Computational Methods in the Study of Political Behavior

Thesis by
Claudia K. Kann

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
Defended May 8, 2023

Claudia K. Kann
ORCID: 0000-0002-8318-4890

ACKNOWLEDGEMENTS

There are so many people to thank for helping me reach this milestone. First and foremost I would like to thank my advisors Mike Alvarez and Jonathan Katz for taking a chance on a Mechanical Engineer who knew little about the Social Sciences. They were a constant source of support and I have learned so much from both of them. In that vein, I would like to thank Marcia O'Malley, my undergraduate advisor for showing me how fun research can be, encouraging me to go to graduate school, and supporting me when I moved on from engineering.

To my partner in life, Josh, thank you for constantly supporting me; for calming me when I was stressed, believing in me more than I believed in myself, and working around my crazy schedule. I so appreciate everything you do for me. To my partner in academia, Maegan, I don't think I could have made it through six years at Caltech without you by my side, and I am so happy we did it together. I am so excited to see the professor you become.

I want to thank my committee for your advice and pushing me to produce better work; Jonathan Katz, Mike Alvarez, Rod Kiewiet, and Phil Hoffman. I also want to thank all of my collaborators and co-authors, those who contributed to the work in this dissertation and those who didn't; Melina Much, Michelle Feng, Danny Ebanks, Mike Alvarez, Jonathan Katz, Zachary Steinhert-Threlkeld, Sarah Hashash, and Jacob Morrier.

Finally, I would like to thank my parents and siblings for always encouraging me, believing in me, and telling me I could do anything I wanted to do, even when they didn't understand what that was.

ABSTRACT

In this thesis, I explore how individual-level actions contribute to aggregate political outcomes. In each chapter, I aim to understand an observed political behavior using data or methodologies previously unused in their contexts. The subject matter ranges from protest activity and vote choice to theoretical opinion models and re-examining how socioeconomic class is understood in quantitative work.

In the first two chapters I employ novel datasets to understand phenomena where popular theories differ from empirical observations. In Chapter 1 I examine protest behavior, which is not the equilibrium prediction of models of collective action. I investigate what aspects of published language can predict protest participation and how these change leading up to and following protests. Specifically, I collect and, using natural language processing methods, analyze 4 million tweets of individuals who participated in the Black Lives Matter protests during the summer of 2020. Using geographical and temporal variation to isolate results, I find evidence that interest in the subject, measured as percentage of online time discussing the matter, is correlated with protest behavior. However, I also find that collective identity, measured through pronoun use, does not have a strong relationship with protest behavior.

Next, in Chapter 2, I use a survey—which I helped to develop and field—to understand the 2020 midterm elections' surprising results. While most accepted models of midterm elections predicted massive Democratic losses (averaging around 40 seats in the House), these predictions were not met. In fact, the Democratic party did well—they did not lose a single state legislature, expanded some majorities, and lost only 9 seats in the House of Representatives. Testing various models of midterm elections, I show that the 2020 midterms were issue-based elections, where views on abortion had a large impact on vote choice.

In the second half of the thesis I focus on methodologies. Specifically, in Chapter 3, I expanded on mathematical models of consensus building to better mimic reality. Bounded confidence models have historically been used to explain convergence of opinions. In this chapter I add a repulsive element, modeling the inclination to differentiate oneself from someone who otherwise has similar beliefs. With this added component, convergence is no longer assumed. I explore both analytical

and simulated numerical results to understand the dynamics of opinions in this new context.

Finally, in Chapter 4, I introduce a method for operationalizing socioeconomic class as a latent variable in regression models. While there has been a plethora of research which shows that class affects opinions, views, and actions, the definition of class is nebulous. I argue that this is a result of the nature of class, which is context dependent. Therefore, rather than explicitly determining class, I present using class within a mixture model framework. This allows for the exact definition of class to change within the context being analyzed and enables researchers to use class within their work. Following the theoretical arguments, I present the efficacy of the approach using the American National Election Studies survey from 2020 to show how class differs when related to views of the U.S. Immigration and Customs Enforcement agency and the Black Lives Matter movement.

PUBLISHED CONTENT AND CONTRIBUTIONS

At the time of submission, none of the work presented in this dissertation has been published. However, portions of the work were done in conjunction with others.

The contents of Chapter 1 includes work done with Zachary Steinert-Threlkeld, Sarah Hashash, and R. Michael Alvarez. The methodology used was presented in Kann et al., 2023 and a version of the chapter is currently under review. In this project, C.K.K. was in charge of data collection, methodology, data processing and analysis and writing.

Chapter 2 is based on work done in conjunction with Daniel Ebanks, R. Michael Alvarez, and Jacob Morrier. A version of the chapter is currently under review. C.K.K. worked on the survey design and implementation, conceptualization and methodology, data preprocessing, data analysis and writing of the chapter.

The work presented in Chapter 3 was done with Michelle Feng. C.K.K. lead conceptualization, writing and simulated results while M.F. was responsible for the proofs of the various theorems.

Finally, Chapter 4 presents work done with Melina Much. Conceptualization and writing were worked on by both authors. C.K.K. was responsible for the methodology and empirical results, while M.M. took leadership on the interpretation of results and situating the approach in previous work.

TABLE OF CONTENTS

Acknowledgements iii
Abstract iv
Published Content and Contributions vi
Table of Contents vi
List of Illustrations ix
List of Tables xiv
Chapter I: Collective Identity in Collective Action: Evidence from the 2020 Summer BLM Protest 1
1.1 Introduction 1
1.2 Collective Identity and Protest Participation 4
1.3 Research Design 8
1.4 Results 19
1.5 Discussion 23
Chapter II: Persuadable Voters Decided the 2022 Midterm: Abortion Rights and Issues-Based Frameworks for Elections 26
2.1 Introduction 26
2.2 Midterm Elections and American Politics 29
2.3 Data and Methods 32
2.4 Results 40
Chapter III: A Repulsive Bounded-Confidence Model of Opinion Dynamics in Polarized Communities 52
3.1 Introduction 52
3.2 Background and Motivation 53
3.3 Model Statement 55
3.4 Analytical Results 58
3.5 Numerical Results on Synthetic Networks 64
3.6 Future Work and Conclusions 72
Chapter IV: Incorporating Latent Class Identities in Quantitative Work 73
4.1 Introduction 73
4.2 Approaches to Class as an Identity and Predictor 76
4.3 Class as a Latent Variable 79
4.4 Empirical Results 83
4.5 Discussion and Conclusion 92
Appendix A: Collective Identity in Collective Action 108
A.1 Data Collection 108
A.2 Topic Analysis 112
A.3 Regression 119
Appendix B: The 2022 U.S. Midterm Election 126
B.1 Survey Question Wording 126

B.2 Crosstabulations 129
B.3 Regression Result Figures 141
B.4 Regression Result Tables 141
B.5 Pooling Tests 151
Appendix C: A Repulsive Bounded-Confidence Model 152
C.1 Additional Proofs 152
Appendix D: Incorporating Latent Class Identities 160
D.1 ANES Details 160
D.2 Stan Implementation 167
D.3 Results 175

LIST OF ILLUSTRATIONS

Number *Page*

1.1 Overview of tweets collected for the summer of 2020. The top panel shows the total tweets collected, the middle panel shows the percent of tweets for each state collected on a date and the bottom panel shows the Google Trends data for the keyword "BLM" in the country as a whole as well as vertical lines for protests which were investigated in this chapter. The grey area represents the time before the murder of George Floyd. 13

1.2 Changes in interest and identity when protesting 21

2.1 House seat change for presidential parties during midterm years. The three times where seats were gained (1934, 1998, and 2002) can be seen. In addition, it is clear the 9 seats lost in 2022, represented by the red dashed line, was a historically mild loss. Source: History, Art & Archives, U.S. House of Representatives, (https://history.house.gov/Institution/Party-Divisions/Party-Divisions/). 27

2.2 This figure shows the weighted percentage of members of each party who voted for congressional candidates of each party. From this figure it is clear that party cohesion was strong—most people voted with their party. This supports the idea that most of the interesting variation will come from those who identify as Independents. Independents as a group make up about half as many individuals as Democrats or Republicans, indicating that they could be the swing in the election. Democrats were significantly more loyal to their party than Republicans. 36

2.3 Vote choice by party identification and views of the national economy and personal finances as compared to the previous year. The width of the bars represent what proportion of individuals (weighted) fell into each grouping. This helps to see how the answers are related to vote choice but also how party identification relates to the response. . 37

2.4 For issues, how different party identifiers voted based on whether they thought it was important or not. The color represents partisan identification while the size is the weighted number of individuals who fit the category. It is clear that most partisans stuck to their party, regardless of their views on issues. Independents were swayed by the issues they viewed as important. 38

2.5 Economic evaluations' effect on voting Republican for pooled and party-based logit models 42

2.6 Average marginal effect (with 95% confidence intervals) of viewing policy areas as important on the probability of voting for the Republican congressional candidate. The plots show the results for the pooled model. 43

2.7 Average marginal effect (with 95% confidence intervals) of viewing policy areas as important on the probability of voting for the Republican congressional candidate. The plots show the results for the party-based model for Independents. 45

2.8 Average marginal effect from thinking abortion, immigration, and violent crime were important on the probability of voting Republican for the 2020 and 2022 survey. There is a clear break in the relationship between importance attributed to abortion and voting decisions for Independents and Republicans. There is no such break for any partisans for immigration or violent crime. 48

3.1 In this figure an Erdős–Renyi Random Graph is created with connection probability of 25%, the evolution three confidence intervals (0.05, 0.2, 0.6) are shown in order to demonstrate the three steady states of the model. 56

3.2 Example with three nodes, where the top and bottom node repulse each other, while the central node attracts both of the others. Without the incorporation of distance scaling, the central node pulls the two outer nodes towards it until they are within confidence of each other, at which point they push each other away until they are no longer within confidence of each other. However, because they are still within confidence of the central node, they are pulled back in, and the cycle repeats so that the model never converges. Using the value of $M_{ij}(t)$ represented in Equation 3.3. 57

3.3 This image shows a simulation where final opinion width was wider than initial opinion width. The edges are created as described in Equation 3.6 where $p_1 = 40\%$, $p_2 = 80\%$ and a confidence bound of 1.6. 58
3.4 An example of the generation of the ER network with attractive and repulsive edges 65
3.5 For all plots, $p_1 = .4$. The horizontal axis represents p_2, while the vertical axis represents c. Note that as p_2 increases, the range of final opinions gets wider. For low values of p_2, as c increases, the range of final opinions becomes narrower (closer to consensus). By contrast, for high values of p_2, as c increases, the range of final opinions becomes wider. 66
3.6 Heat map of opinion spread as a function of the probability of a connection in G_1 and G_2 iterated over confidence bound c. Data drawn from the mean of 100 trials for each set of parameters with parameters $p_1 \in \{0.2, 0.4, 0.6, 0.8, 1\}$, $p_2 \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$, and $c \in \{0.05, 0.4, 0.8, 1.2, 1.6\}$. 66
3.7 An example of the generation of the SBM network with attractive and repulsive edges 68
3.8 Example of the paths taken when the graph is distributed according to the Stochastic Block Model scheme laid out in Equation 3.9. In this case $p_1 = 0.8$ and $p_2 = 0.2$ are both locked the confidence bound varies between the rows and the percent of cross group links varies over columns. Each color represents a different group. There are five groups in these examples. 69
3.9 Heat map of the opinion spread of Stochastic Block Model trials. The rows represent ρ, the percent of positive(negative) connection values that are out (in) group, while the columns are c the confidence bound. 70
3.10 Histograms of metrics at the the start of each trial, useful for comparing the final results. 70

3.11 Proportional spread for Stochastic Block Model trials. The rows represent p_2, the base probability of an out-group negative connection, while the columns are p_1 the base in-group positive probability. The x-axis is ρ the percent of cross group edges and the color represents the confidence bound. A value of 1 means that nodes are equally as close to in and out-group nodes. Smaller values mean groups are increasingly clustered. 71
4.1 Predicted probabilities of class membership for ICE, BLM, and K-modes models. The diagonals show the density of the predicted membership by model. The lower triangle shows the point estimate for the dataset and a generalized linear fit for the correlation (red) as well as a best linear fit (black). The upper triangle shows a heat map of the same information. 85
4.2 Predicted coefficients for class-determining covariates assuming class is split my K-modes or uncovered using a mixture model using the ICE or BLM thermometer as the output. 86
4.3 The estimated class membership when the data is split by a subset of the class-dependent covariates. 1 refers to a 100% probability of having a higher class status while 0 refers to a 100% probability of belonging to the lower class. 87
4.4 Random effects on age for models. 88
4.5 Estimated fixed effects. 89
4.6 Expected ICE thermometer values for strong partisans and independents split by race and gender. 90
4.7 Expected BLM thermometer values for strong partisans and independents split by race and gender. 91
A.1 Coherence metrics for various numbers of topics. 114
A.2 sentTopic WordClouds. 115
A.3 Average distribution of senTopics over time. 116
A.4 Comparison of protest tweets and other tweets by protesters in an effort to validate BLM measure. 117
A.5 Boxplots for the average score for each tweet based on the hand coders based on whether they were within the group closest to 0, 50, and 100 percent related to BLM according to the RJST model. 117
A.6 Tweeting frequencies. 119

B.1 Average marginal effect (with 95% confidence intervals) of viewing policy areas as important on the probability of voting for the Republican congressional candidate. The plots show the results for the party-based model for each party as well as the pooled model results. 141

D.1 Comparison of clustering from k-modes to the estimated group probability from the logit. 169

LIST OF TABLES

Number *Page*

1.1 Example tweets. 11
1.2 Author generated labels for RJST topics 15
1.3 Example calculations: tweets. 18
1.4 Example calculations: RJST output and results. 19
1.5 OLS regression results with day and individual fixed effects. 22
1.6 APEs for logit model with daily fixed effects 23
A.1 Keywords for protesters. 109
A.2 Overview of raw Twitter data: Los Angeles. 110
A.3 Overview of raw Twitter data: Houston. 110
A.4 Overview of raw Twitter data: Chicago. 111
A.5 Correlation matrix between the hand coded response of the four people and then RJST model. It is clear that the unsupervised model is as close to the individuals as they are to each other. This helps to validate our model and lends support to the conclusions drawn using it 118
A.6 OLS regression results with day and individual fixed effects and interaction terms. 121
A.7 OLS BLM regression with control comparison. 122
A.8 Log OLS BLM regression with control comparison. 123
A.9 Percent plural OLS regression with control comparison. 124
A.10 Log percent plural OLS regression with control comparison. 125
B.1 Weighted congressional party vote choice by party ID. 130
B.2 Unweighted congressional party vote choice by party ID. 130
B.3 Weighted demographics by party ID. 131
B.4 Unweighted demographics by party ID. 133
B.5 Weighted most important issue by party ID. 135
B.6 Unweighted most important issue by party ID. 137
B.7 Weighted ability by party ID. 138
B.8 Unweighted ability by party ID. 139
B.9 Weighted views on financial situation by party ID. 140
B.10 Unweighted views on financial situation by party. 140
B.11 Comparison of pooled models. 146
B.12 Final regression results. 150

B.13 We report Wald statistics testing whether party-level interactions are jointly 0 in a pooled model. Large test statistics suggest we can reject the null that coefficients desegregated in the pooled are jointly 0. We instead report the Average Marginal Effects from logit specifications dis-aggregated by party ID. 151
D.1 Number and percent of respondents who responded in each of the eliminated category or were kept in the analysis (0-100). 167
D.2 Estimated mean and standard deviation for logit of k-mode class on class covariates as well as the standard deviation for the priors given to the final response. 170
D.3 Estimates from models when artificially split into two groups. 171
D.4 Class-determining coefficient estimates from mixture models. 175
D.5 Regression results for the Mixture and Pooled models for ICE thermometer rating. 177
D.6 Regression results for the Mixture and Pooled models for BLM thermometer rating. 178

Chapter 1

COLLECTIVE IDENTITY IN COLLECTIVE ACTION: EVIDENCE FROM THE 2020 SUMMER BLM PROTEST

With contributions from Zachary Steinert-Threlkeld, Sarah Hashash, and R. Michael Alvarez.

Does collective identity drive protest participation? A long line of research argues that collective identity can explain why protesters do not free ride and how specific movement strategies are chosen. Quantitative studies, however, are inconsistent in defining and operationalizing collective identity, making it difficult to understand under what conditions and to what extent collective identity explains participation. In this paper, we clearly differentiate between interest and collective identity to isolate the individual level drivers of collective action. We argue that these quantities have been conflated in previous research, causing overestimation of the role of collective identity in protest behavior. Using a novel dataset of Twitter users who participated in Black Lives Matter protests during the summer of 2020, we find that contingent on participating in a protest, individuals have higher levels of interest in BLM on the day of and the days following the protest. This effect diminishes over time. There is little observed effect of participation on subsequent collective identity. In addition, higher levels of interest in the protest increases an individual's chance of participating in a protest, while levels of collective identity do not have a significant effect. These findings suggest that collective identity plays a weaker role in driving collective action than previously suggested. We claim that this overestimation is the consequence of misidentifying interest as identity.

1.1 Introduction

In the summer of 2020, protests erupted in the United States in reaction to the murders of Breonna Taylor and George Floyd. Their deaths embodied the systematic racism Black Americans experience in the United States. These protests sparked continued interest in the Black Lives Matter movement's demands for racial justice. Black Lives Matter (BLM) was officially founded by Alicia Garza, Patrisse Cullors, and Opal Tometi as a Black-centered political movement in 2013 in response to the acquittal of George Zimmerman in the shooting of Trayvon Martin in 2012.[1] While estimating the exact number of people involved in the 2020 Black Lives Matter protests is difficult, they were likely the largest in American history (Buchanan et al., 2020). According to a poll conducted by Gallup between June 23 and July 6, 2020, 11% of American adults said that they had "participated in a protest about racial justice and inequality" in the past 30 days (Long & McCarthy, 2020), indicating a greater level of expressed support than seen for previous BLM protests. The Gallup data indicate that the racial justice and equality protesters were significantly more diverse than previously, with 18% of Black adults, 20% of Asian adults, 13% of Hispanic adults, and 10% of White adults saying they participated (Fisher, 2020; Olteanu et al., 2015). Formal theory predicts that collective action on this scale should be extraordinarily difficult to organize as it involves a collective good—achieving racial justice in the United States (Olson, 1965). What factors explain the widespread participation in the 2020 Black Lives Matter protests?

One mechanism thought to enable participation in collective action at this scale is collective identity; the sense of belonging individuals have to a broader community or institution with a shared perception of group status and goals (Polletta & Jasper, 2001). This group status can originate externally, with outsiders grouping individuals together, such as organizers or entrepreneurs using identities such race, ethnicity, religion, gender, or partisanship as mobilization rubrics. Alternatively, this understanding can originate internally, with individuals seeing that there is a shared sense of purpose or shared ideology. Regardless, by definition, collective identity requires that individuals accept status as part of the group and feel a loyalty to enhancing the status of the group as a whole. By sustaining this sense of belonging and loyalty, working towards the group's goal becomes individually rational and free riding diminishes (Chong et al., 2004; Conover, 1988). Importantly, race in America provides a source of collective identity that has motivated previous episodes of collective action (McClain et al., 2009; Sanchez & Vargas, 2016).

In this chapter, we develop measures that distinguish between collective identity and collective interest when expressed in short online texts. The most common method of operationalizing collective identity is via common hashtag or shared imagery (Driscoll & Steinert-Threlkeld, 2020; Freelon et al., 2016; Metzger et al., 2016). This operationalization, however, approximates a quantity closer to topic interest than to collective identity. In the online world, the focus of this chapter, we

[1] For more details on the on BLM movement and racial inequality in the United States please refer to Bunyasi and Smith, 2019.

define *interest* as discussion of relevant topics, while *identity* by the use of language signifying a sense of belonging (for instance, increased use of plural pronouns such as "we", "us", and "them"). Since choosing to identify with a group gives important insights into the individual's perception of themselves as well as the group's status (Shayo, 2009), this explicit version of collective identity should have a stronger alignment with protest participation than interest. Interest implies that an individual is engaging with a topic or group while identity is a sense of belonging to that group.

We propose a formal model that generates three hypotheses of how true signals of collective action should interact with protest behavior. First, individuals with higher signal values are more likely to protest. Second, individuals should have higher signal values on the day they protest. Finally, going to a protest should increase the signals' value. We test each of these hypotheses using measures of both interest and identity in order to identify if they are signals of collective action.

We test these hpotheses using a new panel dataset of 3,040 Twitter accounts of people likely to have joined BLM protests in Los Angeles, Houston, or Chicago. We then use natural language processing techniques, specifically a Reverse Joint Sentiment Topic model, to analyze each of the accounts' 3.8 million tweets from the summer of 2020, generating separate measures of interest and identity. An ordinary least squares model with day and individual fixed effects is then used to help test the hypotheses derived from the formal model. Results show that contingent on participating in a protest, individuals have higher interest levels the day of and the days following the protest, although this effect diminishes over time. There is a similar pattern for identity, but it is on a smaller scale and has lesser statistical significance. In addition, higher interest in BLM-related topics increases an individual's chance of participating in a protest, while collective identity does not have a significant effect. Overall, these results suggest that for individuals who protest at least once, interest levels have a higher correlation with protesting than identity.

This article joins a growing body of work using digital trace data to understand mobilization around the BLM movement. Social media data has been used to study public opinion about the Black Lives Matter movement (Dunivin et al., 2022), to trace the subtopics discussed (Crowder, 2020; Giorgi et al., 2022; Ray et al., 2017; Tong et al., 2022), as well as to measure the initiation and dispersion of support through social networks (Crowder, 2021; Jackson & Foucault Welles, 2016). These digital studies join a similarly growing body of scholarship that uses offline data, primarily surveys, to understand opinions towards and participation in the movement. Some scholars examine co-ethnic mobilization in support of Black Lives Matter using other pre-existing organizations (Arora & Stout, 2019). Others have similarly used survey data to look at how the protests might have affected public opinion towards police violence (Reny & Newman, 2021; Shuman et al., 2022). Other studies used administrative data to draw the connection between protests and police violence (Williamson et al., 2018) and ethnography to document how other social movements interact with BLM (Petitjean & Talpin, 2022). As far as we are aware, ours is the first study to use social media data to study individual-level motivations for participating in BLM protests.

We proceed as follows. In Section 1.2, we introduce our model of protests, generating expectations of the signals of collective action. In Section 1.3, we describe the research design, including data collection, operationalization of the hypotheses and discussion of our methodology. In Section 1.4, we present our results. Finally, Section 1.5 concludes with a discussion of implications.

1.2 Collective Identity and Protest Participation

Researchers have long struggled to reconcile the reality that large-scale collective actions occur against the theoretical expectation that they should rarely arise, since they involve the production of collective goods (Chong, 1991; Ostrom, 1990; Tilly, 1977). This disconnect between theory and reality has led to considerable theorizing about incentives for individual involvement in collective action, such as protest as a consumable good (Tullock, 1971) and concerns about reputation (Gerber et al., 2008). These alternative theories are similar in that they both posit that individuals are less concerned with their potential to be pivotal. Instead, incentives for private benefits can cause participation in collective action. These incentives can also arise from being part of community, notions of morality, the emotions evoked by collective participation, or having a collective identity (Gause, 2022; Jasper, 1997; Johnston & Klandermans, 1995; Miller et al., 1981; Sanchez, 2006; Stokes, 2003). Collective identity, which we focus on, is a measure of the extent an individual feels like they belong to a group. The original linkage between collective identity and collective action arose when scholars usied the former to explain otherwise nonrational behaviour (Fireman & Gamson, 1977; Teske, 1997). These theories suggest that there is a private benefit to individuals for participating when they feel to be a part of a community.

In addition, this concept is closely related to that of linked fate. These theories suggest that the prominence in racial identity in the social stratification, particularly within the United States, has lead to an increased perception of how the status of the group as a whole impacts the individuals status (Dawson Michael, 1994; McClain et al., 2009; Tate, 1994). While the theories developed to tie collective identity and collective action together, we implicitly assume a degree of linked fate in order to drive the process. In this chapter we are not measuring the degree to which individuals believe the status of the group will affect them as individuals (linked fate), but how strongly they associate with the group (collective identity). This distinction is necessary in order for our measure to make sense.

Going forward, we assume that we have individuals $i \in \{1, ..., I\}$ and days $t \in \mathcal{T}$. In addition, for each individual-day pair we have a collective action signal value $y^*_{it} \in (0, 1)$ for which higher values imply a stronger signal value (i.e., higher levels of interest or collective identity). Finally, we also have an indicator on whether or not individual i protests on day t represented by $x_{i,t}$.

For the original turnout game, we assume that individuals contribute to a public good, such as protesting, when their net utility is nonnegative. If a threshold (q) is met then everyone receives the public good (a policy change resulting from a large enough protest), if not, no one does. For the most basic model, we assume that everyone has the same cost (c) of protesting and benefit (β) from the subsequent policy change if enough individuals protest ($\mathbb{1}$). The utility for protesting is thus:

$$u_i(x_i) = \beta \mathbb{1}_{\sum_i x_i \geq q} - c x_i. \tag{1.1}$$

In this case, since everyone is identical, we look for symmetric equilibria. The symmetric equilbria are mixed strategy responses, that is everyone has a probability p of protesting. For a mixed strategy, we need the payoff for protesting to be the same as not protesting. Thus, we have that the cost to protesting must equal the benefit times the probability that the individual is pivotal. Generally, the probability of being pivotal is so small that the benefit must be massive or the cost minuscule.

In our version of the game, we argue that individuals have a private individual benefit (y^*_{it}) from the act of protesting at time t, which represents the pure benefits from collective identity or interest. For this, we take inspiration from the global games literature which studies games in which actions are influenced by the uncertain actions of others (Bueno De Mesquita, 2010; Little, 2016; Shadmehr & Bernhardt, 2011). In that case, each individual's utility function can be rewritten as

$$u_{it}(x_{it}) = \beta \mathbb{1}_{\sum_j x_{jt} \geq q} - \underbrace{cx_i + y_i^* x_i}_{c_i x_i}. \tag{1.2}$$

For the sake of simplicity, we assume that y_{it}^* is normally distributed, however for any known distribution the proof continues in the same manner. If we assume a cutoff strategy, such that individuals protest if their individual cost is less than some value k^*, then we can solve for this cutoff by solving the equation:

$$\binom{n}{q-1} \Phi(k^*)^{q-1} (1 - \Phi(k^*)^{n-q+1} \beta = k^*. \tag{1.3}$$

In reality, however, the measures we are observing are noisy signals for identity and interest, therefore, the value we see is instead

$$y_{it} = y_{it}^* + y_t + \epsilon_{it} \tag{1.4}$$

where y_t is a daily fixed effect and ϵ_{it} is the, normally distributed, daily noise given the individual. With this information, we have the probability that the true value is greater than the cutoff increases with the measured value. This leads us to our first hypothesis.

