperation in social networks*, Physical Review E **78** (2008), no. 2, 026117.

[120] Feng Fu, Martin A. Nowak, Nicholas A. Christakis, and James H. Fowler, *The evolution of homophily*, Scientific Reports **2** (2012), no. 1, 845.

[121] Feng Fu and Long Wang, *Coevolutionary dynamics of opinions and networks: From diversity to uniformity*, Physical Review E **78** (2008), no. 1, 016104.

[122] Paul A. Gagniuc, *Markov chains from theory to implementation and experimentation*, Wiley, 2017.

[123] Francis Galton, *Vox populi*, Nature **75** (1907), no. 1949, 450–451.

[124] Michael R. Garey and David S. Johnson, *Computers and intractability*, Freeman, 1999.

[125] Alan S. Gerber, Donald P. Green, and Ron Shachar, *Voting may be habit-forming: Evidence from a randomized field experiment*, American Journal of Political Science **47** (2003), no. 3, 540–550.

[126] E. N. Gilbert, *Random graphs*, The Annals of Mathematical Statistics **30** (1959), no. 4, 1141–1144.

[127] David F. Gleich, *Pagerank beyond the web*, SIAM Review **57** (2015), no. 3, 321–363.

[128] Amir Goldberg and Sarah K. Stein, *Beyond social contagion: Associative diffusion and the emergence of cultural variation*, American Sociological Review **83** (2018), no. 5, 897–932.

[129] Brad T. Gomez, Thomas G. Hansford, and George A. Krause, *The Republicans should pray for rain: Weather, turnout, and voting in U.S. presidential elections*, The Journal of Politics **69** (2007), no. 3, 649–663.

[130] J. Gómez-Gardeñes, M. Campillo, L. M. Floría, and Y. Moreno, *Dynamical organization of cooperation in complex topologies*, Physical Review Letters **98** (2007), no. 10, 108103.

[131] A. Grabowski and R. A. Kosiński, *Ising-based model of opinion formation in a complex network of interpersonal interactions*, Physica A: Statistical Mechanics and its Applications **361** (2006), no. 2, 651–664.

[132] P. Grassberger, *On the critical behavior of the general epidemic process and dynamical percolation*, Mathematical Biosciences **63** (1983), no. 2, 157–172.

[133] Charles Miller Grinstead and J. Laurie Snell, *Introduction to probability*, American Mathematical Society, 2006.

[134] Jens Großer and Thomas R. Palfrey, *Candidate entry and political polarization: An antimedian voter theorem*, American Journal of Political Science **58** (2014), no. 1, 127–143.

[135] Jean-Loup Guillaume and Matthieu Latapy, *Bipartite graphs as models of complex networks*, Physica A: Statistical Mechanics and its Applications **371** (2006), no. 2, 795–813.

[136] Jeremy Gunawardena, *Models in biology: 'accurate descriptions of our pathetic thinking'*, BMC Biology **12** (2014), no. 1, 29.

[137] Andrew B. Hall and Daniel M. Thompson, *Who punishes extremist nominees? Candidate ideology and turning out the base in US elections*, American Political Science Review **112** (2018), no. 3, 509–524.

[138] Thomas H. Hammond, *Intelligence organizations and the organization of intelligence*, International Journal of Intelligence and CounterIntelligence **23** (2010), no. 4, 680–724.

[139] Eric R. Hansen and Andrew Tyner, *Educational attainment and social norms of voting*, Political Behavior **43** (2021), no. 2, 711–735.

[140] Pierre Hansen and Michel Delattre, *Complete-link cluster analysis by graph coloring*, Journal of the American Statistical Association **73** (1978), no. 362, 397–403.

[141] Garrett Hardin, *The tragedy of the commons*, Science **162** (1968), no. 3859, 1243–1248.

[142] Russell Hardin, *Collective action*, Johns Hopkins Univ. Press, 1982.

[143] Matthew He and S. V. Petukhov, *Mathematics of bioinformatics: Theory, methods and applications*, Wiley-Blackwell, 2011.

[144] James J. Heckman and James M. Snyder, *Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators*, The RAND Journal of Economics **28** (1997), no. 0, S142–S189.

[145] Dirk Helbing, Attila Szolnoki, Matjaž Perc, and György Szabó, *Punish, but not too hard: How costly punishment spreads in the spatial public goods game*, New Journal of Physics **12** (2010), no. 8, 083005.