Hypothesis 1 (H1)**.** *Individuals who have larger private individual benefits (which correspond to higher valued signals of collective action—interest or identity), leaving all else equal, are more likely to participate in protest behavior.*

$$P(x_{i,t} = 1 | x, y_{i,t-1}) \geq P(x_{i,t} = 1 | x, y'_{i,t-1}) \iff y_{i,t-1} \geq y'_{i,t-1} \tag{1.5}$$

This hypothesis is further supported by existing empirical work. The earliest work, of which we are aware, in which collective identity of this form is measured is provided by Matthews and Prothro, 1966, in which they use two different survey questions to ascertain the closeness their Black participants had to the community as a whole. Subsequent work found that higher levels of group consciousness in Black Americans correlated with higher levels of political participation, generally forms of collective action (Olsen, 1970; Verba & Nie, 1987). If these theories and studies hold, then people with higher levels of collective identity should be more likely to protest, holding all else equal. There has been less work suggesting that interest has a similar role to play.

In addition to the assumption that those with higher valued signals of collective action have an increased likelihood of protesting, we present two more hypotheses. These are derived from the assumption that individuals are the average of their social network in terms of opinions and associations (Hegselmann & Krause, 2002; Siegel, 2009). Thus, given some network $\mathcal{I}$ which represents the contacts of individual i, we have that

$$y_{it} = \frac{1}{|\mathcal{I}|}\sum_{j\in\mathcal{I}} y_{jt}. \tag{1.6}$$

There has been work that suggests that the act of protesting reinforces existing identity through the interactions with other like-minded individuals (Madestam et al., 2013). The protest facilitates the creation of new network structures that expand the reach of the movement and density of connections. Therefore, we propose that protesting should increase the values of the signals observed.

Hypothesis 2 (H2)**.** *The act of protesting solidifies an individual's support for the cause, increasing the expected levels of the signals of collective action observed for the days following the protest action compared to the non-protesting expectation.*

$$E[y_{i,t+j}|x_{i,t} = 1] > E[y_{i,t+j}|x_{i,t} = 0], \forall j \in \{1, ...N\} \tag{1.7}$$

This is driven by the fact that the new connections created by protesting should have higher levels of the signals in the mean (as they are selected to be above a signal threshold). As a result, given new connections $\tilde{\mathcal{I}}$ who, on average, have higher signal values, the average signal value of an individual's connections will increase. This, in turn, increases their signal value.

$$\begin{aligned}
y'_{it} &= \frac{1}{|\mathcal{I}| + |\tilde{\mathcal{I}}|}\left(\sum_{j\in\mathcal{I}} y_{jt} + \sum_{j\in\tilde{\mathcal{I}}} y_{jt}\right) \\
&\geq \frac{1}{|\mathcal{I}| + |\tilde{\mathcal{I}}|}\left(\sum_{j\in\mathcal{I}} y_{jt} + \frac{\mathcal{I}}{\tilde{\mathcal{I}}}\sum_{j\in\mathcal{I}} y_{jt}\right) \\
&= \frac{1}{|\mathcal{I}|}\sum_{j\in\mathcal{I}} y_{jt} \\
&= y_{it}.
\end{aligned}$$

For our final hypothesis, it seems reasonable to assume that individuals who are protesting will exhibit higher than usual amounts of discussion during that time. This sets us up to formally present our final hypothesis:

Hypothesis 3 (H3)**.** *The act of protesting increases the expected levels of collective action signals observed during that day compared to the non-protesting expectation.*

$$E[y_{i,t}|x_{i,t} = 0] < E[y_{i,t}|x_{i,t} = 1] \tag{1.8}$$

These three hypotheses, all which build off of previous work, will enable us to test whether collective identity and collective action operate in the way found by previous literature, or if interest has instead been a more appropriate indicator.

1.3 Research Design

According to the Armed Conflict Location and Event Dataset (Raleigh, 2010), between May 26th and August 22nd, there were over 7,750 BLM demonstrations in over 2,440 locations in all 50 states. These protests were some of the most well attended and longest lasting in American history (Putnam et al., 2020). Thus, they created an environment ideal for analyzing the dynamics of collective identity as a movement develops and spreads. In this section, we discuss how we collect and analyze our unique social media dataset. First, we selected three cities for analysis and found Twitter users we classify as protesters. These accounts are classified as protesters if they were likely at protests in their city based on keywords and location provided from Twitter. We then collected the entire Twitter timeline for each of these protesters for the summer of 2020. In order to measure both signals of collective action, we estimated a Reverse Joint Sentiment Topic (RJST) model, which is a weakly supervised natural language processing model. Finally, we use the results from the RJST model to test our hypotheses for how collective identity and interest each may have helped resolve the collective action problems facing the BLM movement. We discuss each step in detail in the following sections of our chapter.

Data Collection

We choose to analyze the BLM movements in Los Angeles, Chicago, and Houston. Cities were not chosen for geographic or political reasons, as we do not expect the role of identity to vary based on the location or median preferences of a city. Instead, we chose to focus on three of America's four largest cities because they account for a significant number of protests and participants during the period of this study.[2,3]

[2] New York City is excluded because the amount of data would have introduced significant data storage issues and computational complexities.

[3] These three cities make up 14% of the protesters and 2% of the protests accounted for by the CCC during this time. Houston made up 9% of the people but only 0.2% of the protests while Los Angeles and Chicago were both about 2% and 1% for protesters and protests, respectively.

Having determined locations to analyze, the next decision involved data collection. Social media was chosen over participant observation or surveys because they give researchers the ability to observe individuals before, during, and after treatment across disparate locations at much lower cost than in-person studies and do not require researcher foreknowledge of an event. In addition, the generation of social media, which occurs outside of the purview of researchers, implies there is no observer effect. Surveys face difficulties that arise from the spontaneity of these events; they are often not known far enough in advance for a research group to pull together a proposal and get the funding and individuals in place to create an effective survey. In addition, people at a protest are often uninterested in responding to a long list of questions when they are focused on their bigger goal. Finally, it is difficult to sample research subjects for surveys conducted at a protest location in a way that produces a scientifically representative sample.[4]

These issues in the collection of data can easily lead to biased responses (Westwood et al., 2022). Additionally, survey methods are unable to dynamically track these values over time (Chenoweth et al., 2022). Even in the case of panel data, the researchers have at most two or three points for each individual over time. Most importantly, perhaps, is that they rarely have information on the individuals before the first protest, and are thus unable to compare how the protest affected them, and whether those effects were lasting. These shortcomings make real time and in-person data collection almost impossible, especially for large scale protests. By using social media data, we are able to retroactively access the conversations of protesters before they protest, giving us a baseline for their activities prior and subsequent to their action. In addition, the nature of the 2020 BLM protests means that we were able to obtain data from a series of protests from the same locations and with the same basic subject matter, but over a varying period of time. A major benefit of collecting time series cross section (TSCS) data is we can factor out day-specific effects. Finally, there has been significant research connecting the use of social media with protest behavior (Valenzuela, 2013) making it an appropriate venue for this work.

From the universe of social media platforms, Twitter is best suited for this research. It is a widely used social media platform, with individuals who use it frequently checking their feed (Duggan & Smith, 2013). In addition, it is used both to coordinate political activities and to discuss everyday events, giving us a more complete picture of the individuals (boyd et al., 2010). Twitter has also emerged as a primary

[4]While this is also the case for Twitter which is a skewed subset of the population as a whole, our results do hold for the Twitter population.

tool used by social movement organizers to engage individuals in collective action (Clark-Parsons, 2022). Importantly for this study, while only 13.5 percent of the United States population is Black, they make up 25 percent of users on Twitter (Brock, 2012), which allows us to more heavily weigh the population for whom this movement is most likely to be salient. In addition, there has already been substantial research using Twitter use to study the BLM movement (Cox, 2017; Freelon et al., 2018; Ince et al., 2017) which provide references to compare our results with. Researchers have also used Twitter to study protests across the globe, in autocracies and democracies (Burns & Eltham, 2009; Larson et al., 2019; Rahimi, 2011; Steinert-Threlkeld, 2017), for the study of the Black Lives Matter movement in the United States (Hsiao, 2021; Ray et al., 2017), and for the study of feminist social movements like MeToo (Clark-Parsons, 2022). Finally, Twitter is easily accessible with two different APIs which allow researchers to systematically find tweets and users relevant to a particular study as well as collect data on tweets and users relating to the movement.

There are, however, some important concerns about measuring collective identity using social media data. The nature of the data means that we do not have access to relevant sociodemographic information which would ideally be used in determining collective identity strength. The conclusions drawn, in addition, can only be applied to other Twitter users who geotag their Tweets. Social media in general provides a sample that is not necessarily representative of the population as a whole and geotagged tweets make up less than 1% of total tweets (Ajao et al., 2015). There is work suggesting that users with geotagged Tweets are statistically different than those who do not (Karami et al., 2021); despite these shortcomings, we believe that the data used has fewer limitations than other data sources. Finally, it is worth noting that in this case we are selecting on the dependent variable. We only observe individuals who protest at least once. In future work, we hope to include a baseline of non-protesters as well.

For this study, a protester is defined as anyone who uses keywords related to the Black Lives Matter movement from Los Angeles, Houston, or Chicago during a subsample of those cities' summer 2020 protests. The list of keywords and their justification can be seen in Table A.1. In Table 1.1 we provide a sample of tweets collected this way and used to identify individuals as protesters.[5] Other research

[5]While we refer to each Twitter user as an *individual*, we understand that organizations play a large and important role in Black social movements. However, the differentiation between individual and organization is beyond the scope of this project.

has used cell phone location records to verify the accuracy of using keywords to measure protest participation in the United States (Sobolev et al., 2020). Generating panel data after an event possibly introduces selection bias since accounts could become private, delete tweets, or no longer exist on Twitter. We are aware of no research quantifying this decay rate, but studies using Twitter and Facebook in China, Colombia, and Uganda have found no differences in results when comparing this chapter's method to data collected in real time (Boxell & Steinert-Threlkeld, 2022; Chang et al., 2022; Morales, 2021).

Table 1.1: Example tweets.

City	Date	Text
Los Angeles	05-28	We're posted on hill and 2nd street downtown Los Angeles #GeorgeFloyd #BlackLivesMatters #ICantBreathe
Los Angeles	05-28	Over 100 protesters facing down cops in the 2nd st tunnel #downtownLA #losangeles #protest #GeorgeFloyd
Los Angeles	06-06	#BlackLivesMatter protests all around Santa Monica yesterday including a march from the Venice pier to the Santa Monica pier, a paddle out of surfers, and a protest in front of city hall for #GeorgeFloyd (not my photos)
Chicago	05-29	They never cared... they never will... this is America #LandOfTheFree #GeorgeFloyd #Minnesota #PoliceBrutality Chicago, Illinois
Chicago	05-29	Take a walk with us tomorrow virtually or in person (we do have one or two places available for those of you ready to venture out) deep listening in the neighborhood olivagallery vankanegan I'll be streaming live. . .
Chicago	09-24	A couple hundred people are gathered in Chicago's Palmer Square Park to demand justice for Breonna Taylor. Small groups are talking among themselves.
Houston	05-29	Thousands of people here. Eerily quiet as people stream towards City Hall. People are angry, as we should be. Peaceful so far. #blacklivesmatter #georgefloyd Houston City Hall
Houston	05-30	Before the arrests tonight in Houston Chief Art Acevedo was right in the middle of #GeorgeFloyd demonstrators saying its about holding bad cops accountable
Houston	06-02	Discovery Green is filling up and numbers are expected to be in the thousands We joined because every voice for justice counts We wanted to add ours physically and verbally #georgefloyd #racialjustice #Discovery

These tweets and the associated users were found using the Version 2 Twitter API and the Python package TwitterAPI.[6,7] These tools allow us to enter a time period, location bounding box around the protest city, and keywords to search for and

[6] https://developer.twitter.com/en/docs/twitter-api
[7] https://github.com/geduldig/TwitterAPI

return the desired information for all tweets that meet the criteria. For this project, we requested the author ID, time the tweet was written, geolocation information (which can be in the form of coordinates, a bounding box, or a city name), public metrics (likes, retweets, etc.), entities (hashtags, mentions, symbols, and URLs), and the tweet text. We choose protests listed in the Crowd Counting Consortium (Chenoweth & Pressman, 2017). From Los Angeles, we choose 14 protests from which we draw 2,348 protesters, from Houston we have 273 protesters from 8 protests, and from Chicago we have 391 protesters from 24 protests (see Tables A.2-A.4 in the Supplementary Materials).

Next, we downloaded all available tweets from each protester from May 20th 2020 until October 1st 2020 using the package gatherTweet (Kann et al., 2023).[8] We again used the Version 2 Twitter API and TwitterAPI to pull the entire timeline for all of these accounts. These tweets provide the conversations of all the selected individuals from five days before the murder of George Floyd through the end of the summer. Figure 1.1 shows the number of tweets we collected on each day from each city. While there are significantly more tweets from Los Angeles than the other two cities—a result of larger protests in Los Angeles than the other two cities—when we look at the distribution of tweets they follow similar patterns. These approximate similarities between the cities provides preliminary support for the assumption that we can pool the protests from the three cities in our analysis. Tables A.2-A.4 show summary statistics for the protests.

Ethical Considerations

The collection and analysis of the data we use in this chapter was reviewed and approved by the Institutional Review Board at the California Institute of Technology. In this study, we did not ask Twitter users for permission to observe their Twitter history or use this data in our analysis. This is consistent with other work using similar social media data. By joining Twitter and using a public account, individuals are accepting the Twitter Terms of Use that specifically state that their content is public information. There is an additional concern, however, that use of Twitter data in research or publishing tweets with identifying information could put users at risk. While tweets, in general, are public information, users may expect that their public tweets will remain within their individual social sphere. Thus, if

[8]The data was collected roughly a year after the protests occurred, in that time if people delete their tweets or accounts the tweets will not show up in our dataset. In addition, some accounts are set to private. Those tweets and accounts will also not show up in our set.

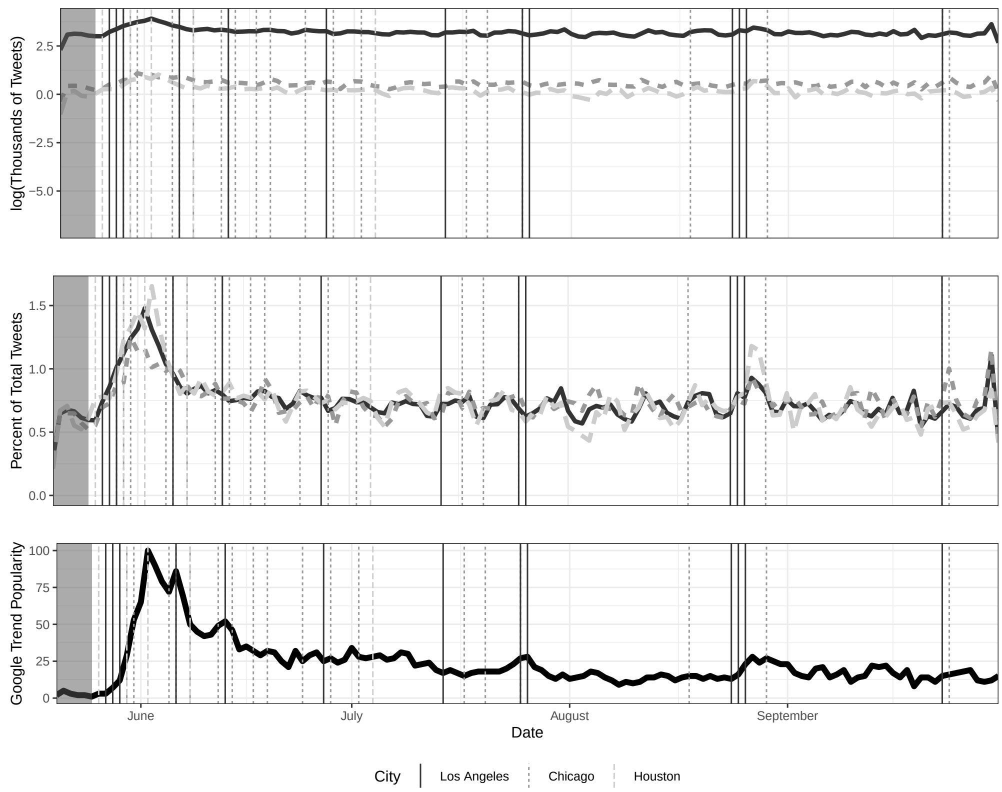


Figure 1.1: Overview of tweets collected for the summer of 2020. The top panel shows the total tweets collected, the middle panel shows the percent of tweets for each state collected on a date and the bottom panel shows the Google Trends data for the keyword "BLM" in the country as a whole as well as vertical lines for protests which were investigated in this chapter. The grey area represents the time before the murder of George Floyd.

researchers expose the views of vulnerable individuals in their research, it could lead to harassment or retaliation. This is particularly a concern when the topic is polarizing and contentious or the individuals in question belong to a group that has a history of being exploited. In this study, we use three strategies to mitigate these risks. First, the social media data we collected is analyzed and presented at the aggregate level—we do not present nor publish individual tweets along with identifying information. Second, we do not attempt to discover the true identities of the users. Finally, upon publication we will share only the tweet identification numbers, consistent with the terms of academic use of these data. A final concern comes from using the geolocation information provided. Again, users do choose how much of their location to share, a setting that can be changed for each tweet individually or for the account as a whole. However, we still do not overly identify individuals based on their location. While we use the location of tweets to identify protesters, we do not use the information beyond that point and never track the movement of the individuals. For the timeline tweets, we don't pull the geolocation information.

Reverse Joint Sentiment Topic Analysis

Text data in general is difficult to quantify. Our raw data consists of 3,810,307 pieces of text. In order to test our hypotheses, we need to find a way of reducing the dimensionality of our text data. We do this by classifying the tweets as belonging to certain clusters. Specifically, we use a Reverse Joint Sentiment Topic Model (RJST) as presented in Lin et al., 2011 to define each tweet by a lower dimension topic and sentiment. RJST works by finding clusters of words that are used frequently together in order to define groupings. RJST, while based on a Latent Dirichlet Allocation (LDA) model, includes a second latent layer that allows us to account for additional structure that the simple LDA model may overlook. A detailed discussion of RJST, our results, and the diagnostics regarding topic selection and validation can be found in Appendix A.2.

The final model used generates 5 topics and 3 sentiments for a total of 15 groupings. The list of author-generated labels for each group can be seen in Table 1.2. For each tweet, there is a probability measure θ which represents the proportion of the tweet belonging to each topic. Within each document and topic, there is a probability measure π which represents the distribution of sentiment within each topic in the document. Thus, by multiplying the probability measures we are able to get a value for how much of each tweet is in each topic sentiment pair (for instance $\theta_1\pi_{12}$ is how much the tweet is in Topic1Sentiment2). These values will be important for analyzing the content of the tweets going forward. In addition, we label the four senTopics which begin with "BLM" as the relevant topics for the analysis; these topics will form the foundation for our analysis.

The validity of these labels is tested in multiple ways, the details of which are presented in the Supplemental Materials. First, we look at the distribution of the topics over time, the topics labeled as related to BLM clearly follow the same pattern as the Google Trend data on the topic. This can be seen in Figure A.3. Next, we look at the percent related to BLM the tweets are which were found using keyword and location information and compare it to the distribution of those in the individuals timelines in general. The results, seen in Figure A.4 show that those tweets we know are related to BLM score high, while the overall tweets are distributed much lower.

Finally, we took a sample of 800 tweets and had four individuals rate the percent they believe the tweet is related to BLM, the results can bee seen in Figure A.5. The correlation between the RJST result and the average hand labeling is 80%. Overall, these three tests lead us to be confident in the RJST model accurately labeling the relevance of tweets to the BLM movement.

BLM	Label
Yes	BLM George Floyd/Breonna Taylor
Yes	BLM General
Yes	BLM City News
Yes	BLM Police Violence
	Public Programs
	Vote General
	Pop Culture
	Media
	Covid Believers/Wear Masks
	Political Confrontation
	Sadness/Nostalgia
	Music
	2020 Presidential Election
	Family
	Anger/Frustration

Table 1.2: Author generated labels for RJST topics

Operationalizing the Hypotheses

In order to convert our data into a form that is suitable for testing the hypotheses, we need to determine the best method of both quantifying the various terms as well as determining the expected values. For each tweet collected, we label individuals as having protested for those days in which their tweets are collected. For all other protests, we mark the individuals as not protesting. This binary variable is the most straightforward we use. The protest dates, the number of protests drawn and the estimated size of each protest can be seen in Tables A.2–A.4.

We will now walk through the general application of our operationalization, and in the next section we will show the same calculations using an example set of tweets. Given that on day t individual i tweets N times, for each $n \in \{1, ..., N\}$, we have a topic distribution $\theta_{n,t,i} \in R^5$ and a sentiment distribution for each topic in each tweet $\pi_{n,t,i,\ell} \in R^3$. In order to get the senTopic distribution, we multiply the sentiment distributions by the corresponding element in the topic distribution. For each tweet, we then take the mean of the sums of the senTopic distributions multiplied by a BLM indicator, and this gives each tweet an interest score. Thus in general we have:

$$y_{i,t,n}^{interest} = \sum_{\ell=1}^{5} \sum_{k=1}^{3} \theta_{n,t,i}(\ell)\pi_{n,t,i,\ell}(k)\delta_{\ell,k}. \tag{1.9}$$

Specifically, for our data we have that $\delta_{\ell,k} = 1$ for the pairs $(1,1),(1,2),(4,2)$, $(4,3)$ and is zero for the rest. Therefore, the score for each tweet is the sum of the BLM scores:

$$y_{i,t,n}^{interest} = \theta(1)\pi_1(1) + \theta(1)\pi_1(2) + \theta(4)\pi_4(2) + \theta(4)\pi_4(3) \tag{1.10}$$

In order to get the daily score, we take the average score for the day:

$$y_{i,t}^{interest} = \frac{1}{N}\sum_{n=1}^{N} y_{i,t,n}^{interest}.$$

This value represents how much of an individual's daily Twitter time is devoted to discussion of BLM—their daily interest. It is the average BLM score of their daily tweets. In order to find their identity scores, a measure of how closely they identify with the Black Lives Matter movement, we look at the levels of explicit group belonging in the topic-related tweets. We will call this variable $y_{i,t}^{identity}$. This value is found by first categorizing the percent of the pronouns in each tweet that are plural, $c_{n,t,i} \in (0,1)$. This tweet level value is a representation of how closely an individual identifies with the subject matter of the tweet. We then take the weighted average, using the interest score over the tweets for each day, to observe to what extent the individual is discussing the topic of the protests as part of the group rather than as the individual. This lets us weight tweets that are more relevant to the topic more.

$$y_{i,t}^{s} = \frac{\sum_n c_{n,t,i} y_{i,t,n}^{interest}}{N y_{i,t}^{interest}} \tag{1.11}$$

These daily scores are our values of interest as we proceed. To test our three hypotheses, we look at how these scores interact with protest attendance. First, for Hypotheses 2 and 3, we run a time and individual fixed effect OLS model with indicators for the relative date of the tweet compared to a protest event the individual participated in if the relative date is between –4 and 4 inclusive. Thus, given that an individual protests at time τ we are solving for:

$$y_{i,t}^{interest(s)} = \alpha_0 + \alpha_1\delta_{t=\tau-2} + \alpha_2\delta_{t=\tau-1} + \alpha_3\delta_{t=\tau} + \alpha_4\delta_{t=\tau+1} + \tag{1.12}$$

$$\alpha_5\delta_{t=\tau+2} + \eta_i + \beta_t + \epsilon_{i,t}. \tag{1.13}$$

The values for α_{1-5} represent the change in signal value if the individual protests at relative time 0 compared to the counterfactual that they did not protest. Statistically significant positive values for α_4 will provide evidence in support of Hypothesis 2. If α_3 is positive and statistically significant, this provides evidence in support of Hypothesis 3.

To test Hypothesis 1, we create a prediction of whether an individual protests based on their signal values, that is the daily *interest* and *identity* values for each user. In order to do this, we use a logit model with day and individual fixed effects. First, we segment the data to only include days in which protests occurred—this is to prevent null results on the days in which protests do not occur. We then run the model solving for:

$$Pr(x_{i,t} = 1) \propto \Phi(\eta_i + \beta_t + \alpha_0 + \alpha_1 y_{i,t}^{interest} + \alpha_2 y_{i,t}^{identity}). \tag{1.14}$$

The value and significance of α_1 and α_2 indicate the effect of the levels of these signals on protesting.

Example Tweet Calculations

In order to clarify the process above, we will talk through how the values would be calculated for three tweets. In addition, this should help illustrate how the RJST algorithm classifies tweets. We begin with three tweets from our sample (displayed in Table 1.3). These tweets are all from the same user, however we have picked them specifically to suit our exercise.

Reading these tweets, it is clear that Tweets 1 and 2 are related to Black Lives Matter while Tweet 3 is discussing Covid. We therefore expect 1 and 2 to be high on the interest score and 3 to be low. Tweet 1 should also score high on collective identity—the user is identifying with the group claiming, "People are angry, as **we** should be" (emphasis added). On the other hand, Tweet 2 seems more observational, and we would expect it to score lower in terms of identity. Finally, while Tweet 3 is not related to BLM, with respect to being a Houstonian, there is a high level of identity. Ideally, our algorithm should filter this out when applying the weighting.

The RJST model outputs the percent that each tweet falls into each topic and within each topic, and each sentiment. The scores for each of the three example tweets

TweetId	Tweet Text
1	Thousands of people here. Eerily quiet as people stream towards City Hall. People are angry, as we should be. Peaceful so far. #blacklivesmatter #georgefloyd @ Houston City Hall
2	Philly police chief to the 57 Buffalo police officers who resigned in protest over the two officers who are now going to be criminally charged for shoving a man to the ground and then ignoring his injury: "BYE FELICIA". feliciaforever
3	Hi Houston, please listen to this doctor in charge of a COVID-19 unit tell us what is happening and masks and reopening schools is deadly. Houston Hospital Struggles To Manage Surge Of COVID-19 Cases

Table 1.3: Example calculations: tweets.

can be seen in Table 1.4. The BLM related sentiment topic pairs are bolded. From looking at the distributions, we can see that Tweet 1 is related to the city news category while Tweet 2 is related to the George Floyd/Breonna Taylor topic as well as the police violence one. Tweet 3 is almost entirely related to Covid. These characterizations are sensible when looking at the content of the tweets and these examples give us confidence in the reliability of our topic modeling. Adding up the distributions in the BLM labeled topics, we get the tweet level interest value (y_{itn}). The tweet level identity scores are also as expected—Tweets 1 and 3 are high while Tweet 2 is low.

Assuming these three tweets came from a single day, and they were the user's only tweets for the day, the daily interest and identity scores are calculated as:

$$y_{it}^{interest} = \frac{1}{N}\sum_{N} y_{itn} = \frac{1}{3}(0.976 + 0.981 + 0.006) = 0.654 \tag{1.15}$$

$$y_{it}^{identity} = \frac{\sum_N c_{itn} y_{itn}}{\sum_N y_{itn}} = \frac{1 * 0.976 + 0 * 0.981 + 1 * 0.006}{0.976 + 0.981 + 0.006} = 0.500. \tag{1.16}$$

Both of these scores make sense when looking at the three tweets chosen. About 2/3 of the tweets are clearly related to BLM. In addition, of the tweets that are related to BLM, TweetID 1 has what would be considered a strong collective identity score while the other is weak. The identity values of non-BLM related tweets should barely come into play.

TweetId	**George Floyd/ Breonna Taylor** $\theta_1\pi_{11}$	Vote General $\theta_2\pi_{21}$	Covid $\theta_3\pi_{31}$	Music $\theta_4\pi_{41}$	2020 Pres Elec $\theta_5\pi_{51}$
1	**0.002**	0.002	0.002	0.002	0.002
2	**0.651**	0.002	0.002	0.002	0.002
3	**0.001**	0.001	0.980	0.001	0.001

TweetId	**BLM** $\theta_1\pi_{12}$	Pop Culture $\theta_2\pi_{22}$	Political Confrontation $\theta_3\pi_{32}$	**City News** $\theta_4\pi_{42}$	Family $\theta_5\pi_{52}$
1	**0.002**	0.002	0.002	**0.969**	0.002
2	**0.002**	0.002	0.002	**0.002**	0.002
3	**0.001**	0.001	0.001	**0.001**	0.001

TweetId	Public Programs $\theta_1\pi_{13}$	Media $\theta_2\pi_{23}$	Sadness Nostalgia $\theta_3\pi_{33}$	**Police Violence** $\theta_4\pi_{43}$	Anger Frustration $\theta_5\pi_{53}$
1	0.002	0.002	0.002	**0.002**	0.002
2	0.002	0.002	0.002	**0.326**	0.002
3	0.001	0.001	0.001	**0.001**	0.001

TweetId	Identity c_{itn}	Interest y_{itn}	Interest $y_{it}^{interest}$	Identity $y_{it}^{identity}$
1	1.000	0.976		
2	0.000	0.981	0.654	0.500
3	1.000	0.006		

Note: The bolded values are those that are categorized as part of the BLM discussion. The Identity and Interest columns are calculated at the individual tweet level (c and y_{itn}) as well as if the three tweets were the individuals corpus for the day ($y_{it}^{interest}$ and $y_{it}^{identity}$).

Table 1.4: Example calculations: RJST output and results.

1.4 Results

We now estimate and discuss the tests outlined in the previous section. We find that interest strongly supports all three hypotheses. In addition, hypotheses 1 and 2 are supported by identity, although the magnitude of the results are smaller. In order to verify that any significant result is not spurious, we also create two placebo tests by setting the protest day to 10 days prior and subsequent to actual protest action. The OLS results can be seen in Table 1.5, while the placebo tests can be seen in

Tables A.6-A.9. Throughout the rest of this section we go into further detail on these results.

In the regression with interest as the dependent variable, where interest is what percent of an individual's daily tweets are in the topics labeled as about the BLM movement, we see significant positive results the day before, day of, and two days after the protest. Following this, the results become insignificant. In addition, the F statistic is significant at the 0.01 level, indicating a good fit of the model. This result suggests that individuals spend about 1.4% more of their Twitter time discussing BLM the day before they protest than they would if they were not going to protest. On the day of a protest, their interest level is on average 6.7% more relevant than it would be otherwise (supporting Hypothesis 2 for interest) and 10% more relevant the day after (supporting Hypothesis 3). By two days after, there is still an increase (3.4%), but the interest level is returning back to non-protesting levels. While we see that in the location-pooled model there is a sustained increase three and four days after the protest, when including interaction terms for protest location, this result varied by location. The significant results for the fully interacted model can be seen in Appendix A.3. As the average amount the sample talks about BLM in the time period ranges from about 20-60%, we view these results as substantially significant in addition to statistically significant.

In addition to the interest-level dynamics related to the hypotheses, it is interesting to note that before protesting, people begin tweeting slightly more about the topic. On the day of the protest, the amount they talk about BLM increases substantially. This trend continues through the day after the protest, after which the results begin to dissipate. When the same test is ran for a placebo protest date 10 days before the real protest, none of the results are significant. When the test is run around relative day 10 there are still some slight increases on days 8 and 9 (1.4% and 1.3%, respectively), but these values are only significant at the 0.1 level. Overall, the results combined with the placebo test supports both Hypotheses 2 and 3 for interest.

We next turn our attention to the results for identity. There is a 1.6% increase of identity the day of protests given protesting. This result is significant at the 0.05 level. While a smaller increase than for interest, the placebo test produces null results. This increases the credit given to the small jumps in identity the day of the protest. There are no significant results for the rest of the protest-relative days. In Section 1.4, we plot the coefficient values around the date of protest and report the

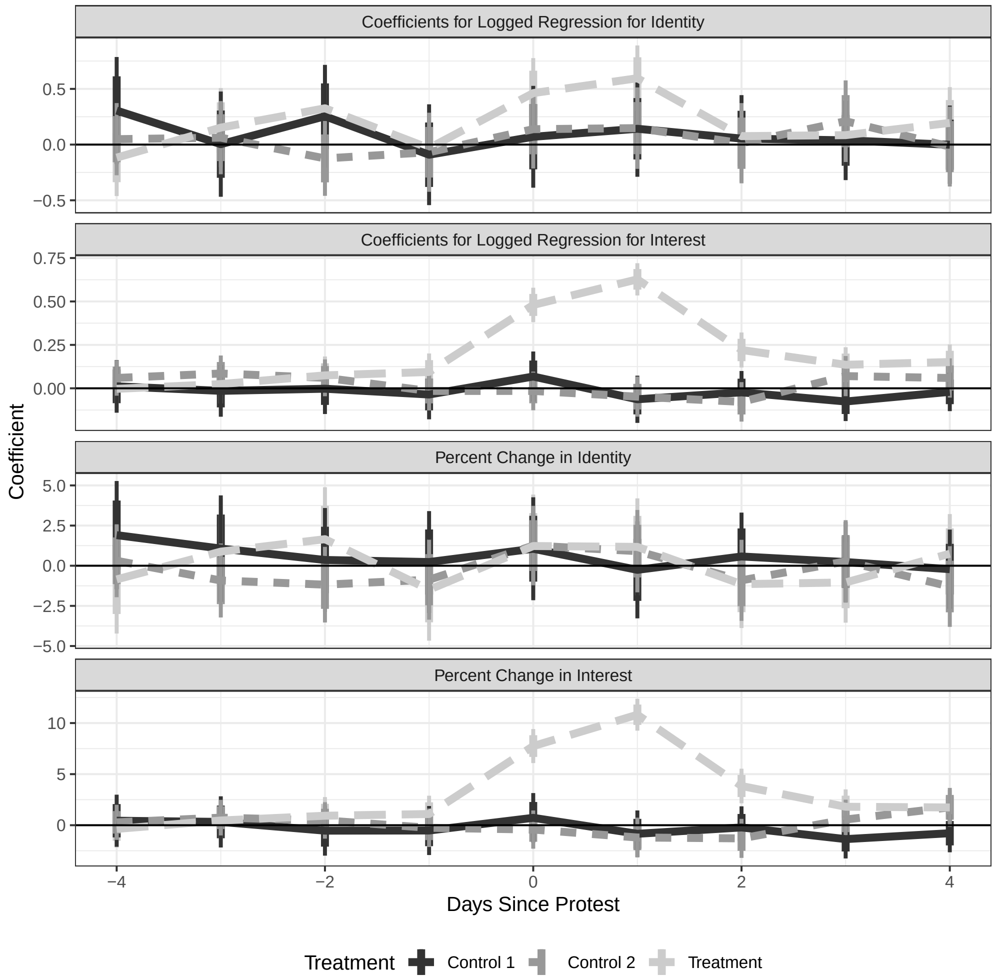


Figure 1.2: Changes in interest and identity when protesting

Note: The thick error bars are the 90% confidence interval while the thinner one is 99%. The scales of the plots are different. The first and second are the coefficients for the log OLS, while less intuitively interpretable, they reflect a similar trend to the third and fourth which reflect a percent change in interest or identity. These results visually represent the regression information found in Table 1.5.

Table 1.5: OLS regression results with day and individual fixed effects.

	Dependent variable:			
	Interest	log(Interest)	Identity	log(Identity)
	(1)	(2)	(3)	(4)
Protest Day - 4	−0.188	0.006	−0.342	−0.023
	(0.622)	(0.037)	(0.826)	(0.118)
Protest Day - 3	0.630	0.034	0.106	0.032
	(0.636)	(0.038)	(0.844)	(0.120)
Protest Day - 2	0.829	0.060	1.254	0.256^{**}
	(0.619)	(0.037)	(0.823)	(0.117)
Protest Day - 1	1.184^{*}	0.099^{***}	-1.328^{*}	−0.045
	(0.607)	(0.036)	(0.806)	(0.115)
Protest Day	6.735^{***}	0.432^{***}	1.599^{**}	0.517^{***}
	(0.569)	(0.034)	(0.755)	(0.108)
Protest Day + 1	10.002^{***}	0.590^{***}	0.838	0.622^{***}
	(0.537)	(0.032)	(0.713)	(0.102)
Protest Day + 2	3.440^{***}	0.194^{***}	−0.797	0.116
	(0.575)	(0.034)	(0.764)	(0.109)
Protest Day + 3	1.562^{***}	0.121^{***}	−0.538	0.127
	(0.578)	(0.034)	(0.767)	(0.109)
Protest Day + 4	1.549^{***}	0.122^{***}	1.060	0.207^{*}
	(0.581)	(0.035)	(0.771)	(0.110)
Observations	165,301	165,301	165,301	165,301
R^2	0.003	0.003	0.0001	0.0004
Adjusted R^2	-0.015	-0.015	-0.019	-0.018
F Statistic df = 9; 162280	57.766^{***}	60.314^{***}	1.626	7.609^{***}

*p<0.1; **p<0.05; ***p<0.01

95% confidence interval. In this presentation, the association between protesting and interest is very clear, while it is less so for identity.

Analysis of both signals support Hypothesis 1, although interest has a larger coefficient. In Table 1.6, the average partial effects are displayed for the logit model using both identity and interest as well as the two independently. City fixed effects are included in the table due to their significance. There were no additional significant terms when interactions were included. The model was also evaluated using a truncated version of the model—only using individuals who tweeted during a significant number of protests—but this truncation did not change the results. In the combined model, it can be seen that changing an individuals interest from 0 to 1 causes a 9% increase in the probability that they protest, while changing the identity score from 0 to 1 has a 1.4% increase in the probability of protesting. The substantive result for political interest is robust controlling for identity. This robustness supports the notion that conflating identity and interest would produce substantively different results. These results support Hypothesis 1, that individuals with higher signal levels are more likely to protest, for both signal types. The result is stronger for interest than identity. This is counter to the expectation given existing literature which in general suggests that collective identity is a major driver for political participation. This evidence together supports the supposition that researchers that conflate identity for interest dramatically overestimate the effects of identity on protest behavior. In assuming that interest was collective identity, other researchers may easily have biased results.

Table 1.6: APEs for logit model with daily fixed effects

	(1)	(2)	(3)
Identity	1.4*	2.1**	
	(0.7)	(0.7)	
Interest	9.1***		9.2***
	(1.4)		(1.4)
Chicago	−5.5***	−5.4***	−5.5***
	(1.3)	(1.3)	(1.3)
Houston	−6.0***	−5.8***	−5.9***
	(1.5)	(1.5)	(1.5)
Observations	17,782	17,782	17,782

*p<0.1; **p<0.05; ***p<0.01

Note: The model was estimated using a probit instead of a logit and with various interaction terms to test for the validity of pooling—the results remained consistent. In addition the data is truncated to only include individuals who tweet during more than 3,5, and 7 protests and the results do not change significantly.

1.5 Discussion

In this chapter, we merge literatures on identity and mobilization with digital studies of protest mobilization. There are two primary contributions. First, we distinguishes between interest and identity as similar but distinct signals of collective action. We then introduce a formal model then generates three hypotheses about how the two signals should correlate with collective action mobilization. Second, we extend previous studies of mobilization using online data by separately operationalizing interest and identity. Previous work assumes that an account using a hashtag or certain images identifies with the movement with which those trackable entities are associated. This assumption is too broad and likely explains why this chapter finds results different from previous studies using digital trace data.

Using this operationalization of both values, we find new results which differ from past research. We find that individuals are more likely to protest given higher levels of interest and slightly more likely to protest given higher levels of collective identity. One of our most interesting results is that interest levels increase the day of a protest and peak the day after, before slowly returning back to base levels. This suggests that the effect of the protest is more transient and less longstanding. These results suggest a minimal impact of collective identity on protesting and protesting on collective identity.

By improving the measurement of identity with online data, we build on previous quantitative, non-social media research into identity and collective action in several ways. Collective identity is salient during the mobilization process in authoritarian settings (Pearlman, 2018; Pfaff, 1996). This contrast with the 2020 BLM protests suggests that identity may be less salient in settings where citizens have other means of of organizing. In settings such as the United States, identity may therefore not be an axis on which to build boundary-spanning movements (Wang et al., 2018). The difficulty of mobilizing around identity is further heightened when the identity is race and there are prevailing biases against the group mobilizing (Manekin & Mitts, 2022)

In previous research, the use of surveys leads to biases in both selection and response (Westwood et al., 2022). The spontaneous nature of protests, in addition, makes pre-measurement difficult and, in most cases, impossible (Chenoweth et al., 2022). In our research, these issues are not as problematic. The frequent use of social media by many members of society provides researchers a window into the minds, and histories, of individuals. In using social media the probability of bias is minimized as individuals cannot retroactively add tweets and are not aware they are being observed for this purpose. We have argued that, previous research using social media data has defined collective identity too broadly. In following the usage of hashtags and other trackable symbols, the quantity measured has instead evolved into interest rather than collective identity.

Moving forward, there are three avenues of future research to pursue. In order to further validate the results found in this chapter, measuring interest and identity for other social movements should be performed. Other movements, such as the Yellow Vests in France, have different contexts and can be used to see if our results are general or specific to the Black Lives Matter movement. The second extension is to include individuals who did not protest as a baseline in order to see if there are clear differences in the interest and identity of those who protest and those who never protest. Third, online identity appears highly salient in motivating changes in online behavior (Munger, 2016; Siegel & Badaan, 2020; Taylor et al., 2022). This chapter's results suggest that identity is less important in changing offline protest behavior, and future work should continue to explore the differential effects of identity.

This chapter provides a framework in which to study protest movements and individual signals of collective action. It enables the contextualization of much of the previous quantitative work on the subject and takes a step towards unifying it into a singular conversation. While there is clear future work to be done, this chapter provides a first step in these efforts.

Chapter 2

PERSUADABLE VOTERS DECIDED THE 2022 MIDTERM: ABORTION RIGHTS AND ISSUES-BASED FRAMEWORKS FOR ELECTIONS

With contributions from Daniel Ebanks, R. Michael Alvarez, and Jacob Morrier.

Leading to the 2022 midterm elections, fundamentals-based forecasts and conventional wisdom among pundits pointed to a strong Republican wave. The incumbent president had a low approval rate, and the national economy was struggling. However, the Republican Party did not perform as well as models and conventional wisdom had suggested. This has led some to suggest the 2022 midterm elections are an "asterisk election," with idiosyncratic, unpredictable results. Still, previous research shows that factors beyond the fundamentals can help predict election results. For instance, unexpected variations in some issues' public salience may lead voters to consider factors they normally disregard. Using a nationally representative sample of registered voters interviewed immediately after the November 2022 midterm elections, we show that abortion was a decisive and highly salient issue in this election. Comparing these results to analogous ones from a November 2020 survey shows this was not a foregone state of affairs. This leads us to believe that abortion's increased salience is attributable to the Supreme Court's decision to overturn *Roe v. Wade* in June 2022. The 2022 midterm elections seem to have been swayed by this exogenous shock to the political system..

2.1 Introduction

One of the regularities in contemporary American politics is that the president's party loses seats in Congressional midterm elections (Jacobson & Carson, 2019). For the past five decades, political scientists have argued that this phenomenon arises as voters treat midterm elections as a referendum on the economic performance of the president and their party (Kramer, 1971; Tufte, 1975). Recent research has shown that even the unusual political situation following Trump's rise to the Presidency in 2016 led to a referendum-style 2018 midterm, with the Republicans losing 40 seats in the House of Representatives while gaining two seats in the Senate (Jacobson, 2019).

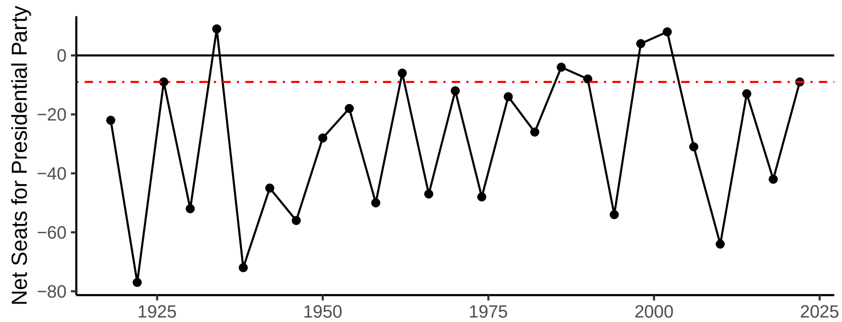


Figure 2.1: House seat change for presidential parties during midterm years. The three times where seats were gained (1934, 1998, and 2002) can be seen. In addition, it is clear the 9 seats lost in 2022, represented by the red dashed line, was a historically mild loss. Source: History, Art & Archives, U.S. House of Representatives, (https://history.house.gov/Institution/Party-Divisions/Party-Divisions/).

Pre-election analyses of the November 2022 midterm elections based on retrospective factors, like the president's approval and the state of the national economy, suggested that the Democratic Party would lose seats in both the House and Senate.[1] The American economy was experiencing widely reported and significant levels of inflation, and the incumbent president, Joe Biden, was relatively unpopular ("Historical Inflation Rates: 1914-2023: US Inflation Calculator", 2023; "Presidential Job Approval Center", 2023). That said, the Democrats only lost nine House seats, while expanding their majority in the Senate to 51 seats. As seen in Figure 2.1, this is a historically mild loss, inconsistent with the models typically used to predict midterm elections' outcomes.

Additional evidence suggests that the 2022 midterm elections were an anomaly. For the first time since 1934, the party in the White House did not lose control of a single state legislature and, in fact, secured full control of the state government in a key swing state, Michigan. The Democrats also expanded their legislative majorities in Nevada and California. Finally, Democrats won key gubernatorial races in multiple swing states, including Arizona, Michigan, Pennsylvania, and

[1] Of course these predictions vary in what they argued could be the magnitude of Democratic seat losses. For example, one analysis (dated August 30, 2022) predicted that the Democrats would lose 30 House and 3 Senate seats (Woolley, 2022). Jacobson noted that the typical predictive model relying on presidential popularity and economic factors would have predicted a Democratic House seat loss of about 45 seats, though he noted that due to partisan loyalties and issues like abortion that the Democratic seat loss could be considerably lower than that (Jacobson, 2022). Larry Sabato's "Sabato's Crystal Ball" predicted in their final pre-election forecast that the Republicans would gain a seat in the Senate and 24 seats in the House https://centerforpolitics.org/crystalball/articles/final-ratings-for-the-2022-election/. The range in the forecasts given by the academic pundits for the 2022 midterm elections was considerable (Edsall, 2022).

Wisconsin. That said, Republicans also experienced electoral successes, including gains in Florida and Texas. New York also swung to the right, giving the Republican Party its entire margin of control in the House of Representatives. Despite these achievements for the Republicans, the Democrats' midterm performance positively defied expectations.

The Republican Party's under-performance in the 2022 midterm elections is the focus of this paper. Given a context in which presidential approval and the economy were linked up in the Republicans' favor, they should have gained a substantial number of seats in the House and the Senate. They did not. We argue that factors outside of President Biden's and Congress' control explain the 2022 midterm elections' apparently anomalous outcome. These factors, especially the U.S. Supreme Court's *Dobbs v. Jackson Women's Health Organization* decision, led segments of the midterm electorate to cast their votes according to an issue-based model, prioritizing opinions on policy issues like abortion rather than approval of President Biden or the state of the national economy.

One of our key contributions is to point out that, in a polarized America, it is necessary to look at what factors motivate the voting decisions of persuadable voters—those who do not identify with either political party. After all, political polarization implies that Democratic identifiers tend to overwhelmingly support Democratic candidates and Republican identifiers to overwhelmingly support Republican candidates. We show below that one of the keys to understanding the 2022 midterm elections is taking this polarization into account. Accordingly, we separately study the factors driving the voting decisions of three groups of partisan identifiers: Democratic, Republican and, most importantly, Independent voters. Using a nationally representative survey of 2,109 registered voters, we provide evidence that Democratic voters overwhelmingly voted along party lines, whereas the issue of abortion persuaded enough Independents and Republicans in key districts to fight the House to a draw.

The rest of the paper proceeds as follows. First, we present the data collection process and survey methodology. Next, we go over some basic descriptive statistics, which we use to justify subsequent modeling decisions, and discuss our regression strategy. Then, we report key results, justify our model specification, and compare these results to analogous results for the 2020 election, in order to confirm the conclusions drawn from the analysis. Finally, we discuss the results and their implications for studying elections.

2.2 Midterm Elections and American Politics

Early research on midterm elections in the United States focused on the so-called "surge-and-decline" theory (Campbell, 1966). Building on the same concept of partisanship that was articulated in the seminal book *The American Voter*, the basic argument was that during presidential election years the winning president's party would gain seats due to the short-term salience of partisanship. But in the midterm elections, the salience of party would recede and thus the out-party would gain seats in midterm, with midterm elections reflecting the partisan equilibrium in the nation. As such, the seat distribution following the midterm elections should be normally distributed around this stable equilibrium.

However, while in general there is a swing in House seats away from the president's party, there is historically a great deal of volatility in the magnitude of this midterm swing (Jacobson, 1987). There is much more volatility than is explained by the return to a normal vote (Bafumi et al., 2010). Clearly, for many midterm elections, there is another component to the seat shift away from the president's party which accounts for the magnitude of loss. Compelled by this volatility, researchers started to focus on the association between other retrospective performance factors and election performance.

The first model of this sort was the midterm-as-a-referendum model, which linked the magnitude of the downward swing with the president's approval and the state of the national economy before the midterm elections (Kramer, 1971; Tufte, 1975). Voters were seen to punish the party of the President for their view of him as well as for the how, what they viewed as his actions, were affecting the country. This model has seen general support from historical data (Jacobson, 2007, 2019; Kramer, 1971; Tufte, 1975) and is often how midterm elections predictions are made.

The midterm-as-a-referendum model generally focuses on two factors: the president's approval rating, and the state of the national economy. In situations where the president's approval rating is low and where economic performance is poor, the president's party should lose a large number of Congressional seats. When these factors are ambiguous, for example when a president's approval is low but the economy is performing well (like in the context of the 2018 midterm election), election losses for the president's party may be more muted.

Building on this retrospective perspective, various scholars have proposed a balancing model of voter behavior to explain the mechanisms underlying the midterm backlash (Alesina & Rosenthal, 1995; Fiorina, 1992; Mebane, 2000). This model synthesizes both retrospective and prospective elements. Voters retrospectively consider the president's legislative agenda and deem it extreme. They then reward the opposition and punish the president's party to constrain her future legislative options. Here, the median voter exploits the checks and balances of a presidential system to forcibly moderate the president by handing control of the legislature to the opposition party. If the president and legislature wish to pass laws, they will need to find consensus.

Evidence used to test the balancing model often relies on aggregated measures of electoral outcomes (Alesina & Rosenthal, 1995; Fiorina, 2003), even in comparative contexts (Kern & Hainmueller, 2006). In aggregate, it is not clear that the 2022 midterm elections can easily be explained by balancing models. The model's overall prediction—that voters would likely want to balance the second two years of President Biden's term by giving Republicans strong majorities in the House and Senate—did not occur. Other researchers have tested the balancing model using different approaches and the empirical evidence does not tend to provide support for the model (Algara et al., 2022; Alvarez & Schousen, 1993; Lacy et al., 2019). Finally, balancing theories also imply that voters engage in a complicated cognitive process—involving both retrospective and prospective elements. These assumption seems at odds with empirical research that shows that voters are generally poorly informed and unsophisticated (Bartels, 1996; Downs, 1957).

Existing models of U.S. midterm elections seem to not explain the outcome of the 2022 midterm, which means we must turn to other models of voter decision making. If the 2022 midterm elections were not decided by retrospective evaluations of President Biden's or the Democratic Party's performance, nor the state of the national economy, nor by sophisticated strategizing about balancing the power of the two parties across the three branches of government, what other theories of voter decision making might help explain this election?

The other factors often used to explain voting behavior in American federal elections are partisanship and issues. Partisanship is a powerful factor in American politics and has long been shown to be a key decision variable for voters (Campbell et al., 1960). However, in recent elections in the United States for many in the electorate their party affiliation has become synonymous with their voting decisions (Mason, 2018). Virtually all Democratic identifiers vote for Democratic candidates, while virtually all Republican identifiers voter for Republican candidates.

Thus, partisanship is a key part of our story for the 2022 midterm elections: since partisan identifiers vote for their party's candidates, we need to study those who do not identify with a party, those who are Independents (Klar & Krupnikov, 2016). In the context of today's highly polarized political environment in the United States, where party identification is synonymous with voting decisions, the political independents are the potentially persuadable voters.

This is where political issues enter the story. Political independents lack the pull of partisanship, and if retrospective factors are not pushing their voting decisions, then perhaps highly salient political issues will dictate how they vote in midterm elections. Research has shown that uncertain voters may cast their ballots based on issue information (Alvarez, 1998) and here we note that in 2022 there were highly social and policy salient issues like gun policy, COVID-19, foreign policy, racial and ethnic inequalities. But the issue that we argue was largely the focal point of the 2022 midterm elections was abortion.