[146] Melvin J. Hinich, *Equilibrium in spatial voting: The median voter result is an artifact*, Journal of Economic Theory **16** (1977), no. 2, 208–219.

[147] Melvin J. Hinich, John O. Ledyard, and Peter C. Ordeshook, *Nonvoting and the existence of equilibrium under majority rule*, Journal of Economic Theory **4** (1972), no. 2, 144–153.

[148] Melvin J. Hinich and Peter C. Ordeshook, *Abstentions and equilibrium in the electoral process*, Public Choice **7-7** (1969), no. 1, 81–106.

[149] ———, *Plurality maximization vs vote maximization: A spatial analysis with variable participation*, American Political Science Review **64** (1970), no. 3, 772–791.

[150] Shigeo Hirano and James M. Snyder, *The decline of third-party voting in the United States*, The Journal of Politics **69** (2007), no. 1, 1–16.

[151] Jennifer L. Hochschild, *If democracies need informed voters, how can they thrive while expanding enfranchisement?*, Election Law Journal: Rules, Politics, and Policy **9** (2010), no. 2, 111–123.

[152] J. Hofbauer and K. Sigmund, *Adaptive dynamics and evolutionary stability*, Applied Mathematics Letters **3** (1990), no. 4, 75–79.

[153] Josef Hofbauer and Karl Sigmund, *Evolutionary game dynamics*, Bulletin of the American Mathematical Society **40** (2003), no. 4, 479–519.

[154] Petter Holme and M. E. Newman, *Nonequilibrium phase transition in the coevolution of networks and opinions*, Physical Review E **74** (2006), no. 5, 056108.

[155] Harold Hotelling, *Stability in competition*, The Economic Journal **39** (1929), no. 153, 41–57.

[156] M. O. Jackson and Y. Xing, *Culture-dependent strategies in coordination games*, Proceedings of the National Academy of Sciences **111** (2014), no. 3, 10889–10896.

[157] Thomas Jefferson and J. Jefferson Looney, *The papers of Thomas Jefferson, retirement series*, Princeton University Press, 2011.

[158] David S. Johnson, Cecilia R. Aragon, Lyle A. McGeoch, and Catherine Schevon, *Optimization by simulated annealing: An experimental evaluation; part II, graph coloring and number partitioning*, Operations Research **39** (1991), no. 3, 378–406.

[159] Matthew I. Jones, Scott D. Pauls, and Feng Fu, *The dual problems of coordination and anti-coordination on random bipartite graphs*, New Journal of Physics **23** (2021), no. 11, 113018.

[160] ———, *Random choices facilitate solutions to collective network coloring problems by artificial agents*, iScience **24** (2021), no. 4, 102340.

[161] Matthew I. Jones, Antonio D. Sirianni, and Feng Fu, *Polarization, abstention, and the Median Voter Theorem*, Humanities and Social Sciences Communications **9** (2022), no. 1, 43.

[162] S. Judd, M. Kearns, and Y. Vorobeychik, *Behavioral dynamics and influence in networked coloring and consensus*, Proceedings of the National Academy of Sciences **107** (2010), no. 34, 14978–14982.

[163] Jonas S. Juul and Mason A. Porter, *Hipsters on networks: How a minority group of individuals can lead to an antiestablishment majority*, Physical Review E **99** (2019), no. 2, 022313.

[164] Mari Kawakatsu, Yphtach Lelkes, Simon A. Levin, and Corina E. Tarnita, *Interindividual cooperation mediated by partisanship complicates Madison's cure for "mischiefs of faction"*, Proceedings of the National Academy of Sciences **118** (2021), no. 50, e2102148118.

[165] Toshihiro Kawakatsu, Kyozi Kawasaki, Michihiro Furusaka, Hirofumi Okabayashi, and Toshiji Kanaya, *Late stage dynamics of phase separation processes of binary mixtures containing surfactants*, The Journal of Chemical Physics **99** (1993), no. 10, 8200–8217.

[166] Grace Kay, *A majority of Americans surveyed believe the US is in the midst of a 'cold' Civil War*, Jan 2021.

[167] Michael Kearns, Siddharth Suri, and Nick Montfort, *An experimental study of the coloring problem on human subject networks*, Science **313** (2006), no. 5788, 824–827.

[168] Tommy Khoo, Feng Fu, and Scott Pauls, *Spillover modes in multiplex games: Double-edged effects on cooperation and their coevolution*, Scientific Reports **8** (2018), no. 1, 6922.

[169] Ken Kollman, John H. Miller, and Scott E. Page, *Adaptive parties in spatial elections*, American Political Science Review **86** (1992), no. 4, 929–937.

[170] Jeremy Kun, Brian Powers, and Lev Reyzin, *Anti-coordination games and stable graph colorings*, Algorithmic Game Theory, Springer Berlin Heidelberg, 2013, p. 122–133.

[171] Casimir Kuratowski, *Sur le problème des courbes gauches en topologie*, Fundamenta Mathematicae **15** (1930), 271–283.

[172] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain, *The science of fake news*, Science **359** (2018), no. 6380, 1094–1096.

[173] Ro'ee Levy, *Social media, news consumption, and polarization: Evidence from a field experiment*, American Economic Review **111** (2021), no. 3, 831–870.

[174] Thomas M. Liggett, *Coexistence in threshold voter models*, The Annals of Probability **22** (1994), no. 2, 764–802.

[175] Linjie Liu, Xiaojie Chen, and Matjaž Perc, *Evolutionary dynamics of cooperation in the public goods game with pool exclusion strategies*, Nonlinear Dynamics **97** (2019), no. 1, 749–766.

[176] Linjie Liu, Xiaojie Chen, and Attila Szolnoki, *Evolutionary dynamics of cooperation in a population with probabilistic corrupt enforcers and violators*, Mathematical Models and Methods in Applied Sciences **29** (2019), no. 11, 2127–2149.

[177] J. Lorenz, H. Rauhut, F. Schweitzer, and D. Helbing, *How social influence can undermine the wisdom of crowd effect*, Proceedings of the National Academy of Sciences **108** (2011), no. 22, 9020–9025.

[178] Irving Lorge, David Fox, Joel Davitz, and Marlin Brenner, *A survey of studies contrasting the quality of group performance and individual performance, 1920-1957.*, Psychological Bulletin **55** (1958), no. 6, 337–372.

[179] Alfred J. Lotka, *Analytical note on certain rhythmic relations in organic systems*, Proceedings of the National Academy of Sciences **6** (1920), no. 7, 410–415.

[180] L. Lovász and M.D. Plummer, *Matching theory*, North-Holland, 1986.

[181] R. Duncan Luce and Howard Raiffa, *Games and decisions: Introduction and critical survey*, Wiley, 1957.

[182] Donald A. MacKenzie, *An engine, not a camera: How financial models shape markets*, MIT, 2008.

[183] Thomas W. Malone, Robert Laubacher, and Chrysanthos N. Dellarocas, *Harnessing crowds: Mapping the genome of collective intelligence*, SSRN Electronic Journal (2009), 4732–09.

[184] Gregory J. Martin and Steven W. Webster, *Does residential sorting explain geographic polarization?*, Political Science Research and Methods **8** (2020), no. 2, 215–231.

[185] Roger A. McCain and Richard Hamilton, *Coordination games, anti-coordination games, and imitative learning*, Behavioral and Brain Sciences **37** (2014), no. 1, 90–91.

[186] Mathew D. McCubbins, Ramamohan Paturi, and Nicholas Weller, *Connected coordination*, American Politics Research **37** (2009), no. 5, 899–920.

[187] Mathew D. McCubbins and Nicholas Weller, *Coordination, communication, and information: How network structure and knowledge affect group behavior*, Journal of Experimental Political Science **7** (2020), no. 1, 1–12.

[188] William J. McGuire and Demetrios Papageorgis, *Effectiveness of forewarning in developing resistance to persuasion*, Public Opinion Quarterly **26** (1962), no. 1, 24–34.

[189] Richard D. McKelvey and John W. Patty, *A theory of voting in large elections*, Games and Economic Behavior **57** (2006), no. 1, 155–180.

[190] Miller McPherson, Lynn Smith-Lovin, and James M Cook, *Birds of a feather: Homophily in social networks*, Annual Review of Sociology **27** (2001), no. 1, 415–444.

[191] Karl Menger, *Zur allgemeinen kurventheorie*, Fundamenta Mathematicae **10** (1927), 96–115.

[192] Areeb Mian and Shujhat Khan, *Coronavirus: The spread of misinformation*, BMC Medicine **18** (2020), no. 1, 89.

[193] Stanley Milgram, *The small-world problem*, Psychology Today **2** (1967), no. 1, 60–67.