In the United States, the 1973 U.S. Supreme Court decision in *Roe v. Wade* established abortion as a constitutional right for people with uteruses. This right was largely confirmed by subsequent Supreme Court decisions like *Planned Parenthood v. Casey* in 1992. During this period abortion became a divisive issue, part of the partisan landscape of American politics. (DiMaggio et al., 1996; Lewis, 2017) For decades, while partisan voters had distinct positions on abortion, it did not seem that elected officials had much say in the matter as *Roe v. Wade* generally established the constitutional right to abortion. But in the summer of 2022, as congressional campaigns started to take shape, the Supreme Court shocked the political world by handing down the *Dobbs v. Jackson Women's Health Organization* decision, which held that the the Constitution does not provide a right to abortion. This was a true shock to the American political system, and suddenly the issue of abortion again became salient as legislatures at the state and federal levels became the focus of debate about the future of abortion policy in the United States.

The referendum model has generally done well explaining past midterm elections outcomes. However, some initial support for an issue-based model come from observational results of past elections. In all but three midterm elections since 1916, the president's party has suffered a net loss of seats in the House of Representatives, as seen in Figure 2.1. In each of the three anomalous cases where the president's party gained seats during the midterm, there were clear external factors contributing to the White House's party success (e.g., the Great Depression in 1934, the Clinton impeachment in 1998, and the 9/11 terrorist attacks in 2002). Thus, while it seems that the recent performance of the president's party and the overall state of the national economy help determine the makeup of Congress after a midterm election, other issues may arise that can lead to anomalous outcomes.

2.3 Data and Methods

Surveys

The primary source of data we use in this paper is from a November 2022 nationally representative online survey, designed by our research group and fielded by YouGov. The survey was in the field in the days immediately following the 2022 midterm election, November 9–19, 2022. The sample contains responses from 2,109 U.S. registered voters, who were selected by YouGov from their opt-in survey subject panel.

The survey design was reviewed and approved by Caltech's Institutional Review Board (IRB), and was conducted as part of a larger project studying the opinions and political behavior of the American electorate. Informed consent was waived by the IRB. The survey margin of error is approximately 2.3%. In the analyses below we use the sample weights provided by YouGov, which weight the sample using gender, age, race and education from the American Community Survey; the weights also use information on the 2020 Presidential vote. The weights have a mean of 1.0, standard deviation of 0.4, and a range of 0.1 to 4.2. All of the estimates we report in this paper are weighted. YouGov provided a fully anonymized dataset.

We also below take advantage of a survey our group conducted in November 2020, also a nationally-representative sample of American registered voters, implemented online by YouGov. This survey had many similar questions to those in our November 2022 survey, facilitating comparison between datasets and elections. This survey design was also reviewed and approved by Caltech's IRB. Informed consent was waived by the IRB, The 2020 survey was fielded November 4-10, 2020, using subjects recruited by YouGov from their opt-in survey subject panel and those from an external partner. The total sample size was 5,051, with an estimated margin of error of 2.0%. The survey was weighted on gender, age, race, education, U.S. Census region, state of residence, and 2020 Presidential vote; the weights range from 0.1 to 5.973 with a mean of 1 and standard deviation of 1.

We note that all surveys have error: from sampling bias, selection bias, and non-response. Yougov and our team took statistical and methodological steps to reduce the effects of many potential sources of survey error. For both surveys, YouGov provided a fully anonymized dataset the data are weighted to vote choice, to limit the effects of partisan-response bias. We are following AARPOR standards for survey collection to ensure at least a nationally representative sample of engaged voters, and weight by education, age, race and gender. These strategies ensure the responses we receive are high quality and accurately reflect the opinions of the respondents; however, although we believe contain high quality responses and are properly weighted, all surveys are subject to some concerns from selection and nonresponse bias. For example, surveys such as ours are likely to attract high-information voters and voters who are politically engaged, especially given the survey is time-consuming, even if respondents are compensated. Although we cannot control for this directly, we note education is highly correlated with political engagement and information, which helps limits this bias. A secondary concern is that voters might be influenced by media narratives in the immediate aftermath of the election. We note a few characteristics of the data that help alleviate this concern: first, the majority of respondents returned their survey in the days immediately following the election, before media narratives had reached a consensus. Second, the winners were not declared until a approximately a week after the election in both 2020 and 2022, due to the counting of mail-in votes. This should reduce any biases introduced by voters winners declaring they voted for the winner when they had not, as a winner had yet to be declared.

Returning to our November 2022 survey analysis, from that dataset we use a variety of survey responses, with the full question Appendix B.1. Our primary dependent variable of interest is the Congressional midterm vote, which we use to measure voter preferences in the election. For that we used the straightforward generic ballot question,

- **Voter Preferences:** “In the November 2022 election for U.S. Congress in the district where you live, which candidate did you vote for?”

Respondents could indicate whether they voted for the Democratic candidate, for the Republican candidate, for neither, that they were not sure, or that they did not vote.

To measure incumbent performance, our survey also included two questions to measure economic and national economy evaluations:

1. **Financial Situation:** We are interested in how people are getting along financially these days. Would you say that you and your family living here are better off or worse off financially than you were a year ago?

2. **Economic Situation:** Now thinking about the economy. Would you say that over the past year the nation's economy has gotten better, stayed the same, or gotten worse?

These questions are important for testing the midterm-as-a-referendum model and are similar to the questions used in past research (Kiewiet, 1983; Kinder & Kiewiet, 1979).

There are two primary ways that past research has measured or estimated the importance of issues in elections and voter decision making; there is a healthy literature debating the relative merits of using self-reported issue importance or choice-based issue importance. (Alvarez, 1998; Alvarez et al., 2000; Hanretty et al., 2020; Krosnick, 1988; Leeper & Robison, 2018). Our 2020 and 2022 surveys had a wide array of self-reported issue importance measures, covering a broad range of issues of relevance in each election. From our surveys, we use two different approaches to measure the role of issues in Congressional midterm elections voting. The first was a question asking whether the respondent thought that either the Democratic or Republican party would do a better job on a number of issues: preventing terrorism, mitigating climate change, abortion policy, law enforcement and criminal justice reform, preventing further spread of COVID-19, reducing the federal budget deficit, growing the economy, providing affordable health care, American foreign policy, and inflation. We find these measures to be highly collinear with other partisan inflected indicators in the survey. So, for the bulk of the analysis in this paper, we use a second, more typical "most important issue" question, asking respondents to indicate what issues from a long list were the most important problems that influenced their vote in the midterm elections. From this question, we create a variable used to test various theories of midterm elections. In particular, we use the following question:

- **Issue Importance:** How important, if at all, were each of the following issues for you as you thought about whom you would vote for in the congressional election in your area in November 2022:[immigration, abortion, foreign policy, economic inequality, the COVID-19 outbreak, violent crime, health care, the economy, racial and ethnic inequality, climate change, inflation, gun policy, Supreme Court appointments].

  - *Potential Choices:* Very important, Somewhat important, Not too important, Not important at all, skipped.
  - *Implementation:* To implement this question as a variable of issue importance, we denote somewhat or very important as 1 and not too important and not important at all as 0, for each issue. We treat NAs as missing data and drop them.

The list of issues includes immigration, abortion, foreign policy, economic inequality, the COVID-19 outbreak, violent crime, health care, the economy, racial and ethnic inequality, climate change, inflation, gun policy, and Supreme Court appointments. This traditional measure of issue importance has a much weaker correlation with party identification, but still correlates with voter preferences. These correlative relationships make this measure a useful variable for testing the various theories explaining midterm election outcomes for two reasons. First, party identification will not immediately swamp any potential effects of issue importance by using a question that mentions specific parties, because for such issues, Democrats and Republicans are highly likely to view their own party's abilities favorably. Second, as we show in this paper, voter preferences vary with respect to this importance measure, across a range of issues. This allows us to make inferences about the role the importance of specific issues played in how voters decided to vote in the 2022 midterm elections.

Additionally, from our survey we use responses to a three-point partisanship question, whether the respondent identified as a Republican, Democrat, or Independent. Importantly, this is not a question of registration, but self-identification. This helps us avoid any confusion that could be created from registration due to primary rules and instead gets a measure of party loyalty. Finally, we have a number of other demographic factors; gender, educational attainment, region, race/ethnicity, religious affiliation, and age. These factors allow us to account for other correlates of voting decisions.

Observational Results

As a first step in the analysis, we look at the results from the survey to see if particular trends are apparent. The initial object of interest is a measure of polarization—

did people vote for the party they identified with? These results can be seen in Figure 2.2. Individuals who self-identified as Democrats were extremely likely to vote for the Democratic candidate. The relationship is less strong for Republicans, and Independents voted both ways. This encourages us to center our study on the Independent and Republican voters as they seem to have driven the results away from the Republican party.

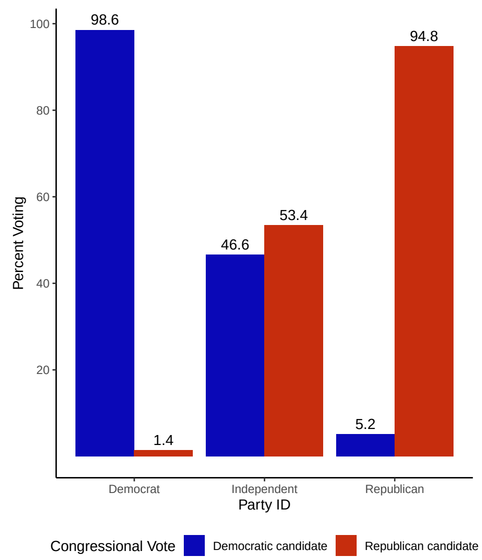


Figure 2.2: This figure shows the weighted percentage of members of each party who voted for congressional candidates of each party. From this figure it is clear that party cohesion was strong—most people voted with their party. This supports the idea that most of the interesting variation will come from those who identify as Independents. Independents as a group make up about half as many individuals as Democrats or Republicans, indicating that they could be the swing in the election. Democrats were significantly more loyal to their party than Republicans.

Since the midterm-as-a-referendum model suggests vote choice is influenced by characterizations of the economy and personal finances, we next look at the distribution of vote choice contingent on those views as well as party identification. These results can be seen in Figure 2.3. Those who identity as Democratic were likely to vote for the Democratic candidate, regardless of their views of the financial and economic situation. Additionally, they had the most positive views of both situations, with less than half of respondents thinking either had gotten worse.

We see that Independents have a higher probability for voting for the Republican candidate when they view either situation as having gotten worse. Over half of Independents thought the economy had gotten worse and almost half thought their personal financial situation had gotten worse. Republicans are slightly more likely

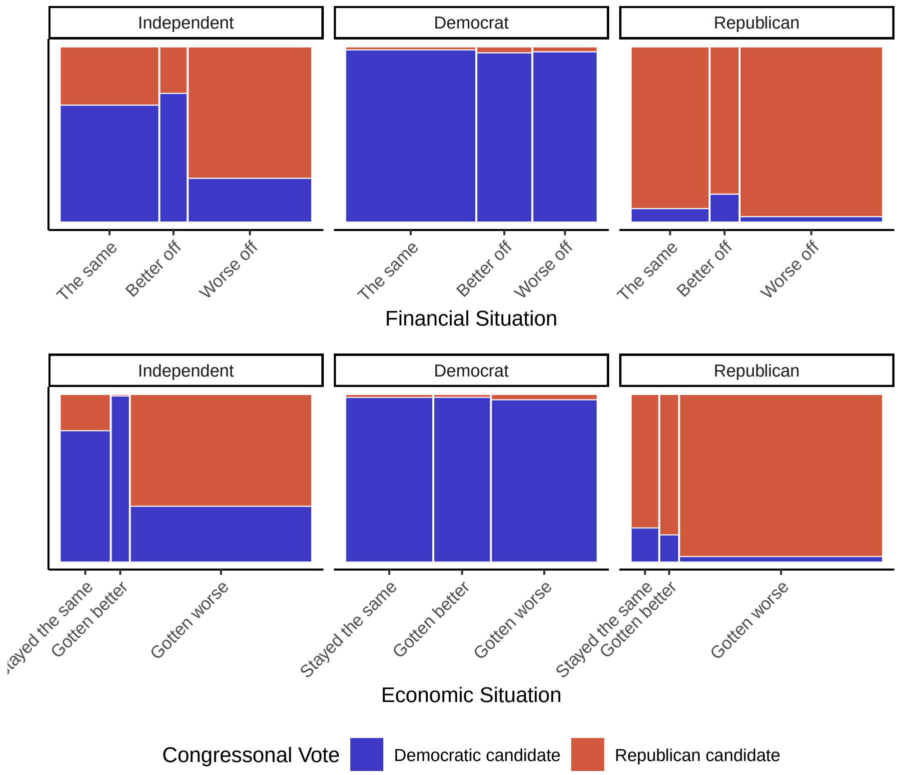


Figure 2.3: Vote choice by party identification and views of the national economy and personal finances as compared to the previous year. The width of the bars represent what proportion of individuals (weighted) fell into each grouping. This helps to see how the answers are related to vote choice but also how party identification relates to the response.

to vote for the Democratic candidate if they think the economic/financial situation has improved or remained the same, but the number of individuals who believe that is the case is relatively small, specifically for the economy. While these observations hint at a midterm-as-a-referendum style model, there seems to be a leaning towards the Democratic party that is not accounted for.

The divide on issues complicates matters slightly more. As seen in Figure 2.4, those who affiliated with the Democratic or Republican party were not particularly swayed by issues. There is a slight exception for the economy, inflation, and violent crime, where viewing the issue as not important as a Republican meant the individual was more likely to vote for the Democratic candidate. These outliers, however, make up a small portion of the population, as evidenced by the size of the "not important" Republican point for the issues. There is a much clearer differentiation among Independents by topic. Finding abortion, climate change, COVID-19, economic inequal-

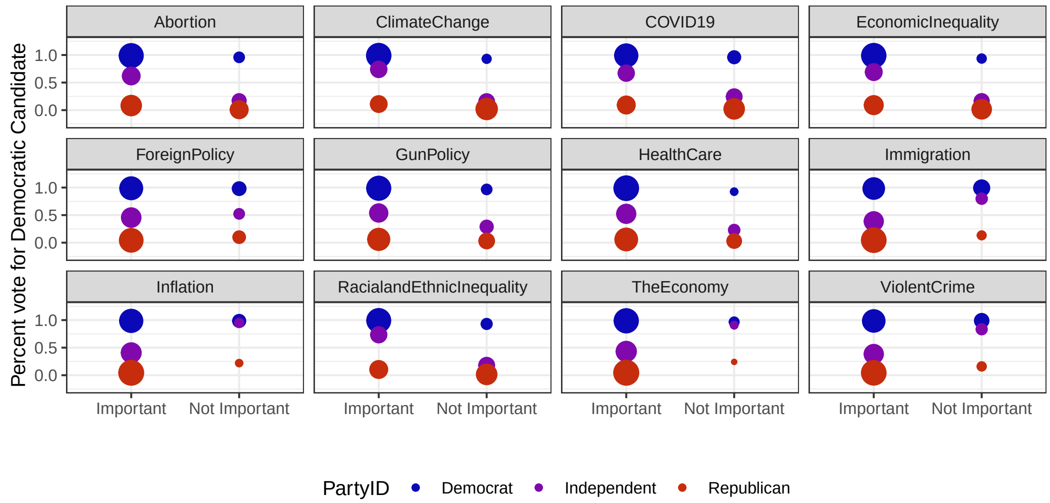


Figure 2.4: For issues, how different party identifiers voted based on whether they thought it was important or not. The color represents partisan identification while the size is the weighted number of individuals who fit the category. It is clear that most partisans stuck to their party, regardless of their views on issues. Independents were swayed by the issues they viewed as important.

ity, gun policy, health care, or racial and ethnic inequality as important increased the probability they voted for the Democratic candidate significantly. Alternatively, finding immigration, inflation, the economy or violent crime important increased the probability of voting for the Republican candidate. These findings further convince us that, for Independents, issues did matter in vote choice and this could be a viable path to follow given the evidence against a midterm-as-a-referendum model fitting these elections' final results.

Multivariate Modeling Strategy

The ideal model to measure the predictive relationship between demographics and opinions with vote choice is a fully specified logit model. First, we estimate a model including a set of demographic controls (race, gender, age, region, educational attainment), a set of variables identifying which party individuals think are better at handling issues, and a set of variables which state whether individuals find specific issues important. The questions regarding which party is better at issues end up being highly co-linear with party identification, so these questions are removed in order to ensure tractability of the results. In addition, this model is reduced through the elimination of questions that are not statistically significant. Details on this process can be seen in Appendices B.4–B.5. This model is referred to as the *pooled* model.

A secondary model—the *party-based* model—is also used. In this model individuals of each party identification are viewed separately. This decision is supported by a series of Wald tests that can be seen in Appendix B.5. This allows us to focus on Independent voters, who seem to have made the difference in the election as well as Republican voters who seem to deviate from the party line more than Democratic voters. In addition, it enables each covariate to effect individuals who identify as party of each group in completely different ways. This decision is supported by the evidence we see in Figure 2.2 of the highly-polarized nature of the contemporary American electorate: nearly all of the Democratic partisan identifiers in our survey reported voting for Democratic House candidates, while nearly all of the Republican partisan identifiers in our sample reported voting for Republican House candidates. This is strong evidence that very few partisan identifiers were likely to be persuadable in the 2022 midterm election. Instead, notice that the partisan Independents show in the middle columns of Figure 2.2. Their support for House candidates was nearly evenly split between Democratic and Republican candidates, which shows the electoral importance of partisan independents in the 2022 midterm elections.

With respect to our modeling strategy, we note that the study is strictly observational and based on predictive evidence. We limit ourselves to claims related to the predictive power of the covariates we study. While we cannot make causal claims in this study, we are able to conduct statistical tests to gather evidence consistent or inconsistent with theories of midterm electoral behavior. This statistical evidence forms the backbone of the inference in our study. To further judge the midterm-as-a-referendum and issue-based theories of midterm elections at the individual level, we look at the average marginal effect of answers to the various questions on vote choice given each logit model. We are unable to quantitatively judge the balancing model; as mentioned earlier this is one of the major issues with the model. However, given the results of the election and that the Democrats controlled the House, Senate, and Presidency, it is clear that in this case the balancing model did not drive the election. If the November 2022 elections subscribed to the midterm-as-a-referendum model we would expect there to be a large average marginal effect of voting Republican when viewing the economy and/or individual financial situation as having gotten worse and a negative marginal effect give views of improvement. Additionally, in this case, when accounting for demographics and party-preference, we would assume there would be no significant effect of importance of issues on vote choice. If the election was issue-driven we should expect there to be certain issues which have statistically significant negative average marginal effects—those which induce voting for the Democratic candidate. In the next section this strategy is applied to the survey data.

2.4 Results

Given the prevailing political and economic environment, under popular theoretical models of the mechanics of midterm elections, we argue that the Republicans under-performed in the November 2022 midterm election. They won 221 seats in the House (an increase by 9 seats from the previous election), 49 seats in the Senate (a decrease of 1 seat from the previous election), and their performance in state level offices was uneven. Under both a midterm-as-a-referendum model and a balancing theory model, the Republicans should have made sizeable gains in this election. Relevant conditions for the midterm-as-a-referendum model were certainly satisfied: the sitting president was unpopular, perceptions of the economy were middling to poor, and the opposition party ran aggressively on presidential performance. The conditions of the balancing model were equally satisfied: Democrats controlled both the House and Senate, as well as the presidency. In the previous session, the Democrats passed trillions of dollars in landmark spending legislation, including investment in many progressive priorities. Yet, the election results confound the predictions of both of these theories—that Republicans should have won large majorities in the House and Senate, and swept state-level offices, especially in swing states.

In order to understand why these theories failed to predict the electoral outcomes of the 2022 election, we subject them to further testing. We focus our study on the U.S House for two reasons. First, all 435 U.S. House districts have U.S. House elections, so we can test the theories on our full sample of U.S. registered voters. Second, candidate effects tend to be weaker relative to the Senate or Governorship, since name recognition is lower. This leads us to believe that prevailing theories of midterm elections are most likely to be apparent in U.S. House elections.

In order to better understand the mechanics of the 2022 midterm elections, we explicitly test the midterm-as-a-referendum model in the U.S. House and show that there is weak support for it at the individual-level of self-reported assessments of the in-party's performance on the economy. The balancing theory is generally more difficult to test, even with our individual-level, high-quality data. An explicit test of this theory would solicit voters' strategic considerations when making a voting decision related to party. In practice, any such solicitation would be highly correlated with party identification, voter ideology, or even covariates that indirectly correlate with party. Given this theory is generally hard to falsify, we argue it is not a strong candidate for explaining the midterm elections in 2022, nor midterm elections in general.

Given the weak empirical support for the midterm-as-a-referendum model and given the conceptual problems with testing the balancing model, we propose an issues-based model for explaining the 2022 midterm election. We show persuadable voters supported Democrats at an unexpected rate in ways that are consistent with their being convinced by issues they held as important. Persuadable voters include those who identify as political and weak partisans—Republicans and Democrats. We find abortion, violent crime, and inflation are issues which persuadable voters found important and were predictive of the party they supported in the election. In particular, we find abortion was a major indicator of the Democratic support among persuadable voters, while crime and inflation were indicators of support for Republicans. Finally, we test to see if support for abortion, violent crime, and immigration where novel in the 2022 midterm elections in their ability to persuade voters. We do this by comparing the results for these three issues against a 2020 benchmark.

Testing the Referendum Model

A popular model for midterm performance, we first explicitly test the midterm-as-a-referendum model in the case of the 2022 midterm election. While the aggregate results clearly do not hold—this is evidenced by the relatively poor performance of the Republican party despite high levels of inflation and President Joe Biden's low approval rating—it is worthwhile to confirm this result at the individual level.

If the midterm-as-a-referendum model held true, we would expect voters' assessment of the President's party to be strongly predictive of a vote against them, especially among persuadable voters. We test this prediction by exploiting two economics-based questions fielded in the survey and the pooled, as well as party-based, logit model (details in Appendix B.1). The first question asks voters the assessment of their personal financial situation, and the second about their assessment of the national economic situation. We note this is an indirect approach of measuring voter's assessments of the White House party's performance. We prefer such measures rather than direct ones because we believe they are less likely to be as strongly correlated with party identification or ideology as measures that directly name a party.

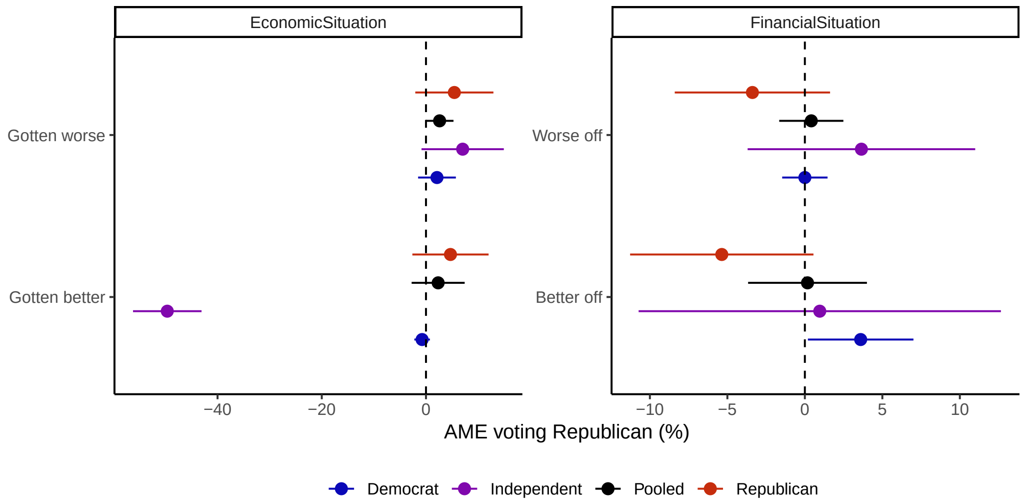


Figure 2.5: Economic evaluations' effect on voting Republican for pooled and party-based logit models

In Figure 2.5, we show weak support for a midterm-as-a-referendum model. It shows that the average marginal effects on voting Republican as a function of evaluations of the national economic situation and assessments of voters' personal financial situations are clustered around 0 percent both in the pooled model and when disaggregated by party identification. There is one notable exception; Independents are more likely to vote for Democrats if they think the economy has improved. However, Figure 2.2, illustrates very few Independents held this view, suggesting this population had a trivial impact on the election.

Taken together, the balance of evidence presented in Figure 2.5 and Figure 2.3 show that assessments of the national economic situation, and personal financial situation were generally not predictive of the vast majority of persuadable voters' party preferences in the House. This finding is even true for partisans who identify with a party. This evidence is not consistent with the predictions of a midterm-as-a-referendum model, where such considerations should be strongly predictive.

Testing the Issue-based Model

Given the weak empirical support for the midterm-as-a-referendum model and the conceptual difficulties in empirically testing a balancing model, we next turn to an issues-based model and the predictions it would generate. If the issues-based model is true, then for at least some issues, persuadable voter's party preferences in the U.S.

House should be correlated with their stated belief that those issues are somewhat or very important. If voters beliefs in issue importance are strongly predictive of how they vote, that provides direct evidence in support of an issues-based model to explain the 2022 midterm elections in the U.S. House. Additionally, if this model holds, we will be able to draw conclusions about and identify the key issues which explain the preferences of voters in these elections.

In Figure 2.6, we present aggregate evidence that issue importance was predictive in this election. In this figure, we show the average marginal effect of stating each issue is somewhat or very important, pooling all voters. While most of the issues are statistically insignificant, violent crime and foreign policy are correlated with individuals being more likely to vote for the Republican candidates while economic inequality and abortion had the opposing relationship—people were more likely to vote for the Democratic candidate if they viewed these issues as somewhat or very important. This is suggestive that in the aggregate, the belief that issues were important was a strong predictor of voters' party preferences.

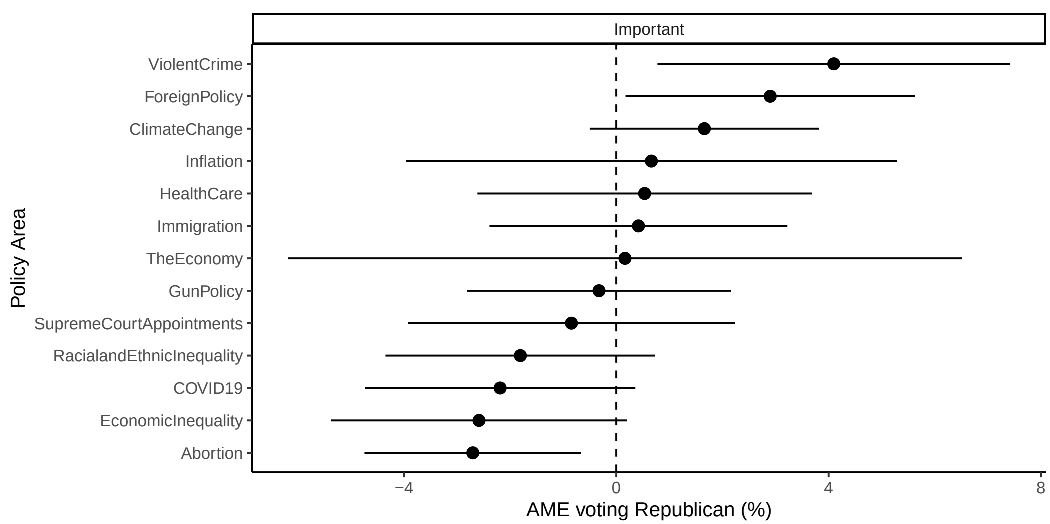


Figure 2.6: Average marginal effect (with 95% confidence intervals) of viewing policy areas as important on the probability of voting for the Republican congressional candidate. The plots show the results for the pooled model.