[194] James Moody, *Race, school integration, and friendship segregation in America*, American Journal of Sociology **107** (2001), no. 3, 679–716.

[195] Mohsen Mosleh, Cameron Martel, Dean Eckles, and David Rand, *Perverse downstream consequences of debunking: Being corrected by another user for posting false political news increases subsequent sharing of low quality, partisan, and toxic content in a Twitter field experiment*, Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (2021), 1–13.

[196] Mohsen Mosleh, Cameron Martel, Dean Eckles, and David G. Rand, *Shared partisanship dramatically increases social tie formation in a Twitter field experiment*, Proceedings of the National Academy of Sciences **118** (2021), no. 7, e2022761118.

[197] Subhayan Mukerjee, Kokil Jaidka, and Yphtach Lelkes, *The ideological landscape of Twitter: Comparing the production versus consumption of information on the platform*, 2020.

[198] Kevin J. Mullinix, *Partisanship and preference formation: Competing motivations, elite polarization, and issue importance*, Political Behavior **38** (2016), no. 2, 383–411.

[199] Raj Rao Nadakuditi and M. E. Newman, *Graph spectra and the detectability of community structure in networks*, Physical Review Letters **108** (2012), no. 18, 188701.

[200] Richard Nadeau, Michael S. Lewis-Beck, and Martial Foucault, *Wealth and voter turnout: Investigating twenty-eight democracies*, Polity **51** (2019), no. 2, 261–287.

[201] Satoshi Nakamoto, *Bitcoin: A peer-to-peer electronic cash system*, Decentralized Business Review (2008), 21260.

[202] Cecilia Nardini, Balázs Kozma, and Alain Barrat, *Who's talking first? Consensus or lack thereof in coevolving opinion formation models*, Physical Review Letters **100** (2008), no. 15, 158701.

[203] John Nash, *Non-cooperative games*, Annals of Mathematics **54** (1951), no. 2, 286–295.

[204] John F. Nash, *Equilibrium points in n-person games*, Proceedings of the National Academy of Sciences **36** (1950), no. 1, 48–49.

[205] Tibor Neugebauer, Anders Poulsen, and Arthur Schram, *Fairness and reciprocity in the hawk–dove game*, Journal of Economic Behavior & Organization **66** (2008), no. 2, 243–250.

[206] John Von Neumann and Oskar Morgenstern, *Theory of games and economic behavior*, Princeton Univ. Pr, 1953.

[207] M. E. Newman, *Clustering and preferential attachment in growing networks*, Physical Review E **64** (2001), no. 2, 025102.

[208] ———, *The structure and function of complex networks*, SIAM Review **45** (2003), no. 2, 167–256.

[209] Mark E. J. Newman, *Networks an introduction*, Oxford University Press, 2018.

[210] Hossein Noorazar, *Recent advances in opinion propagation dynamics: A 2020 survey*, The European Physical Journal Plus **135** (2020), no. 6, 521.

[211] Pippa Norris, *Electoral engineering: Voting rules and political behavior*, Cambridge University Press, 2012.

[212] M. A. Nowak, *Evolutionary dynamics: Exploring the equations of life*, Belknap Press of Harvard University Press, 2006.

[213] Martin A. Nowak, *Five rules for the evolution of cooperation*, Science **314** (2006), no. 5805, 1560–1563.

[214] Martin A. Nowak and Robert M. May, *Evolutionary games and spatial chaos*, Nature **359** (1992), no. 6398, 826–829.

[215] Martin A. Nowak, Akira Sasaki, Christine Taylor, and Drew Fudenberg, *Emergence of cooperation and evolutionary stability in finite populations*, Nature **428** (2004), no. 6983, 646–650.

[216] Hisashi Ohtsuki, Christoph Hauert, Erez Lieberman, and Martin A. Nowak, *A simple rule for the evolution of cooperation on graphs and social networks*, Nature **441** (2006), no. 7092, 502–505.

[217] Hisashi Ohtsuki and Martin A. Nowak, *The replicator equation on graphs*, Journal of Theoretical Biology **243** (2006), no. 1, 86–97.

[218] J Oitmaa, *The square-lattice Ising model with first and second neighbour interactions*, Journal of Physics A: Mathematical and General **14** (1981), no. 5, 1159–1168.

[219] Mancur Olson, *The logic of collective action*, Harvard Univ. Pr., 1977.