Returning to Figure 2.4, the majority of individuals in all parties viewed violent crime, abortion, and foreign policy as important while economic inequality was important to Democrats and Independents at a higher rate than for Republicans. These proportions imply that the significant shifts from these issues could have an overall effect on the electorate as a whole. This contrasts with the test of the midterm-as-a-referendum model, where those who were represented by the significant results were a small subsection of the population.

As discussed in Section 2.3, we have strong reasons to suspect voters who identify with different parties will perceive different issues as important. Further, those who identify differently have the potential to respond to these beliefs differently. This can be a result of salient issues, heterogeneity within the party, and differing opinions on the solution to concerns. What is more, self-identified Democrats and Republicans may view issues as important or not important in such a way that the net effect is zero, even if such issues were decisive in determining their vote. This will be a byproduct of their preferred media sources, news consumption, social networks, and their political geography. In order to overcome this concern, we apply the party-based model to analyze the average marginal effects by party identification to see how each group responded to different issues.

The average marginal effects for issues for Independents can be seen in Figure 2.7. Given that partisanship is such a strong predictor of vote choice, we focus the analysis on Independent voters as they are the most persuadable segment of the electorate. This will enable us to see if issues were predictive of their party preferences in the election. The full results for all parties as well as the pooled model can be seen in Figure B.1. The overall effects for partisans are significantly smaller than for Independents. There are no significant issue effects for Democrats while for Republicans there are small Republican biased effects for foreign policy and small Democrat biased effects for abortion and economic inequality.

Among Independent voters, we find affirmative evidence consistent with an issues-based model of the election. In Figure 2.7, we highlight all the issues we tested in our survey. While violent crime and abortion remain issues that are predictive of how Independents voted in the election, the magnitude of the effect is notably larger than in the pooled model. The estimated average marginal effect for violent crime in the pooled model is 4.1% while in the Independent party based model it is 23.9%. Similarly, in the pooled model, abortion has an average marginal effect of -2.7%. Among Independents, the average marginal effect is -12.3%. This ten percentage point decrease in the marginal probability of voting for Republicans in meaningful evidence consistent with the predictions of an issues-based model where abortion is a significant issue. Interestingly, results do not depend on stating directly which party is preferable on the issue. The mere fact Independents found the issue important was strongly correlated with voting for Democrats.

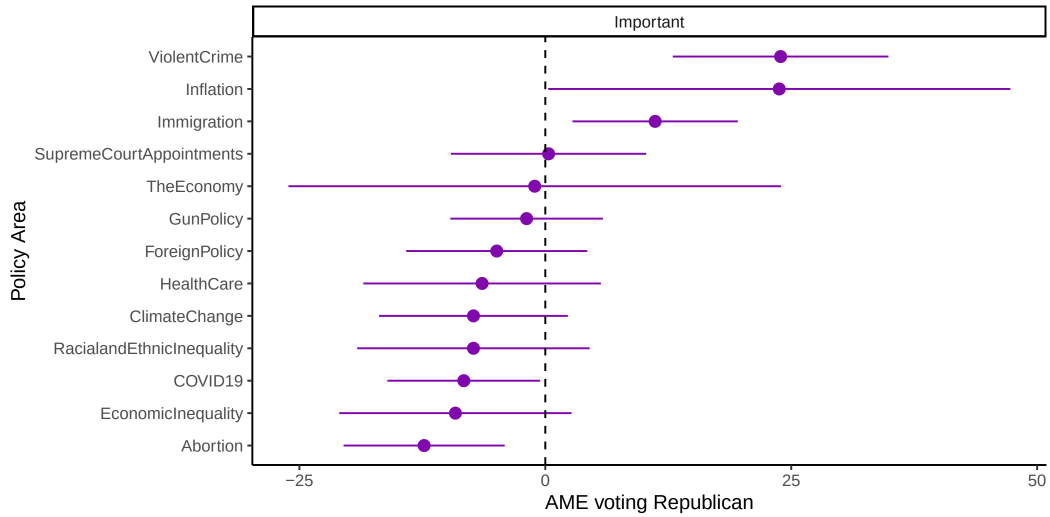


Figure 2.7: Average marginal effect (with 95% confidence intervals) of viewing policy areas as important on the probability of voting for the Republican congressional candidate. The plots show the results for the party-based model for Independents.

Conversely, if Independents stated violent crime and immigration were important issues, they were far more likely to vote Republican, all else equal. The average marginal effects of the three economic issues—inflation, the economy, and economic inequality—all show either high variance (inflation), null effects (economic inequality), or both (the economy in general). These null and high variance findings are consistent with our findings in Section 2.4, which showed the economic evaluations played little explanatory role in this election.

Given that abortion, violent crime, and immigration appear to be major policy issues that affected the election results in the November 2022 midterm, we would like to test to see if these results hold for previous elections. Since the election swung further towards Democrats than expected, we are most interested in the change in how electorates response to abortion changed, as this seems to be the major factor predicting unexpected Democratic House vote choice.

While there is a possibility that this level of effect was normal, and thus the relative importance of these issues is consistent among elections, the result of the Supreme Court ruling in *Dobbs* is an obvious exogenous shock unique to these election. If this level of effect is abnormal, and the ruling raised the importance of the issue for Independent voters, this has important implications for issue-based models. It can begin to explain how issues become important and how persuadable voters form party preferences in the face of the issues. We exploit the exogenous variation offered by the end of *Roe v. Wade* to test how issue importance changes in response to a major policy change outside the direct control of Congress. Additionally, we look to see if the response to violent crime and immigration remained constant in the same period.

2020 Presidential Benchmark

The 2022 midterm elections offer a unique opportunity to explore how exogenous shocks affect voter behavior and disrupt historically observed patterns in midterm elections. In June, before the November 2022 election, the Supreme Court overturned 50 years of Federal legal protection for abortion—a decision over which neither the Democratic congress nor Democratic president had any direct control. Outside of the Senate's role in selecting Supreme Court judges over the previous 50 years, the legislature had little say in the specifics of the decision. In fact, prior to 2022, many Democrats and Republicans assumed that *Roe* was settled law. Because the *Dobbs* decision was a rare surprise policy change outside the control of both the president and congress, we can see how this change affected voter behavior through the channel of increased issue importance. In order to measure the change in the predictive effect of abortion as an issue in the aftermath of *Dobbs*, we calculate a historical benchmark contrasting the 2022 electorate with that of 2020 using the same model as in Section 2.4. In this section, we test the plausibility of the claim that the *Dobbs* decision raised the importance abortion in the minds of voters, and we measured a reduced-form estimate of the magnitude of the change.

For this comparison we make use of our 2020 survey. This allows us to compare the 2022 electorate's opinions on issues and the 2020 electorates opinion on issues to isolate how shifts in opinion on abortion, violent crime, and immigration changed in the relative correlation with vote choice between the two elections. If the magnitudes are the same for persuadable voters in the two elections, than we know that abortion, violent crime, and immigration are consistently important issues. In such a case, while we would not be able to exclude the possibility *Dobbs* had an effect on how voters perceive the issue of abortion, at the same time, we would not have affirmative support that it change how voters perceive the issue. On the other hand, if the magnitudes of the average marginal effects for immigration and violent crime are consistent across the two elections, but abortion goes from null to predictive, than that would be evidence consistent with the Supreme Court's actions in *Dobbs* changing the perceived importance of abortion.

Evidence from this benchmark is consistent with the hypothesis that the Supreme Court's ruling increased the importance of the issue of abortion for persuadable voters in 2022 relative to 2020. To establish the benchmark, we fit the same regression as Section 2.4, including all common variables between the two surveys.[2] If the coefficients in the models are different, the estimated importance of the issue in voting decisions has changed. The average marginal effect on voting for the Republican candidate rather than the Democratic candidate when changing your view from not important (either not too important or not important at all) to important (somewhat important or very important) for each of the three main issues we have isolated as potentially relevant can be seen in Figure 2.8.

When looking at the average marginal effects on voter preferences for stating abortion is a somewhat or very important issue, the role of *Dobbs* is more evident. We report these marginal effects disaggregated by party identification in Figure 2.8. There is a significant change in the average marginal effect of abortion for voting Republican for both Independents and Republicans when comparing 2020 and 2022. In 2020, abortion as an issue seemed to have no effect on vote choice. For Independents, the marginal effect is 4.5%. For Republicans, the marginal effect is 2.5 %. For both, the effect, militates in favor of the Republicans. Moreover, the magnitudes of these marginal effects are small, and neither marginal effect is statistically significant. In contrast to this benchmark, the marginal effects in 2022 for this issue are large in magnitude and notable for their change in direction from 2020. For these two groups, in 2022, the magnitude of the average marginal effect grew to -13% and -6%, respectively. The issue now favored the Democrats for both groups and results are statistically significant. In the world of politics, these are large values, implying the issue played a large role in the election. In contrast, for violent crime and immigration, the average marginal effects between years all overlap. In 2020, Independents were slightly more likely to vote for Republican candidates if they viewed either violent crime or immigration as somewhat or very important issues, and the magnitude of these marginal effects is consistent with the ones calculated for 2022. In both election years, Independents and Republicans were more likely to vote for the Republican candidate if they viewed violent crime as important, however this effect is consistent. The consistency of the effect implies these issues were not responsible for the unexpected nature of the 2022 midterm election.

[2] The surveys have the same question wording on issue importance, with the exception of inflation which was not included in 2020.

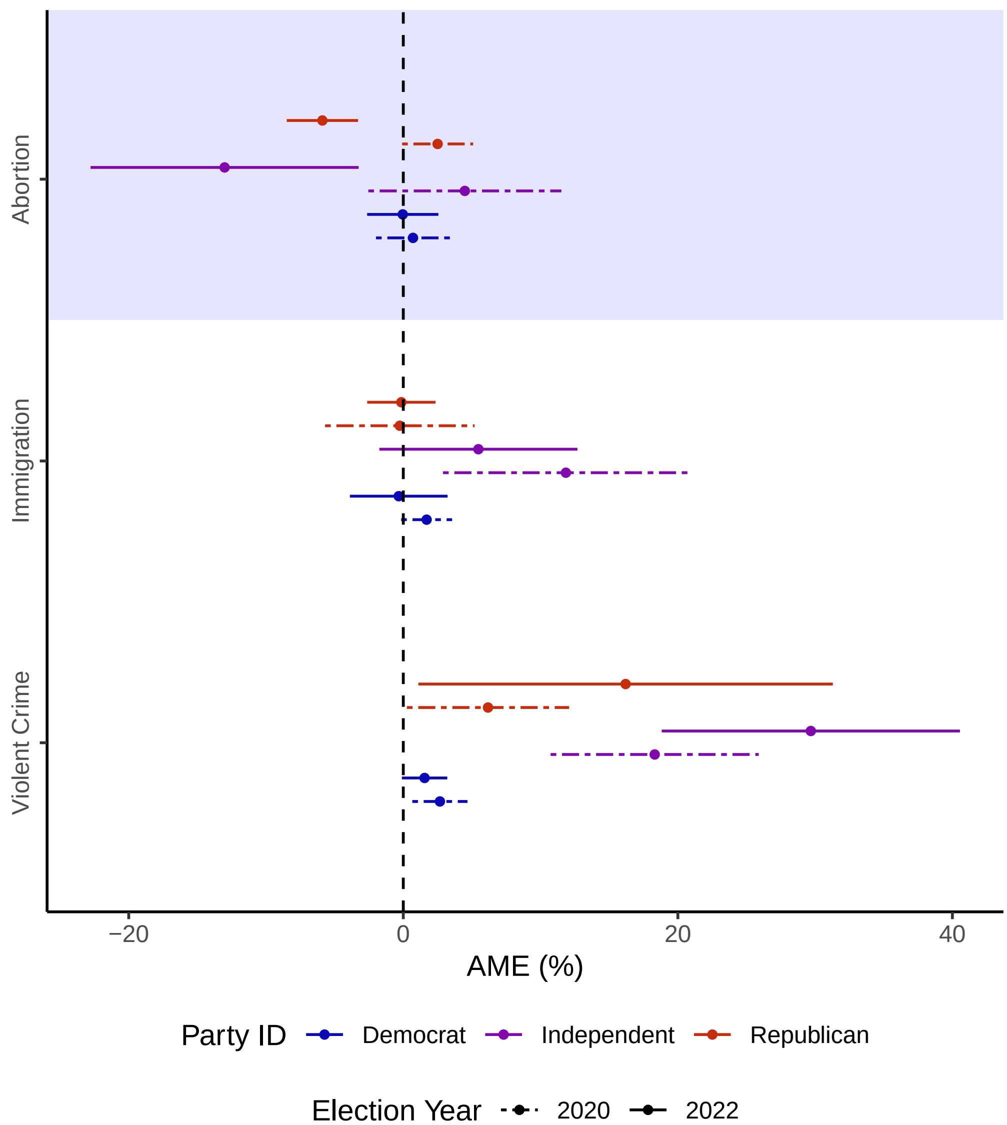


Figure 2.8: Average marginal effect from thinking abortion, immigration, and violent crime were important on the probability of voting Republican for the 2020 and 2022 survey. There is a clear break in the relationship between importance attributed to abortion and voting decisions for Independents and Republicans. There is no such break for any partisans for immigration or violent crime.

It is clear from the results that the abortion issue increased in importance for both Republicans and Independents between 2020 and 2022, and it was one of the only changes between the elections that could have shifted the 2022 midterm results so drastically in favor of the Democratic party. Comparing the results to a similar model for 2020 helps us to validate the results as well as rule out the importance of other issues that could have convinced us otherwise such as immigration and violent crime.

Discussion and Conclusion

Leading in to the 2022 midterm elections, traditional models of midterm elections suggested the Democratic Party would endure significant losses: economic growth was slowing down, inflation rising, and President Joe Biden's approval plummeting. Academic frameworks for understanding midterm elections, such as the midterm-as-a-referendum and midterm balancing models, suggested the President's party risked losing its thin majorities in both houses of Congress. Accordingly, pundits and academics were near-certain Democrats would be defeated, as happened to the President's party in the 2018, 2014, 2010, and 2006 midterm elections. Election results left pundits and scholars startled. The Democratic Party lost a mere nine seats in the House of Representatives and gained one seat in the Senate, numbers comparable to the incumbent party's performance in the 2002 elections following the 9/11 terrorist attacks and the 1998 elections following Bill Clinton's impeachment. Democrats also gained control of Michigan's legislature, the first time the President's party increased the number of state legislatures under its control since 1934. In this paper, we investigate and test plausible theories of behavior in midterm elections. We show that an issue-based model is most consistent with data from a nationally representative survey of U.S. registered voters. Further, we show evidence suggesting that the Supreme Court's decision the *Dobbs v. Jackson Women's Health Organization* case played a key role in raising the salience of abortion.

The midterm-as-a-referendum model is inadequate to explain the 2022 data. Indeed, voters' electoral choices were essentially uncorrelated with their perceptions of the national economic situation and their personal financial situation. A balancing model is difficult to disprove given the complex strategic behavior it assumes on the part of voters. Given these difficulties, we tested the predictions of an issue-based model, in which voters base their voting decisions on the issues they consider salient. Using survey data gathered immediately after the 2022 midterm elections, we found that voters' beliefs about specific issues' importance are strongly correlated with their voting behavior. This is consistent with an issue-based model. Abortion, crime, and immigration are all issues which, if Independent voters believed they were important, influenced their vote in the midterm elections.

On May 2nd, 2022, six months before the midterm elections, the political news website *Politico* published a leaked draft of the Supreme Court's decision in the *Dobbs v. Jackson Women's Health Organization* case. This decision, which overturned the landmark *Roe v. Wade* ruling, sparked massive public interest. Almost two months later, on June 24th, when the decision was officially released, the public took notice. Politicians reacted immediately: California, Colorado, and Vermont introduced state-level amendments that would enshrine the right to an abortion in their state constitutions. Thirteen states had trigger laws in place that automatically banned abortions the moment *Roe v. Wade* was overturned. Rallies, state-level referenda, and countless news segments were devoted to debating the issue of abortion. Republicans argued that abortion would not be a decisive issue in the midterm elections, and that the midterm would instead hinge on crime and inflation. We found that abortion, crime, and immigration were all decisive issues for Independent voters but abortion, unlike the two other issues, has known a significant rise in decisiveness relative to previous elections. In contrast, inflation was only a noisy predictor of Independent voters' choices in the election.

Given *Roe v. Wade*'s rescission, it is not surprising that voters perceived abortion as an important issue. However, it is not evident from cross-sectional data alone that abortion was more decisive in the 2022 midterm elections relative to previous elections. If abortion had always been a decisive issue, then we could not conclude that *Dobbs* had influenced the 2022 midterm elections' outcome. Perhaps, abortion, crime, and immigration have persistently been decisive issues. We show that this is inconsistent with our surveys' results. Indeed, we found evidence that the *Dobbs* decision's shock preceded a substantial increase in the correlation between abortion's perceived importance and voting behavior. The argument that the change in the salience of abortion was due to the appointments of conservative judges such as Neil Gorsuch (2017), Brett Karanaugh (2018), and Amy Coney Barrett (2020) altering the composition of the Supreme Court, is addressed by using the 2020 election as a benchmark. In comparing the 2022 results to the 2020 results we preclude the assertion that these changes were due to changes made on the Supreme Court as the 2020 election was after the major changes in Supreme Court appointments but prior to the Dobbs decision. Although other policy issues, such as crime and immigration, held purchase in the minds of Independent voters, the marginal effect of their stating those issues were somewhat or very important was relatively unchanged relative to the 2020 elections. In contrast, the marginal effect of stating abortion is a somewhat or very important issue increased from 4.5% in favor of Republicans in 2020 percent to 13% in favor of Democrats in 2022, suggesting a substantial structural break.

This change suggests issue importance is related to dramatic, sudden policy changes with far-reaching consequences. Such changes can replace voters' other concerns and lead a new issue to become focal, resulting in a structural break in the political equilibrium. While the Republican Party's base is vocally opposed to abortion, the broader electorate has more moderate views on abortion. When Republicans were campaigning on taking away people with uterus' rights to abortions, many persuadable and cross-pressured individuals were unconvinced. With *Roe* firmly in place, these rights seemed secure, so the issue was tangential to Independent voters' electoral behavior. As the policy environment shifted due to the Supreme Court's *Dobbs* ruling, voters re-calibrated their preferences. It seems that pivotal voters cast their ballots with abortion in mind.

In this paper, we show that an issue-based model offers a general framework for understanding midterm elections. Unlike alternative theories, the issue-based model offers testable predictions in the face of major policy shocks. This framework provides a useful roadmap for future research on electoral behavior. Just as in the 2022 midterm elections, this framework can be used to generate testable hypothesize about key issues in the election, structural breaks in political dynamics, and offers falsifiable explanations consistent with surprises and electoral context. In this sense, through the issue-base lens, the outcomes of the midterm of 2022 were neither surprising nor inexplicable.

Chapter 3

A REPULSIVE BOUNDED-CONFIDENCE MODEL OF OPINION DYNAMICS IN POLARIZED COMMUNITIES

With contributions from Michelle Feng.

Collective opinions affect civic participation, governance, and societal norms. Due to the influence of opinion dynamics, many models of their formation and evolution have been developed. A commonly used approach for the study of opinion dynamics is bounded-confidence models. In these models, individuals are influenced by the opinions of others in their network. They generally assume that individuals will formulate their opinions to resemble those of their peers. In this paper, inspired by the dynamics of partisan politics, we introduce a bounded-confidence model in which individuals may be repelled by the opinions of their peers rather than only attracted to them. We prove convergence properties of our model and perform simulations to study the behavior of our model on various types of random networks. In particular, we observe that including opinion repulsion leads to a higher degree of opinion fragmentation than in standard bounded-confidence models.

3.1 Introduction

Opinions dictate how individuals interact with society. They influence who we are friends with, how we vote, and what we consume. At the individual and collective level, opinions shape our lives and our social interactions. Understanding how opinions are formed and their dynamics provides a framework for studying changes in our society. The role of opinions in politics and governance is a prominent part of public discourse in the U.S. Inspired by discussions of political polarization and partisan politics, this paper presents a mathematical approach to modelling polarized opinion dynamics where individuals feel both a compulsion to become more similar and differentiate themselves from others—these are modeled as attractive and repulsive forces.

The influence of public opinion on politics have been studied by philosophers, sociologists, and social theorists (Blumer, 1948; Habermas, 1991; Speier, 1950). Contemporary approaches to studying opinions frequently seek to quantify them. In this paper, we focus on the dynamics of opinions. We are interested in studying how opinions in a society shift as a result of relationships between individuals. Various models for studying individual opinions exist (Clifford & Sudbury, 1973; Grabowski & Kosiński, 2006; Hegselmann & Krause, 2002; Martins, 2008). Bounded-confidence models are a class of models that suppose individuals change their opinions based on their relationships, when their opinions are already close to those of their peers. That is, if someone's opinion is very far away from my own, even if I have a relationship with them, I will not base my opinions on theirs. Many bounded-confidence models have been developed and studied. They include examinations of consensus formation (Dittmer, 2001; Fortunato et al., 2005), polarization (Hegselmann, 2020; Sirbu et al., 2019), and a large variety of model extensions for application to real-world opinions (Altafini & Ceragioli, 2018; Brooks & Porter, 2020; Hickok et al., 2022; Kan et al., 2021).

We consider polarization, and the notion that individuals may form their opinions by being contrarian. If I have an adversarial relationship with someone, I may specifically choose to hold an opinion that is different from theirs. Similar to other bounded-confidence models, we maintain the idea that individuals are mostly influenced by others whose opinions are already somewhat close to our own. We are most interested in understanding how collective opinions in this model behave. What types of relationships and community structures lead to strong polarization within a society? How might we extend those observations to real-world applications and data?

The chapter is organized as follows. We introduce the motivation for our model in Section 3.2 and define our model in Section 3.3. We present analytical results in Section 3.4, and perform numerical simulations on synthetic networks (Section 3.5). Conclusions follow in Section 3.6.

3.2 Background and Motivation

In this section, we introduce the motivation for our proposed model of opinion dynamics. First, we discuss political science research which motivates our modelling choices. This is followed by an introduction to the Hegselmann–Krause model for opinion dynamics, which we use as a starting point in the formulation of our model.

Political Science motivation

In political science it is common to think of ideologies as points in space, as being on the left or the right, liberal or conservative. This spatial view of politicians and individuals drives much of the work that is done on voting behavior, both at the individual and legislative levels, as well as the models of strategic behavior within Congress. The original conception of this model is often attributed to Downs and his median voter theory (Downs et al., 1957). This work was followed by further theoretical work on legislative organization (Baron, 1994; Hitt et al., 2017; Riker, 1980; Shepsle, 1979), electoral competition (Ansolabehere et al., 2001), and the courts (McNollgast, 1994) to name a few.

The most common method of obtaining ideological spacial estimates for members of congress is NOMINATE (Poole & Rosenthal, 1985). It uses the observed voting choices and an item response model (IRT) to recover spatial distances. This work has been expanded to include bridges over time to estimate changes in the distribution of congressional representatives across congresses (Poole & Rosenthal, 2000). More recently, such bridging techniques and new data sources have been used in order to get consistent measurements for politicians in different chambers as well as candidates who do not win their election (Bailey, 2007; Bonica, 2014; Clinton et al., 2004; Shor et al., 2010).

In this chapter we present a bounded confidence model in which there are both attractive and repulsive links between members, rather than the canonical model which simply has attractive links. The addition of repulsive links is motivated by the idea of varying salience of issues among members of congress. While representatives may have ideological positions that can be uncovered through voting behavior, there is reason to believe that politicians are drawn to fellow representatives with similar priorities. Therefore, working with other members of congress causes their ideologies to converge. In contrast, they make a point of distancing themselves from representatives whose salient issues run in opposition to them, regardless of other similarities. This would cause them to attempt to distinguish themselves. From an electoral perspective, this distinguishing is important and has not yet, to the our knowledge, been accounted for in spatial models.

Bounded-Confidence models

The model we propose is a variant of the Hegselmann–Krause (HK) model (Hegselmann & Krause, 2002). The HK model considers the opinions of a group of interacting agents who influence each other. In the HK model, agents are modelled in a network, with connections between them. Agents who are connected to each other will affect each others' opinions, but only if their opinions are sufficiently close. That is, even if two agents are connected, if their opinions are far apart, they will not take each other into consideration as they form new opinions.

The precise mathematical statement of HK is as follows. Suppose $G = (V, E)$ is a network, with associated adjacency matrix A. Then at each time step t, we denote the opinions of nodes $i \in V$ with the opinion vector $\vec{x}(t)$. We associate to the model a confidence bound c. Opinions are updated according to the following rule:

$$x_i(t+1) = \frac{\sum_{j\in V} A_{ij} x_j(t) \vec{1}_{|x_j(t)-x_i(t)|<c}}{\sum_{j\in V} A_{ij} \vec{1}_{|x_j(t)-x_i(t)|<c}}. \tag{3.1}$$

That is, at time $t+1$, we examine all neighbors of i which are within the confidence bound, and then average their opinions. Note that we can reformulate this as

$$x_i(t+1) = x_i(t) + \frac{\sum_{j\in V} A_{ij} (x_j(t) - x_i(t)) \vec{1}_{|x_j(t)-x_i(t)|<c}}{\sum_{j\in V} A_{ij} \vec{1}_{|x_j(t)-x_i(t)|<c}}. \tag{3.2}$$

While the model formulation given above is for 1-dimensional opinions, by expanding the notation, the same averaging scheme can be used for higher-dimensional opinions as well.

Previous studies of the HK model have found that it converges in polynomial time (Bhattacharyya et al., 2013). In Hegselmann and Krause, 2002, these authors also investigated the steady states of the model. Specifically, the HK model suggests that as the confidence bound increases, there is a transition between three types of steady states. For low confidence bounds, the steady state has many possible opinions and no particularly dominant opinions (we refer to this as *fragmentation*). As the confidence bound increases, steady states begin to exhibit only a small number of dominant opinions (*polarization*). For confidence bounds beyond a certain threshold, we observe only a single dominant opinion (*consensus*). These three different steady states can be seen in Figure 3.1. In later sections, we will discuss how the steady states of our model compare.

3.3 Model Statement

The key mechanism that drives our model is the inclusion of repulsive edges, so that individual nodes can push each other way. Suppose $G = (V, E)$ is a network with associated adjacency matrix A. For any pair of nodes i, j, if $A_{ij} = 1$, there is an attractive edge between them. If $A_{ij} = -1$, there is a repulsive edge. Otherwise, $A_{ij} = 0$ and there is no edge between the nodes. As in the original model, we assume

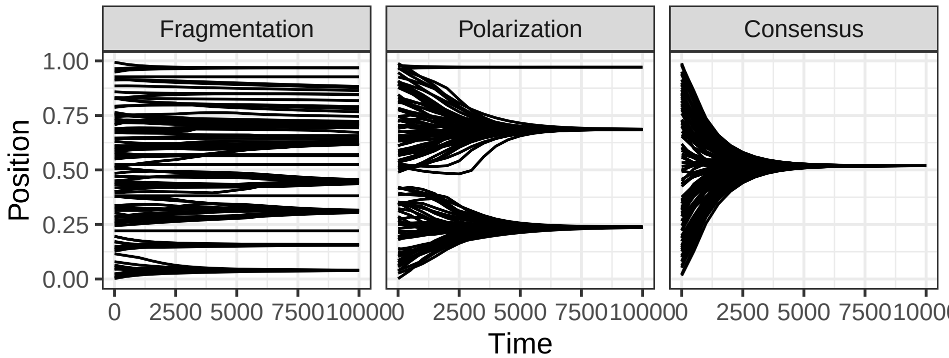


Figure 3.1: In this figure an Erdős–Renyi Random Graph is created with connection probability of 25%, the evolution three confidence intervals (0.05, 0.2, 0.6) are shown in order to demonstrate the three steady states of the model.

that for all i, $A_{ii} = 1$. This assumption ensures that a node's existing position is included in the calculation for the next step. We let $\vec{x}$ be the vector of opinions, with $x_i(t)$ representing the opinion of node i at time t.

We define the variable M as follows:

$$M_{ij}(t) = \begin{cases} x_j(t) - x_i(t) & A_{ij} \geq 0 \\ \text{sign}(x_j(t) - x_i(t))|c - |x_j(t) - x_i(t)|| & A_{ij} = -1, |x_j(t) - x_i(t)| > 0 \\ \text{sign}(j - i)c & A_{ij} = -1, x_j(t) = x_i(t). \end{cases} \tag{3.3}$$

Intuitively, M_{ij} represents a signed distance which node i will potentially travel because of node j. The effect of M is that repulsive forces grow weaker as nodes move farther away from each other. Note that the third row of M_{ij} covers the case where two nodes have the same opinion and repulse each other. In this case, the node with the higher index is pushed towards a higher opinion, while the node with the lower index is pushed towards a lower opinion. In simulation, this situation is unlikely, as it is rare that two nodes which are repulsed share the precise same value. We updated opinions using the following rule:

$$x_i(t+1) = x_i(t) + \frac{\sum_{j \in V} A_{ij} M_{ij}(t) \mathbf{1}_{|x_j(t) - x_i(t)| < c}}{\sum_{j \in V} |A_{ij}| \mathbf{1}_{|x_j(t) - x_i(t)| < c}}. \tag{3.4}$$

Note that if there are no repulsive edges, Equation 3.4 reduces precisely to the HK model as stated in Equation 3.2.

Equation 3.3 is incorporated into the model in order to aid convergence. To see why, suppose that we instead naively replaced $M_{ij}(t)$ in Equation 3.4 with $x_j(t) - x_i(t)$.

We can quickly see from the following three node example in Figure 3.2 that opinions may oscillate forever, with attractions pulling opinions together which then repulse each other when they enter the bounded confidence interval. We further discuss the steady states of this model in Section 3.4 and Section 3.5.

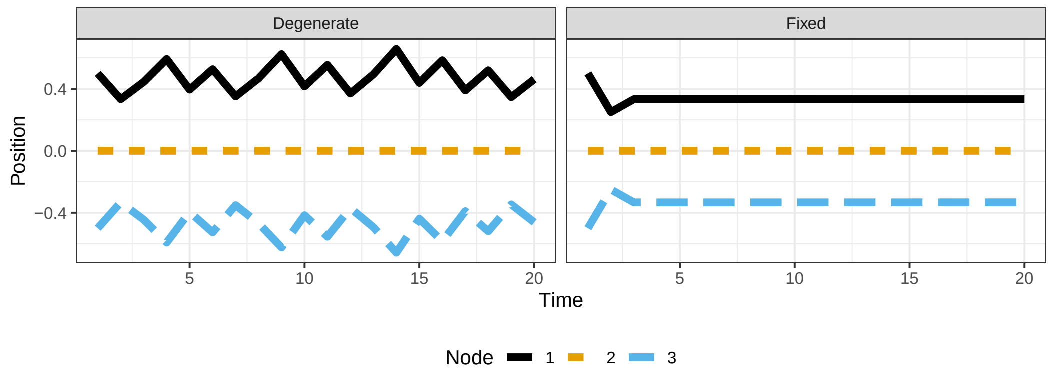


Figure 3.2: Example with three nodes, where the top and bottom node repulse each other, while the central node attracts both of the others. Without the incorporation of distance scaling, the central node pulls the two outer nodes towards it until they are within confidence of each other, at which point they push each other away until they are no longer within confidence of each other. However, because they are still within confidence of the central node, they are pulled back in, and the cycle repeats so that the model never converges. Using the value of $M_{ij}(t)$ represented in Equation 3.3.

When there exist repulsive edges, our model gives rise to several forms of behaviors that differ from the standard HK model. First, the initial range of opinions does not necessarily bound the final set of opinions. In our model, if there exist enough repulsive edges, it is possible for the final opinions to span a much wider range than the initial opinions (as shown in Figure 3.3). We prove a bound on final opinions in Theorem 1.

Second, with repulsive edges, connected nodes within confidence of each other may not converge to a single opinion. In HK, we can consider the *receptivity subgraph*, or the subgraph of G where edges are pruned if they connect nodes outside confidence bound of each other. In HK, the connected components of the receptivity subgraph will converge to single opinions. In our model, because tensions between attractive and repulsive edges exist, it is possible for nodes to converge to an opinion which is different from its neighbors at stopping time. For example, the same three-node example in Figure 3.2 converges to a state where all three nodes are still connected and within confidence of each other, but do not have the same opinion.

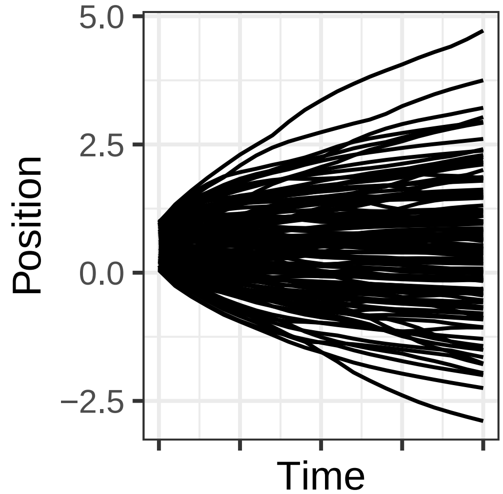


Figure 3.3: This image shows a simulation where final opinion width was wider than initial opinion width. The edges are created as described in Equation 3.6 where $p_1 = 40\%$, $p_2 = 80\%$ and a confidence bound of 1.6.

3.4 Analytical Results

In this section, we give some simple proofs about steady states of our model. We note that while the model converges in most cases (we suspect all, based on the analytical results in Section 3.5), it is not necessarily true that the final opinions are bounded by the initial opinions. The inclusion of negative edges means that repulsive forces between nodes can push the final opinions well outside the bounds of the initial opinions, though this process can only continue so far before nodes are no longer within confidence of each other. To that end, we propose a simple bound on final opinions based on the number of negative edges.

Theorem 1. *Let $G = (V, E)$ be a network with n nodes and confidence bound c. Suppose that G is the complete graph, and that every edge $(i, j) \in E$ is repulsive (that is $A_{ij} = -1$). Suppose also that we have initial opinions $x_i(0)$ such that $|x_i(0) - x_j(0)| < c$. Then the model converges, and* $\max_{i,j} \left|x_i(T) - x_j(T)\right| = (n-1)c$.

In order to provide some basic intuition, we first prove the bound in the simplest case (that of 2 nodes).

Proposition 1. *Let $G = (V, E)$ be a network with 2 nodes , so that $V = \{0, 1\}$. Then the dynamical process described by Equation* 3.4 *converges, and at time of* *convergence T,*

$$|x_0(T) - x_1(T)| \leq \max\{c, |x_0(0) - x_1(0)|\}.$$

Proof. Suppose that there are no edges, or $E = \emptyset$. Then the model converges at time $T = 1$, and

$$|x_0(T) - x_1(T)| = |x_0(0) - x_1(0)|.$$

Now suppose that $E = \{(x_0, x_1)\}$, and $|x_0(0) - x_1(0)| \geq c$. Then the update rule will result in no changes, the model converges at $T = 1$

$$|x_0(T) - x_1(T)| = |x_0(0) - x_1(0)|.$$

Now suppose that $|x_0(0) - x_1(0)| < c$. If $A_{01} = -1$, the two nodes repel each other. Then

$$\begin{aligned}
x_1(1) &= x_1(0) + \frac{(x_1(0) - x_1(0)) + c - (x_1(0) - x_0(1))}{2} \\
x_0(1) &= x_0(0) + \frac{(x_0(0) - x_0(0)) - (c - (x_1(0) - x_0(0))}{2} \\
x_1(1) - x_0(1) &= x_1(0) - x_0(0) + \frac{2(c - (x_1(0) - x_0(0)))}{2} \\
&= c
\end{aligned}$$

and $x_1(1) - x_0(1) >= c$, so that after this time, these two nodes will no longer affect each other, and cannot push each other further, so the model has converged, and $\max_{i,j} |x_0(T) - x_1(T)| \leq c$.

If $A_{01} = 1$, the two nodes attract each other, and the model is equivalent to standard Hegselmann–Krause, so that we have convergence to a single point and

$$|x_0(T) - x_1(T)| = 0.$$

This covers all possible cases, and the proposition is proven. □

The main point to note from this two-node proof is that the repulsive forces between any two nodes will contribute to pushing them apart to a distance of precisely c. Also note that any node cannot move more than c in any direction over the course of one timestep, because $|M_{ij}(t)| \leq c$. In order to prove Theorem 1 we need several lemmas and corollaries. The proofs of these can be found in Appendix C.1. We state them here so that we can use them for the final proof.

Lemma 1. *Suppose $i \in V$ a node. Define the following sets:*

$$V_i^+(t) = \left\{ j \in V : A_{ij} = 1 \text{ and } |x_j(t) - x_i(t)| < c \right\}$$
$$U_i(t) = \left\{ j \in V : A_{ij} = -1 \text{ and } \left[(0 < x_j(t) - x_i(t) < c) \text{ or } \left(x_j(t) = x_i(t) \text{ and } j > i \right) \right] \right\}$$
$$L_i(t) = \left\{ j \in V : A_{ij} = -1 \text{ and } \left[(0 < x_i(t) - x_j(t) < c) \text{ or } \left(x_j(t) = x_i(t) \text{ and } i > j \right) \right] \right\}.$$

Then

$$x_i(t+1) = \frac{\sum_{j \in V_i^+(t)} x_j(t) + \sum_{j \in U_i(t)} (x_j(t) - c) + \sum_{j \in L_i} (x_j(t) + c)}{|V_i^+(t)| + |U_i(t) + |L_i(t)|}. \tag{3.5}$$

Intuitively, this lemma tells us that the update rule moves $x_i(t)$ to $x_i(t+1)$ by taking an average of several opinions. The set $V_i^+(t)$ contains nodes i is attracted to at time t. The set $U_i(t)$ contains nodes which repulse i at time t, and which will push i's opinion lower. The set $L_i(t)$ contains nodes which repulse i at time t, and which will push i's opinion higher. Equation 3.5 tells us that we can take the average of $x_j(t)$ for $j \in V_i^+(t)$, $x_j(t) - c$ for $j \in U_i(t)$, and $x_j(t) + c$ for $j \in L_i(t)$ to determine $x_i(t+1)$.

Lemma 2. *Let $i \in V$ at time t, and let $W(t) \subset V$ be a set of nodes such that $W(t)$ is completely contained in $V_i^+(t) \cup U_i(t) \cup L_i(t)$. Define*

$$\overline{W}(t) = \frac{\sum_{j \in W(t)} x_j(t)}{|W(t)|}$$

to be the average of $x_j(t)$ for all $j \in W(t)$. Then we can rewrite Equation 3.5 *as*

$$x_i(t+1) = \frac{\sum_{j \in \left(V_i^+(t) \cup U_i(t) \cup L_i(t)\right) \setminus W(t)} x_j(t) + \sum_{j \in W(t)} \overline{W}(t) + (|L_i(t)| - |U_i(t)|)\, c}{|V_i^+(t)| + |U_i(t) + |L_i(t)|}.$$

This lemma allows us to replace a group of opinion values of individual nodes with the average of opinion values across the group, in certain situations.

Lemma 3. *Let $G = (V, E)$ be a network with n nodes and m edges with confidence bound c. Suppose that every edge in G is repulsive. At time t, suppose $x_i(t) > x_j(t)$ for all other nodes j, so that i is the node with the highest opinion value at time t. Then $x_i(t+1) > x_j(t+1)$ for all j.*

Corollary 1. *Let $G = (V, E)$ be a network with n nodes and m edges with confidence bound c. Suppose that every edge in G is repulsive. At time t, let $M = \{i : x_i(t) \geq x_j(t) \forall j \in V\}$. Then $x_{\max_M i}(t+1) > x_j(t+1) \forall j \in V$.*

Corollary 2. *Let $G = (V, E)$ be the complete network with n nodes with confidence bound c. Suppose that every edge in G is repulsive. At time t, let $M = \{i : x_i(t) \leq x_j(t) \forall j \in V\}$. Then $x_{\min_M i}(t+1) < x_j(t+1) \forall j \in V$.*

Lemma 4. *Let $G = (V, E)$ be a network with n nodes and m edges with confidence bound c. Suppose that every edge in G is repulsive. At time t, suppose $x_i(t) > x_j(t)$ for all other nodes $j \in V$, so that i is the node with the highest-valued opinion at time t. Suppose that there is some node j such that $x_i(t) - x_j(t) < c$, and that j has the highest-valued opinion of all such nodes. Then*

$$\frac{2c}{2 + |L_{ij}(t)| + |L'_{ji}(t)|} \leq x_i(t+1) - x_j(t+1) \leq \frac{(|L'_{ji}(t)| + 2)c}{2 + |L_{ij}(t)| + |L'_{ji}(t)|}.$$

Corollary 3. *Let $G = (V, E)$ be a network with n nodes and m edges with confidence bound c. Suppose that every edge in G is repulsive. At time t, suppose $x_i(t) < x_j(t)$ for all other nodes $j \in V$, so that i is the node with the lowest-valued opinion at time t. Suppose that there is some node j such that $x_j(t) - x_i(t) < c$, and that j has the lowest-valued opinion of all such nodes. Then*

$$\frac{2c}{2 + |U_{ij}(t)| + |U'_{ij}(t)|} \leq x_i(t+1) - x_j(t+1) \leq \frac{(|U'_{ij}(t)| + 2)c}{2 + |U_{ij}(t)| + |U'_{ij}(t)|}.$$

Lemma 4 and Corollary 3 give us precise conditions under which the nodes with the most extreme opinions will no longer be within confidence bound of any other nodes. Specifically, in order for the node with the highest-value opinion to lose connection with all other nodes, it must be true that the only node it is still influenced by is the node with the second-highest-value opinion, and that neither of the two nodes is influenced by any other nodes. Otherwise, they will remain within confidence of each other, even as the node with highest-value opinion remains the most extreme node and continues to have its opinion pushed upward.

We conclude with one more lemma about the bound on the width of the gap between consecutive nodes.

Lemma 5. *Let $G = (V, E)$ be the complete network with n nodes and confidence bound c. Suppose that every edge in G is repulsive. At time t, suppose that i and j are nodes such that $(i, j) \in E$, $x_i(t) > x_j(t)$, and $x_i(t) - x_j(t) < c$, and there exist no nodes k connected to i or j such that $x_i(t) > x_k(t) > x_j(t)$. Then*

$$|x_i(t+1) - x_j(t+1)| \leq c.$$

At this point we have the tools to return us to Theorem 1. As a reminder, the statement of the theorem was:

Theorem 1. *Let $G = (V, E)$ be a network with n nodes and confidence bound c. Suppose that G is the complete graph, and that every edge $(i, j) \in E$ is repulsive (that is $A_{ij} = -1$). Suppose also that we have initial opinions $x_i(0)$ such that $|x_i(0) - x_j(0)| < c$. Then the model converges, and* $\max_{i,j} \left|x_i(T) - x_j(T)\right| = (n-1)c$.

Proof. The intuition for this theorem is as follows: for any repulsive edge (i, j), nodes i and j will repel each other until

$$|x_i(t) - x_j(t)| >= c$$

at some future time t. If every edge is repulsive, we must have a distance at least c between every pair of nodes connected by an edge in order for the model to converge. Intuitively, the nodes will always continue to push each other outward until they reach a distance of c, and no further, so that the final convergent state of the model will occur when there are gaps of at least c between all of the m edges in the original graph. However, from Lemma 4, the gaps will have precisely width c, so that the bound holds.

From Corollary 1 and Corollary 2, at time 1, there must be a highest and lowest-value opinion node. By Lemma 3, for $t > 1$, these nodes will always be the highest and lowest-value opinion nodes. Call these nodes i_{max}, i_{min}.

Because G is the complete graph, and all edges are repulsive, we can observe that i_{max} and i_{min} will have their opinions pushed outward, since initially every node is within confidence of every node. Additionally, from Lemma 2, we can observe that i_{max} will be pushed in the direction of $\overline{\{j \neq i\}}(0) + c$, so that the nodes with opinions much lower valued than the average will start to drop out of confidence of i_{max}. Further, from Lemma 4, i_{max} will remain within confidence of at least one node as long as it is within confidence of at least 2 nodes in the previous timestep. Combining these lemmas, we can see that eventually at time t', i_{max} will be within confidence of exactly one other node.

Let i'_{max} be the singular node for which $x_{i_{max}}(t') - x_{i'_{max}}(t') < c$. Then we can follow the same proof procedure as in Lemma 3 to prove that $x_{i'_{max}}(t'+1) > x_j(t'+1)$ for all $j \in V$ other than $j = i_{max}$, and that none of the remaining nodes can be pushed into confidence of $i_m ax$. We do not include the procedure here because of its similarity to Lemma 3, but the key observation that drives the proof is that there is only a single node i_{max} exerting downward pressure on i'_{max} (if a very high number of nodes were exerting downward pressure on i'_{max}, it would be possible for i'_{max} to lose its position as the node with second-highest-value opinion). This allows us to rewrite the $x_{i'_{max}}$ as an average of values which preserve the order of i_{max}, i'_{max}, and the remaining nodes. Similarly, we can show that there is some time after which the node with the second-lowest-value opinion will always remain the node with the second-lowest-value opinion.

We continue in this manner, proceeding from the nodes with the highest and lowest-value opinions inwards until we show that after some time, the nodes' opinions must remain in a fixed order.

From this point on, we observe that from Lemma 5, the gap between any two consecutive nodes is bounded by c. Because of our initial conditions on $x_i(t)$, it is impossible for any gap between consecutive nodes to be larger at any point. If any two nodes have a gap smaller than c, we will not have converged, as the repulsion between the two nodes will push them apart in the next time step. All nodes will push each other apart until the gap between any two consecutive nodes is precisely c, at which point the model has converged. Because there are n nodes, this tells us

$$\max\{|x_i(T) - x_j(T)|\} = (n-1)c.$$

□

The proof for Theorem 1 relies on all edges being repulsive, thereby preserving the ordering of the nodes. This property does not necessarily hold when there are both attractive and repulsive edges. However, we suspect based on numerics that the following theorem is also true:

Theorem 2. *Suppose $G = (V, E)$ is a network with n nodes, m edges, and confidence bound c. Let m_r be the number of repulsive edges in the network. Then the model* *converges and*

$$\max_{i,j \in V}\{x_i(T) - x_j(T)\} \leq \max\{\max_{i,j \in V}\{x_i(0) - x_j(0)\}, mc\}.$$

Intuition. The worst case for this model assumes that all repulsive nodes end up at least c apart from each other, so if all nodes start out within confidence of each other, the worst case is one in which all nodes with repulsive edges are chained together in consecutive order along a line of m edges, in which case the width of their opinions cannot exceed mc, since the bounds in Lemma 5 should apply and prevent any individual gap from growing wider. The only way a gap could grow wider is if there are attractive nodes pulling the repulsed nodes further apart, in which case those attractive nodes either have repulsive forces between them, and have already been considered in the line, or must have started farther apart to begin with, in which case we look at $\max_{i,j \in V}\{x_i(0) - x_j(0)\}$.

Because we cannot rely on nodes remaining in fixed order in this case, we cannot use the same technique as in Theorem 1 to prove convergence and a bound. However, in practice, we observe that the range of final opinions increases with the number of repulsive edges, and that in practice the bound of mc is not very tight (this is to be expected, as, for example, in the case of the complete graph in Theorem 1, the bound is considerably smaller). To see numerics showing that the range of final opinions scales with number of repulsive edges and c, see Figure 3.5 and associated discussion. □

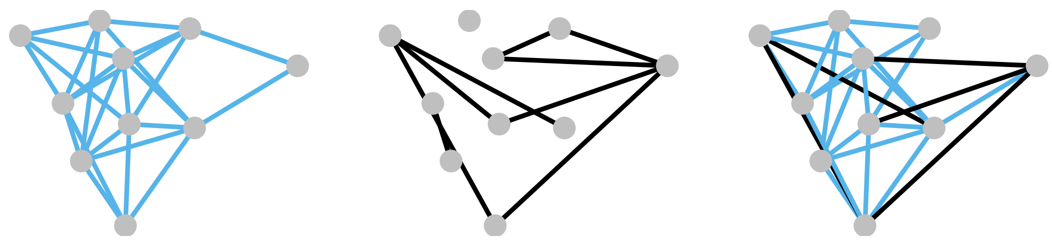
3.5 Numerical Results on Synthetic Networks

In this section we present analysis of numerical simulations on a variety of random networks, chosen for their usage in modelling social structures (Siegel, 2009). We go through each of the structures and their simulation results in turn.

Erdős–Renyi

We begin with an adaptation of Erdős–Renyi (ER) networks as a simple random network model. To achieve a random network with both positive and negative edges, we generate two ER networks, $G_1 = G(n, p_1)$ and $G_2 = G(n, p_2)$, with associated adjacency matrices A_1 and A_2. The total network G, is then the network derived from the adjacency matrix $A_1 - A_2$. A visual of this method can be seen in Figure 3.4. In the subsequent network the probability of each type of edge between any set of nodes can be written as:

$$P((i,j) \in E) = \begin{cases} 0 & (1-p_1)(1-p_2) + p_1 p_2 \\ 1 & p_1(1-p_2) \\ -1 & (1-p_1)p_2. \end{cases} \tag{3.6}$$



(a) G1: positive node network with $p_1 = 0.6$

(b) G2: negative node network with $p_2 = 0.2$

(c) G3: final network with $p_1 = 0.6$ and $p_2 = 0.2$.

Figure 3.4: An example of the generation of the ER network with attractive and repulsive edges

To create the simulation results, 100 trials were run with all combinations of the following parameters:

$$\begin{aligned} p_1 &\in (0.2, 0.4, 0.6, 0.8, 1) \\ p_2 &\in (0.0, 0.2, 0.4, 0.6, 0.8) \\ c &\in (0.05, 0.4, 0.8, 1.2, 1.6). \end{aligned}$$

For each trial, a random set of initial opinions is generated and the model is applied for 10000 iterations. In Figure 3.5, one trial is shown for each set of parameters when p_1 is set to 0.4. This trial was chosen randomly and all other trials qualitatively look the same.

The final range of opinions gets wider with both p_2 and c once repulsive edges are included. These results are in line with expectations. As p_2 increases, so does the number of negative connections, resulting in more repulsive forces between nodes, pushing opinions apart. As c increases, nodes have more neighbors. For $p_2 << p_1$, the attractive forces overpower the repulsive ones, so that higher c leads to more consensus, as in standard HK models. For $p_2 >> p_1$, the opposite is true—repulsive forces overpower attractive ones, and nodes push each other further apart for higher c, resulting in a wider spread of opinions.

In particular, we observe that the proportion of $\frac{p_2}{p_1}$ seems to be the driving factor in the range of final opinions. To draw clearer conclusions, we look at *opinion spread*, which we define as the following quantity:

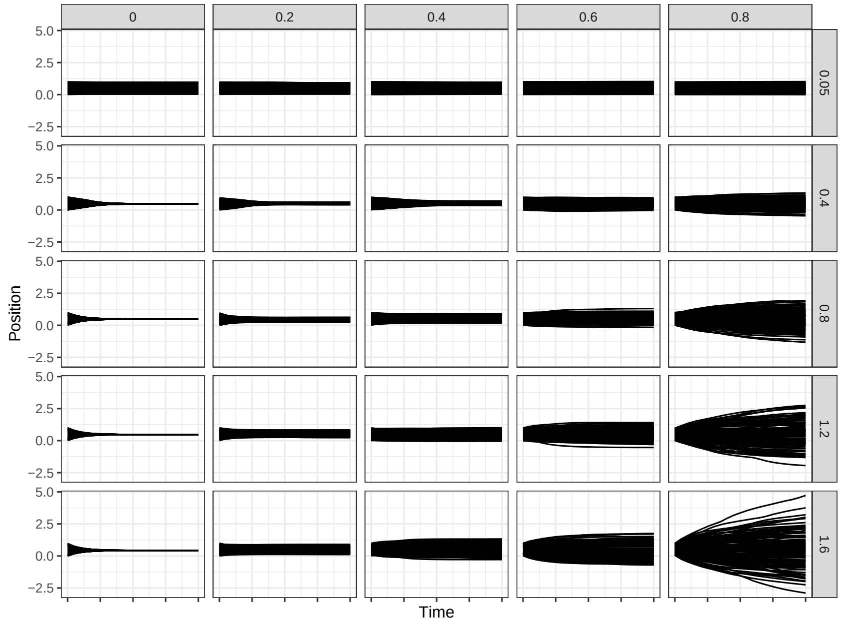


Figure 3.5: For all plots, $p_1 = .4$. The horizontal axis represents p_2, while the vertical axis represents c. Note that as p_2 increases, the range of final opinions gets wider. For low values of p_2, as c increases, the range of final opinions becomes narrower (closer to consensus). By contrast, for high values of p_2, as c increases, the range of final opinions becomes wider.

$$\frac{\max_{i,j} |x_i(T) - x_j(T)|}{\max_{i,j} |x_i(0) - x_j(0)|} \tag{3.7}$$

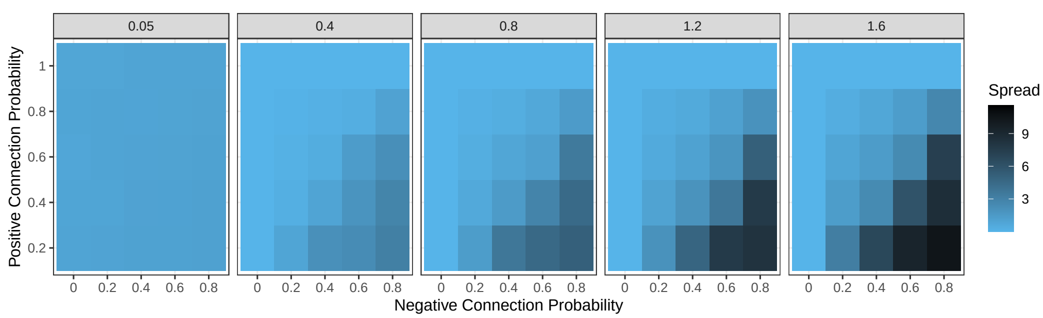


Figure 3.6: Heat map of opinion spread as a function of the probability of a connection in G_1 and G_2 iterated over confidence bound c. Data drawn from the mean of 100 trials for each set of parameters with parameters $p_1 \in \{0.2, 0.4, 0.6, 0.8, 1\}$, $p_2 \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$, and $c \in \{0.05, 0.4, 0.8, 1.2, 1.6\}$.