[220] Peter C. Ordeshook, *Extensions to a model of the electoral process and implications for the theory of responsible parties*, Midwest Journal of Political Science **14** (1970), no. 1, 43–70.

[221] Martin J. Osborne and Ariel Rubinstein, *A course in game theory*, MIT Press, 1994.

[222] Elinor Ostrom, *Collective action and the evolution of social norms*, Journal of Economic Perspectives **14** (2000), no. 3, 137–158.

[223] Thomas R. Palfrey, *Spatial equilibrium with entry*, The Review of Economic Studies **51** (1984), no. 1, 139–156.

[224] Taehoon Park and Chae Y. Lee, *Application of the graph coloring algorithm to the frequency assignment problem*, Journal of the Operations Research Society of Japan **39** (1996), no. 2, 258–265.

[225] Romualdo Pastor-Satorras and Alessandro Vespignani, *Epidemic spreading in scale-free networks*, Physical Review Letters **86** (2001), no. 14, 3200–3203.

[226] Jorge Peña and Yannick Rochat, *Bipartite graphs as models of population structures in evolutionary multiplayer games*, PLoS ONE **7** (2012), no. 9, e44514.

[227] Gordon Pennycook, Adam Bear, Evan T. Collins, and David G. Rand, *The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings*, Management Science **66** (2020), no. 11, 4944–4957.

[228] Gordon Pennycook, Tyrone D. Cannon, and David G. Rand, *Prior exposure increases perceived accuracy of fake news.*, Journal of Experimental Psychology: General **147** (2018), no. 12, 1865–1880.

[229] Gordon Pennycook, Jonathon McPhetres, Yunhao Zhang, Jackson G. Lu, and David G. Rand, *Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention*, Psychological Science **31** (2020), no. 7, 770–780.

[230] Gordon Pennycook and David G. Rand, *Fighting misinformation on social media using crowdsourced judgments of news source quality*, Proceedings of the National Academy of Sciences **116** (2019), no. 7, 2521–2526.

[231] Matjaž Perc, *Double resonance in cooperation induced by noise and network variation for an evolutionary prisoner's dilemma*, New Journal of Physics **8** (2006), no. 9, 183–183.

[232] Matjaž Perc, Jillian J. Jordan, David G. Rand, Zhen Wang, Stefano Boccaletti, and Attila Szolnoki, *Statistical physics of human cooperation*, Physics Reports **687** (2017), 1–51.

[233] Matjaž Perc and Attila Szolnoki, *Coevolutionary games—a mini review*, Biosystems **99** (2010), no. 2, 109–125.

[234] Geoff Peterson and J. Mark Wrighton, *Expressions of distrust: Third-party voting and cynicism in government*, Political Behavior **20** (1998), no. 1, 17–34.

[235] M. Pineda, R. Toral, and E. Hernández-García, *Noisy continuous-opinion dynamics*, Journal of Statistical Mechanics: Theory and Experiment **2009** (2009), no. 08, P08001.

[236] Dennis L. Plane and Joseph Gershtenson, *Candidates' ideological locations, abstention, and turnout in U.S. midterm Senate elections*, Political Behavior **26** (2004), no. 1, 69–93.

[237] Keith T. Poole, *Spatial models of parliamentary voting*, Cambridge Univ. Press, 2005.

[238] Keith T. Poole and Howard Rosenthal, *U.S. presidential elections 1968-80: A spatial analysis*, American Journal of Political Science **28** (1984), no. 2, 282–312.

[239] ———, *A spatial model for legislative roll call analysis*, American Journal of Political Science **29** (1985), no. 2, 357–384.

[240] ———, *Congress: A political-economic history of roll call voting*, Oxford University Press, 1997.

[241] ———, *D-nominate after 10 years: A comparative update to Congress: A political-economic history of roll-call voting*, Legislative Studies Quarterly **26** (2001), no. 1, 5–29.

[242] D. J. Price, *Networks of scientific papers*, The Structure and Dynamics of Networks (2011), 149–154.

[243] Derek De Price, *A general theory of bibliometric and other cumulative advantage processes*, Journal of the American Society for Information Science **27** (1976), no. 5, 292–306.

[244] Carey E. Priebe, Youngser Park, Joshua T. Vogelstein, John M. Conroy, Vince Lyzinski, Minh Tang, Avanti Athreya, Joshua Cape, and Eric Bridgeford, *On a two-truths phenomenon in spectral graph clustering*, Proceedings of the National Academy of Sciences **116** (2019), no. 13, 5995–6000.