In Figure 3.6, we plot average opinion spread across trials as a function of the proportion $\frac{p_2}{p_1}$ and confidence bound c in a heat map. We observe similar trends as in Figure 3.5, with higher proportions $\frac{p_2}{p_1}$ leading to higher values of opinion spread, and the influence of c on opinion spread depending on $\frac{p_2}{p_1}$. In the following examples, we will similarly see that opinion spread is largely controlled by the negative edges in the network, but that the addition of more structure to the network will influence opinion formation in interesting ways.

Stochastic Block Models

Next, we adapt a Stochastic Block Model (SBM) in order to incorporate both positive and negative edges. In these networks, each node is assigned to a group $k \in K$. The probabilities of connections when $i_k = j_k$ is different than when $i_k \neq j_k$. This enforces structure within the network. As in Section 3.5, we generate this network through two sub networks. In this case, the process begins with two SBM networks, $G_1 = G(n, p_1, \rho)$ and $G_2 = G(n, p_2, \rho)$, with associated adjacency matrices A_1 and A_2. The variable p is the probability of having a connection with another node in the same cluster while $p_{1(2)}\rho$ is the probability of having a connection with a node in a different cluster. If G is network represented by the adjacency matrix given by $A_1 - A_2$ we have the generated network edge probabilities:

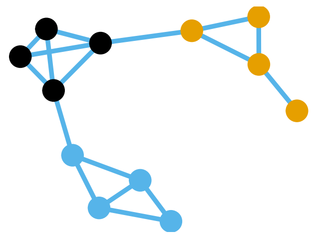
$$P((i,j) \in E)_{i_k = i_j} = \begin{cases} 0 & (1-p_1)(1-p_2) + p_1 p_2 \\ 1 & p_1(1-p_2) \\ -1 & (1-p_1)p_2 \end{cases} \tag{3.8}$$

$$P((i,j) \in E)_{i_k \neq i_j} = \begin{cases} 0 & (1-p_1\rho)(1-p_2\rho) + \rho^2 p_1 p_2 \\ 1 & p_1\rho(1-p_2\rho) \\ -1 & (1-p_1\rho)p_2\rho. \end{cases} \tag{3.9}$$

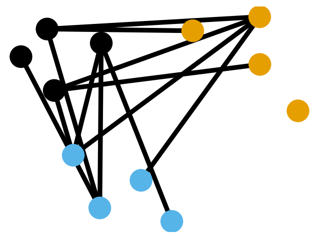
A sample of the generative process can be seen in Figure 3.7, where the blue edges represent positive edges and the black negative. Each color of nodes represents a group $k \in K$.

Again, in order to create simulation results, 100 trials are run for all combinations of the parameters

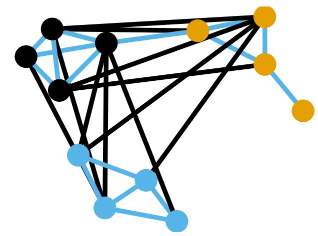
$$\begin{aligned} \rho &\in (0, 0.2, 0.4, 0.6, 0.8, 1) \\ p_1 &\in (0.2, 0.4, 0.6, 0.8, 1) \\ p_2 &\in (0.0, 0.2, 0.4, 0.6, 0.8) \\ c &\in (0.05, 0.4, 0.8, 1.2, 1.6). \end{aligned}$$



(a) G1: positive node network with $p_1 = 0.85$ and $\rho = 0.05$



(b) G2: negative node network with $p_2 = 0.3$ and $\rho = 0.05$



(c) G3: final network $p_1 = 0.85$, $p_2 = 0.3$ and $\rho = 0.05$

Figure 3.7: An example of the generation of the SBM network with attractive and repulsive edges

For each trial, a random set of initial opinions is generated and the model is applied. The full histories of an example run when $p_1 = 0.8$ and $p_2 = 0.2$ can be seen in Figure 3.8. It can be seen that each group fully converges on itself, while the confidence interval determines how dispersed the groups are from each other. As ρ increases, the full set begins to converge. As is the case with the ER models, for certain combinations the spread at steady-state is larger than that at the start. From these values we can see that when p_1 and p_2 are locked, higher confidence bounds result in a larger terminal spread, as does lower percentages of cross-group edges (ρ).

As with the ER model, with the SBM we first look at steady-state opinion spread. The results can be seen in Figure 3.9. Note, that when $\rho = 1$ the SBM model is equivalent to the ER model for the same parameters. We therefore have that the final row is identical to Figure 3.6. We note that as the value of ρ shrinks, the final spread increases. Otherwise, the trends found for the ER model are consistent.

In the case of SBMs, since each vertex $i \in V$ is assigned to a group in k, we are also interested in clustering in addition to opinion spread. We introduce a measures of how close vertices are to in-group vertices versus out-group vertices. In order to calculate this, which we call *proportional spread*, first we find the average in and out-group distances as:

$$\mathcal{I}_T = \frac{1}{|k|} \sum_{k_i \in k} \frac{1}{|k_i|^2} \sum_{j \in k_i \subset V} \sum_{\ell \in k_i \in \subset V} |x_j(T) - x_\ell(T)| \tag{3.10}$$

$$\mathcal{O}_T = \frac{1}{|k|} \sum_{k_i \in k} \frac{1}{|k_i|} \sum_{j \in k_i \subset V} \frac{1}{|k| - |k_i|} \sum_{\ell \in \subset V/k_i} |x_j(T) - x_\ell(T)|. \tag{3.11}$$

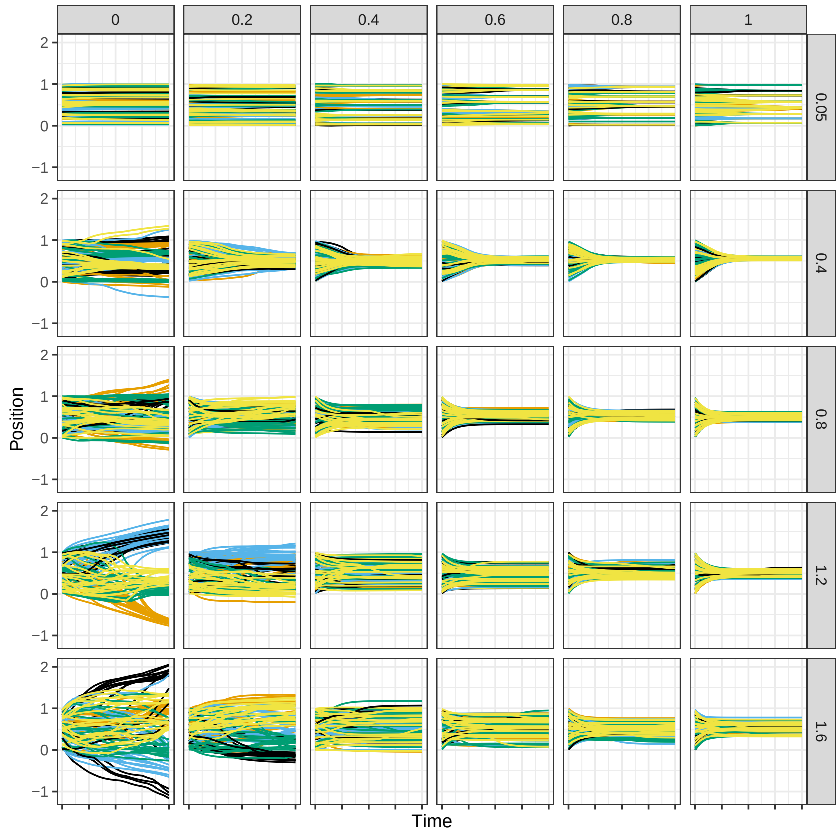


Figure 3.8: Example of the paths taken when the graph is distributed according to the Stochastic Block Model scheme laid out in Equation 3.9. In this case $p_1 = 0.8$ and $p_2 = 0.2$ are both locked the confidence bound varies between the rows and the percent of cross group links varies over columns. Each color represents a different group. There are five groups in these examples.

Since the end-spread of the samples differ, in order to appropriately compare them we look at the ratio, that is

$$\mathcal{PS}_T = \frac{\mathcal{I}_T}{\mathcal{O}_T}. \tag{3.12}$$

Smaller values of proportional spread imply same-group nodes are significantly closer to each other than different-group nodes. Thus, small values imply increased separation by group. In Figure 3.10 the spread as well as $\mathcal{O}_0$ and $\mathcal{I}_0$ can be seen. It is clear that the values range most of the space, in addition $\mathcal{O}_0 \approx \mathcal{I}_0$. Thus, uniformly, at the beginning of the trials we have $\mathcal{PS}_0 = 1$.

The plots of proportional spread at steady-state over the trials can be seen in Figure 3.11. There are many things to note in this figure. First, we observe that clustering is most clear with higher values of p_1 and lower values of ρ. The former

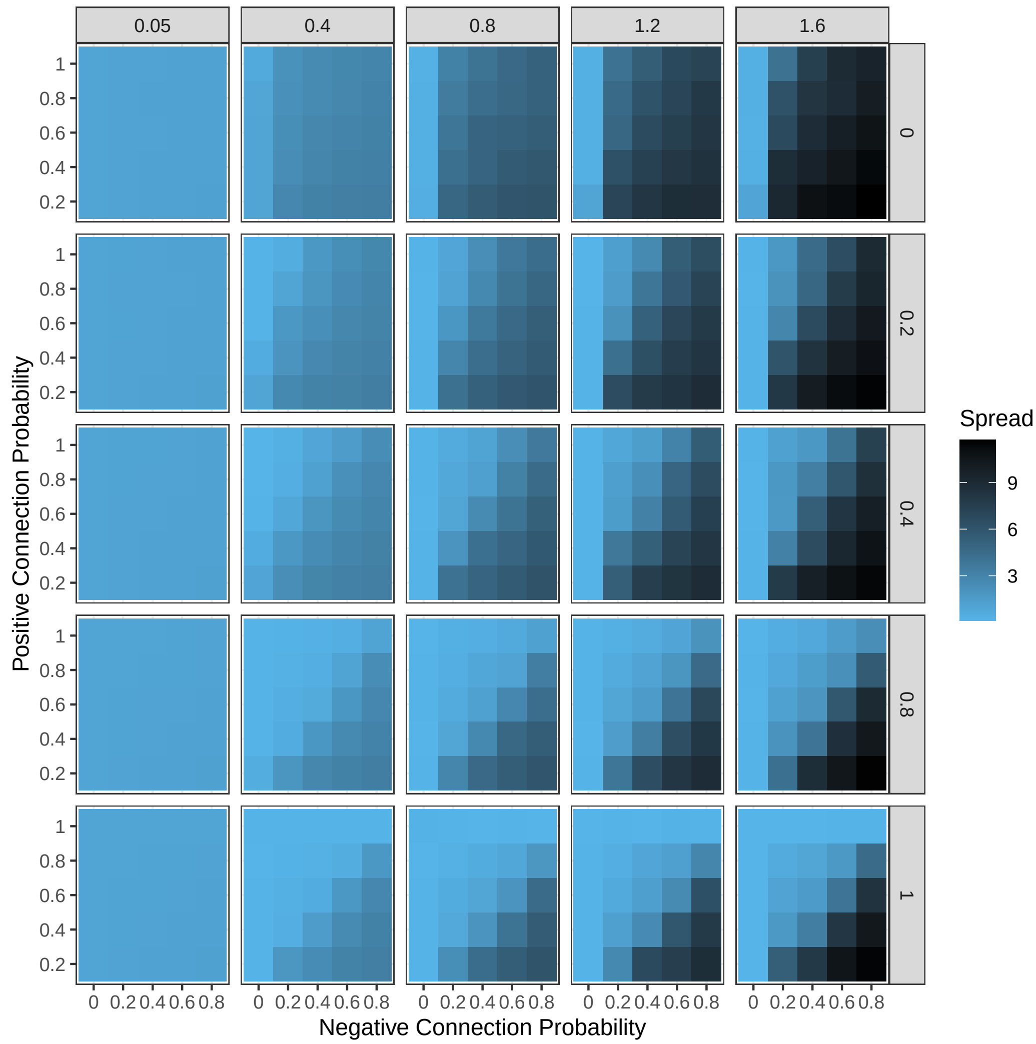


Figure 3.9: Heat map of the opinion spread of Stochastic Block Model trials. The rows represent ρ, the percent of positive(negative) connection values that are out (in) group, while the columns are c the confidence bound.

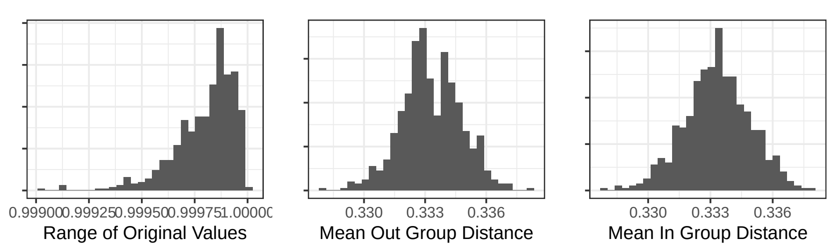


Figure 3.10: Histograms of metrics at the the start of each trial, useful for comparing the final results.

indicates higher level of in-group positive connections while, in terms of in group connections, the latter represents fewer out-group positive connections. This trend

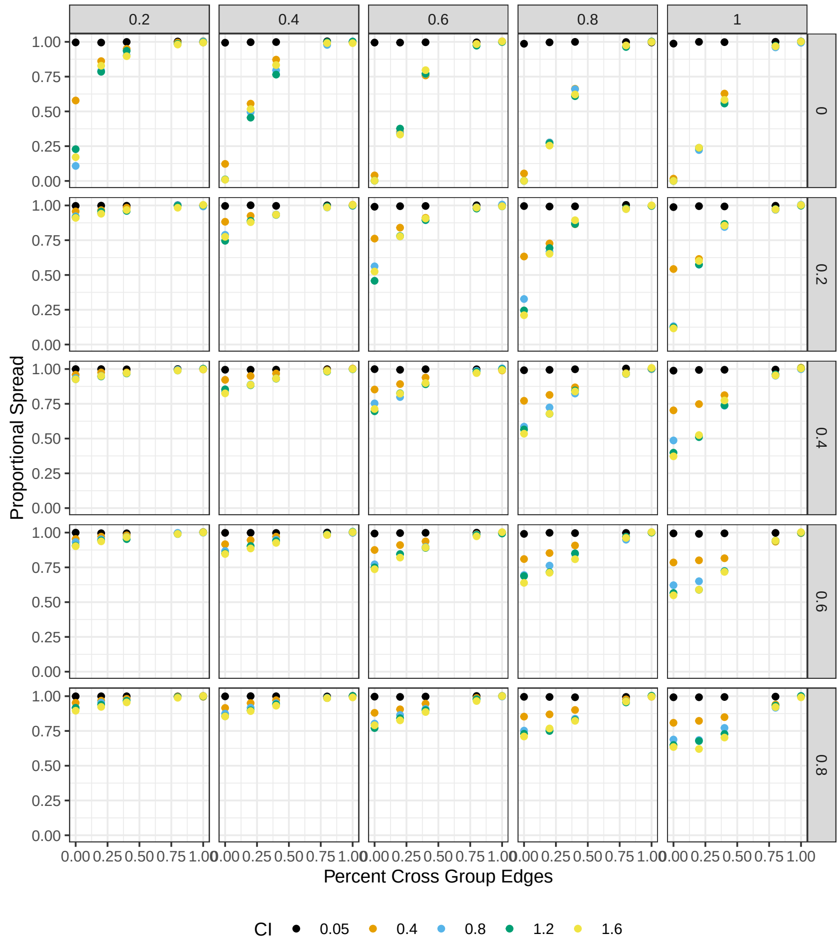


Figure 3.11: Proportional spread for Stochastic Block Model trials. The rows represent p_2, the base probability of an out-group negative connection, while the columns are p_1 the base in-group positive probability. The x-axis is ρ the percent of cross group edges and the color represents the confidence bound. A value of 1 means that nodes are equally as close to in and out-group nodes. Smaller values mean groups are increasingly clustered.

thus is consistent with expectations—higher group connectivity implies a larger probability that the entire group is connected and attracting each other. The low out-group positive connections means that there is a weaker countervailing force pulling nodes towards nodes of opposing groups.

The values of p_2 are measures of repulsive forces. To a lesser extent, clustering also increases with lower values of p_2. This implies that repulsive connections between groups increases polarization. This was the major result we were searching for by including the repulsive elements in the model to begin with. However, the relative importance of positive versus negative connections was not expected. While the negative connection are necessary to create spread in this way, the density of positive connections seems to be more important in determining the level of polarization.

3.6 Future Work and Conclusions

Bounded confidence models improve on existing political science measures of ideology such as ideal point and spatial estimates by enabling researchers to include a time dimension. While most alternative methods are stagnant, these models evolve over time based on the positions of the nodes and connections between them—representing the opinions and connections between individuals, respectively.

However, previous models of opinion dynamics have focused on the attractive nature of network connections. These result in a complete convergence in the steady-state for each receptivity subgraph, which is counter to observed opinion behavior both at the legislative and electorate level. When the term *polarization* is used, it still implies a shrinking of the overall opinion space. The model introduced in this paper, acknowledges that there are circumstances in which individuals seek to differentiate themselves from those in their network. This behavior leads to the possibility of an overall opinion space expansion.

In this paper, the basic bounds of this expansion were proven analytically in certain cases, with intuition provided for the general case. In addition, the steady-state behavior for random networks were analyzed numerically. These simulation results offered insight into the effects of various structural parameters on the model. It was clear that when there was structure in the network links (for instance in the Stochastic Block Model) group-based clustering emerged. This clustering was despite the random initial conditions provided. In addition, the size of the confidence bounds and the density of repulsive edges are both pivotal in the opinion spread.

The model introduced in this paper lends itself to complex opinion dynamics, where politicians or individuals want to differentiate themselves due to factors orthogonal to their expressed opinions. In future work we hope to explore how this model can help us understand political behavior, both at the representative and electorate levels. Initial ideas include using informative initial conditions for congressional networks. Additionally, this model can be used to look at ideological opinions of individuals who are influenced by pop culture associating ideological beliefs with other factors. This would introduce a variable connection to an ideal point, which then attracts or repulses the individual. The addition of repulsive forces make bounded-confidence models increasingly relevant for empirical and theoretical studies of public opinion.

Chapter 4

INCORPORATING LATENT CLASS IDENTITIES IN QUANTITATIVE WORK

With contributions from Melina Much.

Using the intersectional research paradigm, we investigate the best way to incorporate class as a social identity alongside race and gender. Class, while known to be linked to political attitudes and behaviors, is often overlooked in quantitative work because of its unwieldy operationalization. There are two difficulties with class: a lack of clarity in the definition of class within surveys and a sparsity of data. In this paper, we propose treating class as a context-dependent latent variable. This understanding can be modeled using a mixture model as a wrapper around existing methods. Through this approach, we can both identify which aspects of class are most salient for a given output as well as split the data into class-based groups. This overcomes both of the existing obstacles. Both socioeconomic status measures (SES) and subjective social status (SSS) measures of class can be included in the generation of a broader class definition. The model outputs an appropriate definition of class based on the assumption that class divides the dataset into groups. Our conception of class and the mixture model approach is supported by two empirical examples using data from the 2020 ANES. From these examples we confirm our hypothesis that class is output specific—class assignments are correlated but different between the two examples. In addition, through an analysis of the results we see that by pooling class, researchers could produce biased estimates for various coefficient estimates if only focusing on racial and gender identities.

4.1 Introduction

Class has been widely recognized as an important factor in understanding the link between economic context and political attitudes and behaviors as it deals with both identity and material circumstances (Huckfeldt, 1984; Jackman & Jackman, 1983). Its relation to political life more broadly has foundations in canonical social science stemming from Marx and Weber (Wright, 2005). The concept has been measured in a myriad of ways that spans across both socioeconomic status (SES) and subjective social status (SSS) (Diemer et al., 2013). Most often SES is used to determine class through material circumstances by measures of occupation, income, and education (Bartels, 2016; Goldthorpe & McKnight, 2006; Leighley & Nagler, 1992, 2007). Scholars that focus on SSS tend to understand it from the lens of identity and self-awareness/placement into one of the cultural definitions of class (Jackman & Jackman, 1973, 1983). While class's importance to political life is ubiquitous, it remains unclear whether or not SES or SSS should be used to measure the concept. Additionally, little work has been done to prioritize methods that capture both SES and SSS in a parsimonious way.

The literature on how to measure class is complicated by the notion of intersectionality (Collins, 1999; Crenshaw, 1989, 1991). Intersectionality, as articulated by Kimberle Crenshaw in 1989, attempts to show how racism, sexism, and classicism combine to generate unique lived experiences for individuals living at the intersection of marginalized groups. Originally, this was used to draw insight into how working-class Black women were overlooked in the United States legal system (Crenshaw, 1989, 1991). This work was expanded though to show the ways that race, sex, and class as identities shaped sociological and political phenomena (Hancock, 2007b; McCall, 2005; Simien, 2007; Weldon, 2006). It provides a framework for understanding lived experiences with oppressive structural power dynamics (racism, classicism, and sexism) along with a framework to articulate the nature of social identities on the individual level (Dhamoon, 2011; Yuval-Davis, 2015). An intersectional lens demands that class be studied in tandem with race and gender as they are multiply constituted social identities.

While there is a broad discussion on the best ways to include representations of intersectionality writ large in quantitative work, there has been less work on methodologies for how to include a comprehensive measure of class that takes into account its interlocked nature with race and gender. This discussion is necessary, as broadly, when quantitative scholars attempt to incorporate the theory in their work, class is often overlooked. We posit that there are two main reasons for this oversight. The first is that class can be measured in many ways within survey responses as opposed to its alternatives, race, and gender. A scholar can either use SES or SSS, but the measurement tactics will vary across different datasets. Some of the measures of SSS on their own provide skewed information about material circumstances, and different measures of SES exist across datasets.[1] Secondly, intersectional research

[1]While valuable to show perceived social class status, many working-class individuals often inflate their class status, and upper-middle-class individuals deflate it (Jackman & Jackman, 1983; Sosnaud et al., 2013).

is always challenged with the sparsity of data and incorporating class further subsets already small datasets for racial and ethnic minorities of different genders (Barreto et al., 2018; Frasure-Yokley, 2018; Junn & Masuoka, 2008; Much, 2022). When a scholar chooses the intersectional research paradigm and wants to include a four-way SSS measure, it would further split the groups into the data making traditional regression tactics less suitable.

In order to overcome these barriers, we introduce the use of mixture models to serve as a wrapper around existing models as a method for including class. It allows for flexibility in specifying the functional form of intersectionality (ex: interaction terms or multilevel models), while also providing a mechanism for determining class membership using both SSS and SES measures. These models are a form of clustering analysis, generating probabilities of instances belonging to each group. The use of mixture models in order to semiparemetrically uncover distributional shape has been used in a multitude of fields within the social sciences, as well as biological and physical sciences (McLachlan & Basford, 1988; McLachlan et al., 2019). Conveniently, if we assume that the membership is dependent on covariates, we can also uncover an equation that represents the probability of belonging to one group or the other. This enables the model to serve the dual purpose of generating estimates for prediction as well as estimates for class membership. The latter is important for understanding which elements of class are salient in various outcomes—or rather represents which survey questions can appropriately be used as proxies for class for specific outcomes.

Intuitively, this approach can be connected to the fuzzy logic path of intersectionality suggested by Hancock, 2007b. In it, Hancock claims that an appropriate way to incorporate the in-group heterogeneity of individuals resulting from the salience of identities is to view identities as percentages rather than binary. This is because the identities of individuals may affect them differently due to a multitude of contextual factors—for instance, the strength of racial identity can be changed due to the racial composition of the neighborhood. By applying a class-based mixture model, we produce a probability of being in each class for each individual. The implicit assumption of mixture models is that each point belongs to a discrete group with the probability. However, the explicit maximization problem assumes that each point is a linear combination of the two groups, with weightings representing the grouping variables. Thus, by generating estimates for individuals using the combinations of the groups represented by the probabilities, we generate a class scale.

This work is situated within the intersectional research paradigm, which posits race, gender, and class are interlocked concepts and their interrelated nature should be prioritized in the research design phase (Hancock, 2007b). We believe it is important to note that this methodology is not meant to be used to test the existence of intersectionality within separate research projects. This paper is written with the understanding that intersectionality is viewed as a paradigm or lens that informs a way of thinking about social science problems (Hancock, 2007b; Simien, 2007). That is, regardless of the size of the effects of the intersection of race, gender, and class on outcomes, their inclusion is necessary in order for models to be consistent with the lived experiences of individuals. The inclusion of class is a step toward moving models further toward reality. While using mixture models enables us to uncover elements of class that are more salient to specific outcomes, the existence of intersectionality and the importance of all three aspects are not questioned.