[245] Markus Prior, *Media and political polarization*, Annual Review of Political Science **16** (2013), no. 1, 101–127.

[246] Jingtao Qi, Liang Bai, and Yandong Xiao, *Social network-oriented learning agent for improving group intelligence coordination*, IEEE Access **7** (2019), 156526–156535.

[247] Iyad Rahwan, Manuel Cebrian, Nick Obradovich, Josh Bongard, Jean-François Bonnefon, Cynthia Breazeal, Jacob W. Crandall, Nicholas A. Christakis, Iain D. Couzin, Matthew O. Jackson, and et al., *Machine behaviour*, Nature **568** (2019), no. 7753, 477–486.

[248] D. G. Rand, S. Arbesman, and N. A. Christakis, *Dynamic social networks promote cooperation in experiments with humans*, Proceedings of the National Academy of Sciences **108** (2011), no. 48, 19193–19198.

[249] David G. Rand and Martin A. Nowak, *Human cooperation*, Trends in Cognitive Sciences **17** (2013), no. 8, 413–425.

[250] David G. Rand, Martin A. Nowak, James H. Fowler, and Nicholas A. Christakis, *Static network structure can stabilize human cooperation*, Proceedings of the National Academy of Sciences **111** (2014), no. 48, 17093–17098.

[251] Sidney Redner, *Reality-inspired voter models: A mini-review*, Comptes Rendus Physique **20** (2019), no. 4, 275–292.

[252] Jon C. Rogowski and Stephanie Langella, *Primary systems and candidate ideology: Evidence from federal and state legislative elections*, American Politics Research **43** (2015), no. 5, 846–871.

[253] Jon Roozenbeek and Sander van der Linden, *The fake news game: Actively inoculating against the risk of misinformation*, Journal of Risk Research **22** (2019), no. 5, 570–580.

[254] ———, *Fake news game confers psychological resistance against online misinformation*, Palgrave Communications **5** (2019), no. 1, 65.

[255] Ryan A. Rossi and Nesreen K. Ahmed, *The network data repository with interactive graph analytics and visualization*, Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI'15, AAAI Press, 2015, p. 4292–4293.

[256] Nasser R. Sabar, Masri Ayob, Rong Qu, and Graham Kendall, *A graph coloring constructive hyper-heuristic for examination timetabling problems*, Applied Intelligence **37** (2012), no. 1, 1–11.

[257] Silvio R. A. Salinas, *Introduction to statistical physics*, Springer, 2011.

[258] F. C. Santos and J. M. Pacheco, *Scale-free networks provide a unifying framework for the emergence of cooperation*, Physical Review Letters **95** (2005), no. 9, 098104.

[259] Jack Santucci, *Variants of ranked-choice voting from a strategic perspective*, Politics and Governance **9** (2021), no. 2, 344–353.

[260] Ana Lucía Schmidt, Fabiana Zollo, Antonio Scala, Cornelia Betsch, and Walter Quattrociocchi, *Polarization of the vaccination debate on Facebook*, Vaccine **36** (2018), no. 25, 3606–3612.

[261] Amartya Sen, *Development as freedom*, Anchor Books, 2000.

[262] L. Shi and S. K. Singh, *Decentralized adaptive controller design for large-scale systems with higher order interconnections*, IEEE Transactions on Automatic Control **37** (1992), no. 8, 1106–1118.

[263] Jieun Shin, Lian Jian, Kevin Driscoll, and François Bar, *The diffusion of misinformation on social media: Temporal pattern, message, and source*, Computers in Human Behavior **83** (2018), 278–287.

[264] Hirokazu Shirado and Nicholas A. Christakis, *Locally noisy autonomous agents improve global human coordination in network experiments*, Nature **545** (2017), no. 7654, 370–374.

[265] ———, *Network engineering using autonomous agents increases cooperation in human groups*, iScience **23** (2020), no. 9, 101438.

[266] Hirokazu Shirado, Feng Fu, James H. Fowler, and Nicholas A. Christakis, *Quality versus quantity of social ties in experimental cooperative networks*, Nature Communications **4** (2013), no. 1, 2814.

[267] K. Sigmund, C. Hauert, and M. A. Nowak, *Reward and punishment*, Proceedings of the National Academy of Sciences **98** (2001), no. 19, 10757–10762.

[268] Karl Sigmund, Hannelore De Silva, Arne Traulsen, and Christoph Hauert, *Social learning promotes institutions for governing the commons*, Nature **466** (2010), no. 7308, 861–863.

[269] Herbert A. Simon, *On a class of skew distribution functions*, Biometrika **42** (1955), no. 3/4, 425–440.

[270] Harmanjit Singh and Richa Sharma, *Role of adjacency matrix & adjacency list in graph theory*, INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY **3** (2012), no. 1, 179–183.

[271] S. Sivasankari and G. Vadivu, *Tracing the fake news propagation path using social network analysis*, Soft Computing (2021).

[272] Brian Skyrms, *The stag hunt*, Proceedings and Addresses of the American Philosophical Association **75** (2001), no. 2, 31–41.

[273] ———, *The stag hunt and the evolution of social structure*, Cambridge University Press, 2004.

[274] Kaat Smets and Carolien van Ham, *The embarrassment of riches? A meta-analysis of individual-level research on voter turnout*, Electoral Studies **32** (2013), no. 2, 344–359.

[275] Adam Smith and Edwin Cannan, *The wealth of nations*, Bantam Dell, 2003.

[276] J. Maynard Smith and G. R. Price, *The logic of animal conflict*, Nature **246** (1973), no. 5427, 15–18.

[277] V. Sood and S. Redner, *Voter model on heterogeneous graphs*, Physical Review Letters **94** (2005), no. 17, 178701.

[278] Alexander J. Stewart, Antonio A. Arechar, David G. Rand, and Joshua B. Plotkin, *The coercive logic of fake news*, arXiv preprint abs/2108.13687 (2021).

[279] Alexander J. Stewart, Mohsen Mosleh, Marina Diakonova, Antonio A. Arechar, David G. Rand, and Joshua B. Plotkin, *Information gerrymandering and undemocratic decisions*, Nature **573** (2019), no. 7772, 117–121.

[280] Jeffery M. Stonecash, Mark D. Brewer, and Mack D. Mariani, *Diverging parties: Social change, realignment, and party polarization*, Routledge, 2003.

[281] Steven H. Strogatz, *Exploring complex networks*, Nature **410** (2001), no. 6825, 268–276.

[282] Wei Su, Ge Chen, and Yiguang Hong, *Noise leads to quasi-consensus of Hegselmann–Krause opinion dynamics*, Automatica **85** (2017), 448–454.

[283] Cass R. Sunstein, *#republic: Divided democracy in the age of social media*, Princeton University Press, 2018.

[284] James Surowiecki, *The wisdom of crowds*, Anchor Books, 2005.

[285] Pontus Svenson, *From Néel to NPC: Colouring small worlds*, arXiv preprint cs/0107015 (2001).

[286] Briony Swire-Thompson, Ullrich K. Ecker, Stephan Lewandowsky, and Adam J. Berinsky, *They might be a liar but they're my liar: Source evaluation and the prevalence of misinformation*, Political Psychology **41** (2019), no. 1, 21–34.

[287] György Szabó and Gábor Fáth, *Evolutionary games on graphs*, Physics Reports **446** (2007), no. 4-6, 97–216.

[288] György Szabó, Attila Szolnoki, Melinda Varga, and Lívia Hanusovszky, *Ordering in spatial evolutionary games for pairwise collective strategy updates*, Physical Review E **82** (2010), no. 2, 026110.

[289] György Szabó, Jeromos Vukov, and Attila Szolnoki, *Phase diagrams for an evolutionary prisoner's dilemma game on two-dimensional lattices*, Physical Review E **72** (2005), no. 4, 047107.

[290] Attila Szolnoki, Matjaž Perc, and György Szabó, *Topology-independent impact of noise on cooperation in spatial public goods games*, Physical Review E **80** (2009), no. 5, 056109.

[291] Yasuhiro Takeuchi, *Global dynamical properties of Lotka-Volterra systems*, World Scientific, 1996.

[292] Corina E. Tarnita, Hisashi Ohtsuki, Tibor Antal, Feng Fu, and Martin A. Nowak, *Strategy selection in structured populations*, Journal of Theoretical Biology **259** (2009), no. 3, 570–581.

[293] Peter D. Taylor and Leo B. Jonker, *Evolutionary stable strategies and game dynamics*, Mathematical Biosciences **40** (1978), no. 1-2, 145–156.

[294] Josef Tkadlec, Andreas Pavlogiannis, Krishnendu Chatterjee, and Martin A. Nowak, *Limits on amplifiers of natural selection under death-birth updating*, PLOS Computational Biology **16** (2020), no. 1, 1–13.

[295] Christopher K. Tokita, Andrew M. Guess, and Corina E. Tarnita, *Polarized information ecosystems can reorganize social networks via information cascades*, Proceedings of the National Academy of Sciences **118** (2021), no. 50, e2102147118.

[296] Arne Traulsen, Christoph Hauert, Hannelore De Silva, Martin A. Nowak, and Karl Sigmund, *Exploration dynamics in evolutionary games*, Proceedings of the National Academy of Sciences **106** (2009), no. 3, 709–712.

[297] Yuhai Tu, *How robust is the internet?*, Nature **406** (2000), no. 6794, 353–354.

[298] Jay Joseph Van Bavel and Andrea Pereira, *The partisan brain: An identity-based model of political belief*, Trends in Cognitive Sciences **22** (2018), no. 3, 213–224.

[299] John B Van Huyck, Raymond C Battalio, and Richard O Beil, *Tacit coordination games, strategic uncertainty, and coordination failure*, The American Economic Review **80** (1990), no. 1, 234–248.

[300] Federico Vazquez and Víctor M. Eguíluz, *Analytical solution of the voter model on uncorrelated networks*, New Journal of Physics **10** (2008), no. 6, 063011.

[301] V. Volterra, *Variations and fluctuations of the number of individuals in animal species living together*, ICES Journal of Marine Science **3** (1928), no. 1, 3–51.

[302] Friedrich von Hayek, *Individualism and economic order*, University of Chicago Press, 1996.

[303] Philipp Von Hilgers and Amy N Langville, *The five greatest applications of Markov chains*, Proceedings of the Markov Anniversary meeting, Citeseer, 2006, pp. 155–368.

[304] John von Neumann, *On the theory of parlor games*, Mathematical annals **100** (1928), 295–320.

[305] Soroush Vosoughi, Deb Roy, and Sinan Aral, *The spread of true and false news online*, Science **359** (2018), no. 6380, 1146–1151.

[306] Shaoli Wang, Libin Rong, and Jianhong Wu, *Bistability and multistability in opinion dynamics models*, Applied Mathematics and Computation **289** (2016), 388–395.

[307] Xin Wang, Antonio D. Sirianni, Shaoting Tang, Zhiming Zheng, and Feng Fu, *Public discourse and social network echo chambers driven by socio-cognitive biases*, Physical Review X **10** (2020), no. 4, 041042.

[308] Duncan J. Watts and Steven H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature **393** (1998), no. 6684, 440–442.

[309] Steven W. Webster and Alan I. Abramowitz, *The ideological foundations of affective polarization in the U.S. electorate*, American Politics Research **45** (2017), no. 4, 621–647.

[310] D. J. A. Welsh and M. B. Powell, *An upper bound for the chromatic number of a graph and its application to timetabling problems*, The Computer Journal **10** (1967), no. 1, 85–86.

[311] Robert Wilson, *Computing equilibria of n-person games*, SIAM Journal on Applied Mathematics **21** (1971), no. 1, 80–87.

[312] Vicky Chuqiao Yang, Daniel M. Abrams, Georgia Kernell, and Adilson E. Motter, *Why are U.S. parties so polarized? a "satisficing" dynamical model*, SIAM Review **62** (2020), no. 3, 646–657.

[313] Wayne W. Zachary, *An information flow model for conflict and fission in small groups*, Journal of Anthropological Research **33** (1977), no. 4, 452–473.

[314] Damián H. Zanette and Santiago Gil, *Opinion spreading and agent segregation on evolving networks*, Physica D: Nonlinear Phenomena **224** (2006), no. 1-2, 156–165.

[315] John F. Zipp, *Perceived representatives and voting: An assessment of the impact of "choices" vs. "echoes"*, American Political Science Review **79** (1985), no. 1, 50–61.

[316] J. Zoeliner and C. Beall, *A breakthrough in spectrum conserving frequency assignment technology*, IEEE Transactions on Electromagnetic Compatibility **EMC-19** (1977), no. 3, 313–319.

[317] Joshua Zukewich, Venu Kurella, Michael Doebeli, and Christoph Hauert, *Consolidating birth-death and death-birth processes in structured populations*, PLoS ONE **8** (2013), no. 1, 1–7.