To illustrate the empirical applications of our approach, we apply it to two outcomes from the American National Election Study survey from 2020—thermometer scores on the U.S. Immigration and Customs Enforcement agency and on the Black Lives Matter movement. These two outcomes were chosen as immigration attitudes are said to vary by class (Berg, 2010; McDermott et al., 2019). We additionally include Black Lives Matter attitudes as they are racially stratified and often have gender differences, therefore it is a reasonable area to introduce a class-based investigation (Azevedo et al., 2022). We find that, as expected, class memberships are correlated between outcomes, but not identical. For the two outcomes, we have that the salient version of class is slightly different. This approach additionally lets us understand the factors that lead to thermometer scores for each of these topics better.

In this paper, we begin with a review of current methods of quantifying class and their limitations. We then introduce mixture models and how they have the potential to help scholars incorporate a more complete understanding of class into their work. We demonstrate the strengths of this approach through a thorough study of how individuals view the U.S. Immigration and Customs Enforcement agency and the Black Lives Matter movement. We conclude that class is, in fact, dependent on the situation, supporting the hypothesis motivating our approach.

4.2 Approaches to Class as an Identity and Predictor

We understand class as a higher-order construct representing an individual or group's relative position in an economic-social-cultural hierarchy. This reflects both socioeconomic status and subjective social status, or in other words, accounts for both the material circumstances of a person as well as their group contexts and social status based on networks and cultural norms. We take this approach to provide a holistic view of class that can then be combined with racial and gender identities to garner a full look at structural inequality and lived experience (Crenshaw, 1989; Yuval-Davis, 2015). Conventional wisdom has measures of SES represented by a combination of factors such as education, occupation, and family income levels (Diemer et al., 2013; Leighley & Nagler, 1992). We understand SSS as defined by Jackman and Jackman, 1973 with four different categories for respondents to self-ascribe; to low-income, working-class, middle-class, upper-class. Much work has been done to differentiate class and social status, our aim in this paper is not to enter into this conversation. Rather, our work is meant to be a broad, contemporary quantitative approach to operationalizing the spectrum of economic experiences and identities; therefore, approaching these concepts under the umbrella of class allows us to achieve this goal with more theoretical parsimony. The flexibility of our model allows us to combine these measures, and additionally accommodate other measures that we believe to be related to class such as student loans, employment status, union affiliation, and having money in the stock market.

Within political science, many have argued for the use of income as the best measure of SES in terms of predicting voting (Leighley & Nagler, 1992). They argue when class is measured in terms of something like education, it seems as though the higher the class the more likely an individual is to vote. This work came on the heels of work such as Bennett, 1991; Burnham, 1982, 1987 which showed a decrease in turnout from lower-class individuals. They argued if this finding were correct, it would lead to a further elite class bias in public policy. Leighley and Nagler, 1992 show instead that this finding is driven by the operationalization of class rather than class itself. When measuring SES with income rather than education, turnout by class appeared stable. In this paper, we leverage the lessons learned from this piece—the influence of class is susceptible to variance based on the way it is operationalized and care needs to be taken to properly articulate its contours.

Currently, the most common approach of other scholars is to use measures of occupational prestige, or a composite measurement of these SES categories along with occupation to articulate class (Goldthorpe & McKnight, 2006). These measures take into account not only the occupation but also the benefits associated with the job like income security, earnings stability, and long-term prospects. The groups in the schema are as follows: higher and lower professional and managerial classes, the "routine nonmanual class" (typically lower-grade clerical "white-collar workers," the "petty bourgeoisie" (small employers and self-employed), and the "working class" (foremen and technicians, skilled, semi-, and unskilled manual workers) (Evans & Opacic, 2022). Due to the nature of the datasets we are using, we do not use the Goldthorpe schema in this paper, however, we are confident that it can be accommodated by our method of class-based mixture modeling.

Class consciousness and class as a social identity emerged with the work of scholars like Karl Marx who based his theories on where people were situated in the means of production (Evans & Opacic, 2022). This literature is related, but distinct from measuring social status according to Weber (Weber, 1968). According to Weber, social status is based on social hierarchies and cultural perceptions, while class is based on objective material realities (Chan & Goldthorpe, 2007). Our work is situated with the behavioralist researchers that operationalize class through SSS survey measures as a means to predict political outcomes. Jackman and Jackman, 1973, 1983 showed that class identity was a combination of the material as well as social patterns of contact that change the relationship between objective measures of SES and subjective measures and that these relationships predict relevant political outcomes. Jackman and Jackman, 1973 showed that the boundaries of these class identities and patterns of contact then led to distinct out-group views and the development of class-based identity using the work of (Tajfel, 1969). Our work recognizes the importance of the origins of Marx and Weber but broadly situates class as in part a social identity along the lines of Jackman and Jackman, 1973, 1983 where there is a psychological attachment to the self-identified group.

The work by Jackman and Jackman, 1973, 1983 led to a boom in research on the link between subjective social status and socioeconomic status. Scholars such as Evans and Kelley, 2004; Sosnaud et al., 2013 show that the vast majority of people identify as being in the middle class despite their material realities showing otherwise. In the American context, Sosnaud et al., 2013 specifically showed that this divergence in subjective and objective class varied by race and education. This research has shown a distortion between subjective measures and objective measures that could be rooted in desires to distance from segments of society such as the upper or lower class which have certain cultural connotations (Bourdieu, 1987; Lamont, 2002; Stuber, 2006). Our modeling approach recognizes both the strength and weaknesses of SSS by accommodating the measure along with the material context.

Across the disputes on how to operationalize class from Marx, Weber, and beyond, intersectionality provides a method for articulating these stratifications in social positions that recognizes that class does not exist in a vacuum away from other structural oppressions (Yuval-Davis, 2015). The power of intersectionality, broadly, is that it does not limit understandings of stratification to one axis of difference like class on its own, but incorporates power differentials along race and gender as multiply constituted identities (Crenshaw, 1989; Hancock, 2007b; McCall, 2005; Yuval-Davis, 2015). Lived experience in this context is thus a combination of material and cultural economic realities, racial and ethnic dynamics, and gender structures. Yuval-Davis, 2015 argues that situated intersectionality specifically provides a comprehensive manner to studying social inequalities and class in a way that is ignored if one takes the traditional approach of Weber or Marx. We specifically define class as both a combination of SSS and SES in mixture modeling, then use an intersectional research paradigm approach to including race and gender as well (Hancock, 2007b, 2019; Much, 2022). This methodology allows the researcher to incorporate race, gender, and class quantitatively.

4.3 Class as a Latent Variable

Historical work has emphasized the importance of care in the way class is operationalized. While there is little doubt that class membership holds importance in a multitude of areas, findings are often difficult to accept when they can be nullified by an equally reasonable definition of class. Thus, in this paper, we recommend viewing class as a latent variable. In addition, rather than locking down a single definition of class, we argue that it is situationally dependent. An individual's class identity is dependent on a variety of factors, and in different contexts, different aspects of that identity may become more salient.

There exist many methods for uncovering latent classes, some of the most popular unsupervised methods include item response theory models (Lord, 1980; Osteen, 2010) , k-means (MacQueen, 1967), k-modes (Huang, 1997), principal component analysis (Pearson, 1901) and factor analysis. In the case of class, model-based approaches take advantage of the structure of responses in a way that other, fully unsupervised methods, fail to. Individuals of different classes have different opinions, habits, and views on life, not just different characteristics. This variety in outcomes, not just observable characteristics, leads us to choose a model-based approach rather than other latent variable models such as item response models or clustering which are purely unsupervised to uncover class membership. In jointly estimating the latent class with the response variable, the coherence of the response is taken in addition to the clustering of the class-related covariates. Specifically, we are not interested simply in the clusters of class-related responses, but how they interact with the response variable in question.

Furthermore, among model-based approaches, we choose mixture models due to the "soft" classification it provides. That is, for each respondent, the mixture model provides a probability of being in each group. Rather than assigning each respondent to a class, we generate a class spectrum. This conceptualization is consistent with work done within the intersectionality field which suggests the most appropriate characterization for identity is continuous rather than binary (Hancock, 2007a).

In our approach, we assume that there are two class-based groups for ease of understanding and parsimony. Additionally, since we are not assigning discrete class assignments, but probability measures, two classes still allows for a great deal of variability. However, this assumption is without loss of generality and the number of classes can easily be expanded within the model to large groups if needed. A large contribution of this approach is its flexibility and responsiveness to the needs of the researcher. We claim that by using a mixture model, dependent on variables related to class, we achieve the dual goal of including class in addition to uncovering, for specific outcomes, how various factors contribute to the understanding of class. Given any models for outcomes, a mixture model wrapper can be included to integrate class in the analysis.

Mixture Model Description

For respondents $i \in \mathcal{I}$, take y_i to be the outcome of interest. We claim there is an underlying latent class variable denoted as $\kappa_i \in \{1, 2\}$. These correspond to class memberships which are not explicitly observable.

Assuming κ_i, the expectation of the response variable can be expressed as:

$$\mathbb{E}[y_i] = \sum_{j \in \mathcal{J}} \pi_j f_{j,i} \tag{4.1}$$

where $\pi_j = Pr(\kappa_i = j)$ is the probability that individual i belongs to class j and $f_{j,i}$ is the expected outcome for individual i assuming they belong to class j.

In order to extricate respondent class and generate appropriate estimations of the outcome, we split the variables into two groups: one which determines class membership (class-determining variables) and the other which determines outcomes dependent on class (outcome-determining variables). While both groups of variables contribute to the overall outcome, the class-determining variables are not found in the outcome regression. The covariates for respondent i are thus written as (c_i, x_i) where $c_i = (1, c_{i,1}, c_{i,2}, ...)$ are the class-determining variables and $x_i = (1, x_{i,1}, x_{i,2}, ...)$ are the outcome-determining variables. Equation 4.1 can thus be re-written with our definitions as:

$$\mathbb{E}[y_i|c_i, x_i] = \sum_{j \in \mathcal{J}} \pi_j(\alpha c_i) f_{i,j}(\beta_j x_i). \tag{4.2}$$

Without loss of generality, we assume that the class function is defined by a logit link function. This can be replaced by any general linear model that maps the covariate space to the $[0, 1]$ interval. For the logit link function, we have that:

$$\log\left(\frac{\pi_1(c_i)}{1 - \pi_1(c_i)}\right) = \alpha c_i. \tag{4.3}$$

The functional form of the class-group outcome is exogenous to the model description. The flexibility of this approach allows users to choose the class-group model they feel best represents their outcome. The outcome equations can have any form the researcher desires, so long as the equation

$$\mathbb{E}[y_i|c_i, x_i, \kappa_i = j] = f_{i,j}(\beta_j x_i) \tag{4.4}$$

holds. There is no restriction on the relation between $f_{i,1}$ and $f_{i,2}$, they can be different or the same, although for the case of simplicity in our examples we will have them maintain the same functional form as each other.

Empirical approach

In this section we discuss methods for fitting mixture models as well as tests to ensure robustness of solutions. This is meant to aid in the use of the method for future research. Due to the nature of mixture models, we suggest an empirical Bayes approach to fitting the model. Empirical Bayes is a method in which results from similar models are used in order to aid in the solving of a more complicated model.

When solving mixture models using Monte Carlo methods, the results are more robust and solutions converge faster if weakly informative priors are supplied. It can safely be assumed that class is a clustering of various economic and social dimensions (Diemer et al., 2013). Thus, in order to provide the mixture model with an appropriate prior we recommend using a completely unsupervised clustering method on the relevant covariates as an initial guess. Specifically, we suggest k-modes, an extension of k-means which extends its use to categorical data (Huang, 1997; MacQueen, 1967). With the specification that $k = 2$, this assigns a group membership, g_i, of 1 or 2 to each respondent.

Once initial clusters have been uncovered, we fit a logisitic regression to the class-membership variables with the unsupervised membership values as the outcome. This takes the form:

$$Pr(g_i|c_i) = \frac{1}{1 + e^{-\tilde{\alpha} c_i}}. \tag{4.5}$$

The estimates $\tilde{\alpha}$, with their standard errors, serve as the priors for the class membership. The next step is to estimate probabilities of class membership given the regression results. At this point, the respondents can be split into two groups based on estimated probabilities, for instance those for whom the probability of being in a group is greater than 50%.

With these new groups, a simple regression can be run of the form:

$$y_i = f_j(\tilde{\beta}_j x_i). \tag{4.6}$$

The estimates for $\tilde{\beta}_j$ can then be used as the priors for the group-level coefficients β_j. With these priors in hand, the researcher is prepared to solve for the full model using Markov Chain Monte Carlo techniques.

Generally, we recommend including the exact estimates (if not shrinking the priors) in order to choose relevant parameters for solving such as target average acceptance probability (adapt_delta in stan), max binary tree size for the NUTS algorithm (max_treedepth in stan), warmup iterations and sampling iterations. Once those solving parameters have been chosen, decrease the specificity of the priors by increasing the variance. As this goes, you may need to alter the solving parameters as well. For all the priors, we recommend including the estimate found using the above methods but at least doubling if not tripling the variance for each estimating. This decision is to decrease reliance on the prior and enable further movement when the algorithm searches the space.

Empirical Checks

Once the priors have been set, the full model can be estimated. After estimation, there are a few key checks necessary to ensure goodness of fit. The first set is to ensure that the prior results are not too heavily influencing the final results. While the priors are useful for giving the model direction, a key point of the mixture model approach is that the dependent variable helps determine the way that class is defined. Second, like with any model, it is important to check that the fit is appropriate.

The first checks are to ensure that there is an amount of certainty in the class estimates. This is done in a few ways. First, confirm that the distributions of the estimated class have high densities at 0 and 1. If this is not the case, there can be issues from having essentially an empty set. Second, since we are confident that measures of the class should be correlated, confirm that there is a positive correlation between class as found using the unsupervised method and class as estimated after the mixture model has been fit. Finally, it is important that the coefficients of the class-determining variables are not identical between the mixture model and the fit from the unsupervised model (which provided the priors). If these are identical, that is evidence that the outcome did not effect the class definitions and we are simply using the unsupervised model.

Classic ways to confirm the appropriateness of a model include root mean square error, expected predictive accuracy, and visual confirmations. Given many of the outcomes studied in relation to class—for instance, the propensity to vote, views on subjects, etc.—are noisy, we do not expect there to be a significant improvement in fit from this method. The major benefit of this model is that it is theoretically consistent and provides insight into the effects as well as a definition of class in different contexts. However, it is important that the mixture model does not perform worse than a non-mixture version of the same model.

4.4 Empirical Results

For our empirical confirmation of this methodology, we use the American National Election Studies Survey (ANES) from 2020. As our outcome variables, we look at the thermometer scores for the U.S. Immigration and Customs Enforcement (ICE) Agency as well as Black Lives Matter (BLM). In each of these questions, respondents are asked how they would rate each group from 0 to 100. We choose to look at two different outcomes within the same survey to display how the latent class variable is context specific. These questions were chosen since they were thought to have a higher probability of clear class-related variation.

For our model, we use the two thermometer ratings as the outcome variables. For the class-determining variables, we set c = (college, family income, remaining student loans, employment status, union affiliation, money in stock market, occupation type). The variables college, remaining student loans, union affiliation, and money in the stock market are binary. Occupation type and employment status are categorical with 9 and 7 groups, respectively (details can be seen in Appendix D.1). In the survey, income is represented as a series of bins, to convert this value to a continuous variable we assign the income value as the low end of the bin the respondent belongs to. In addition, the income values are centered around the mean and represented in the 10's of thousands in order to generate coefficients of reasonable magnitudes.

For the outcome-determining variables, we have x = (7 point party identification, gender, race, age). Age and party identification are represented as continuous variables centered around zero while gender and race are categorical. The full options for each of these variables can be seen in the Appendix D.1.[2] The functional form we use for $f_{i,j}$ is a multilevel model with groupings based on gender and race combinations and random effects for both intercept and age depending on these groupings. There are additionally fixed effects for the intercept, age, and party identification. Party identification is the only outcome-dependent covariate thought to be independent of intersectional group. The ANES for 2020 had a total of 8,280 respondents. Of these, 5,831 of the respondents had valid responses to all of the variables used. This is the subset of data used to test our approach.

Fit Checks

As described in 4.3, we first generate the priors for the class coefficients as well as outcome coefficients before running the full model. Before analyzing the results, we first run through the empirical checks recommended in 4.3. First, we confirm that all versions of the clustering result in groups with respondents belonging to each. This can be seen from the density plots in Figure 4.1. In addition, it is clear from the figure that not only do all estimates break the respondents into two relatively certain groups, but the groupings are not identical. This suggests that there is information learned from the utilization of the response variable in the model. In the correlation plots it is clear that for both mixture models and the original clustering, there is a positive correlation between membership classifications. This is what we are hoping to see as we don't expect class to be completely different in different situations, we simply expect some aspects to be more important.

In order to confirm this assumption, we next compare the coefficients estimates from the original k-modes clustering and the two mixture models. It can be seen in Figure 4.2 that the estimated coefficients are different between the priors and

[2] The ANES labels those who answer positively to "Are you Spanish, Hispanic or Latino?" as Hispanic. We maintain this convention in the chapter but call the reader's attention to the detail.

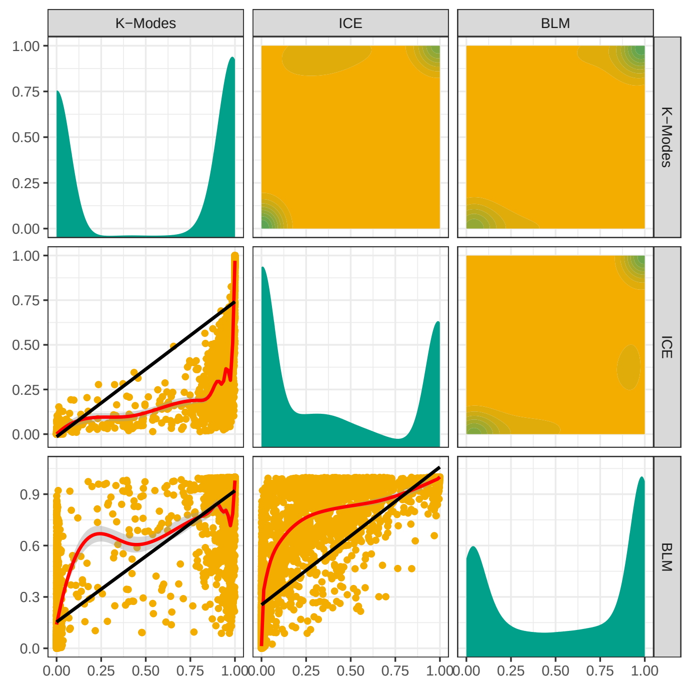


Figure 4.1: Predicted probabilities of class membership for ICE, BLM, and K-modes models. The diagonals show the density of the predicted membership by model. The lower triangle shows the point estimate for the dataset and a generalized linear fit for the correlation (red) as well as a best linear fit (black). The upper triangle shows a heat map of the same information.

the two mixture models. These results show that there was enough strength from the outputs to move the estimates of the class differentiators from their priors to alternative areas. In addition, this was a different addition for the two models.

To analyze the fit of the model, we compare the results to a non-mixture version of the same model. This can be thought of as a mixture model where $\pi_i = 1$ for all i. We call this model the "single class" model as compared to the mixture model. The root mean squared error when ICE is the outcome is 23.98 for the single class model and 23.85 for the mixture model, an improvement but an insignificant one. Similarly, the BLM error is 25.02 and 24.81 for the single and mixture model, respectively. We next look at the expected predictive accuracy of each model using leave-one-out cross-validation. For both outcome variables, the mixture model performs better than the single class model with a difference in expected log pointwise predictive density of 73.7 (standard error 14.1) for the BLM outcome and 39.3 (standard error 10.9) for the ICE outcome. Given these checks, we can feel confident that the model is appropriately estimating both class and the outcomes. We can now move on to analyzing the results.

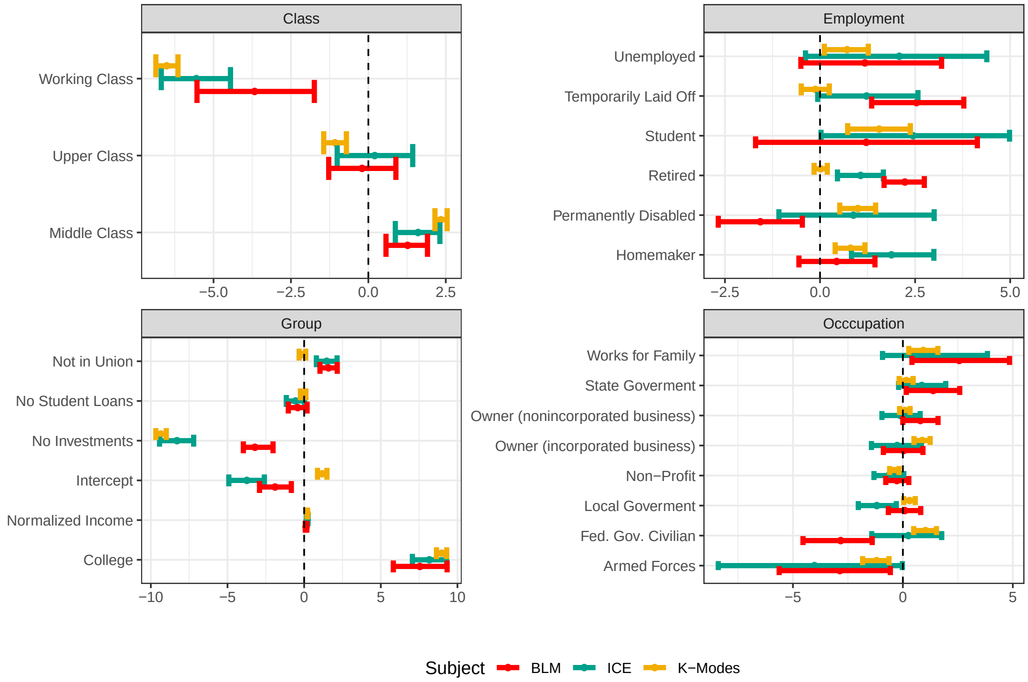


Figure 4.2: Predicted coefficients for class-determining covariates assuming class is split my K-modes or uncovered using a mixture model using the ICE or BLM thermometer as the output.

Outcome Analysis

With the confirmation that the model has been appropriately fit, we now turn to analyze the results. There are two sets of analyses that can be done. The first is how class is understood in each context while the second is how the outputs respond to class. We first look at the definition of class and this will be followed by an analysis of the outcomes.

In order to understand how class is determined in each situation we return to Figure 4.2. With these results, we can see how class differs in the three contexts: naive k-modes, BLM and ICE. For the SSS measures, in which we look at how people self-identify, the mixture model-derived coefficients show them as less differentiating than the k-modes based prior. There is no statistical difference in class membership between those who identify as upper and lower class while the positive and negative effects of middle and working class, respectively, are smaller. There are also differences between the two mixture models. While the impact of being an investor is similar for the k-modes model and ICE mixture mode, the importance is smaller when the model is based on feelings towards BLM. Finally, in some cases all three models have starkly different results, for returned individuals there is no difference between working individuals and retired individuals, but for the ICE mixture model being retired is indicative of more likely having a higher class status and for BLM it is even more likely. This supports our theory that for different outcomes, the salient factors of class are different. That class itself is fluid.

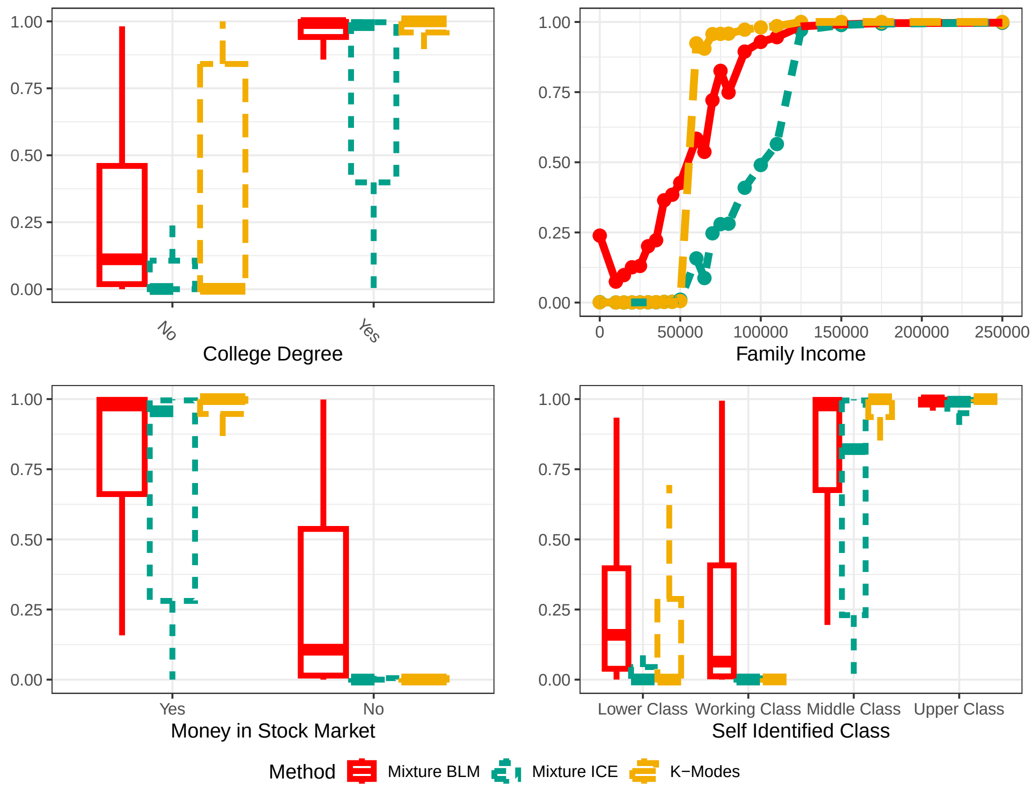


Figure 4.3: The estimated class membership when the data is split by a subset of the class-dependent covariates. 1 refers to a 100% probability of having a higher class status while 0 refers to a 100% probability of belonging to the lower class.

Due to the categorical nature of the covariates, the coefficients themselves are a bit difficult to interpret. As a result, we additionally look at the differentiation between the clusters through their expectation in group membership. We cut the data by each covariate and look at the estimated membership for each clustering method. These results can be seen in Figure 4.3. We see that many of the categories imply a stronger group membership in one direction or the other—for instance, college education implies a higher class and as does higher income levels. However, the strength of these movements is different between models. Self-identified class is significantly less predictive in the BLM model than in the other two, and in both mixture models the income gradient is less stark. In addition, the transition from low to higher class within income happens earlier in the BLM model than in the other two.

We conclude that while there are certain factors that undoubtedly are indicators of class membership—a higher income correlates with higher class, as does a higher level of education—class is topic dependent. With this new understanding of class, we can delve into the conclusions that can be drawn from the sub-regressions.

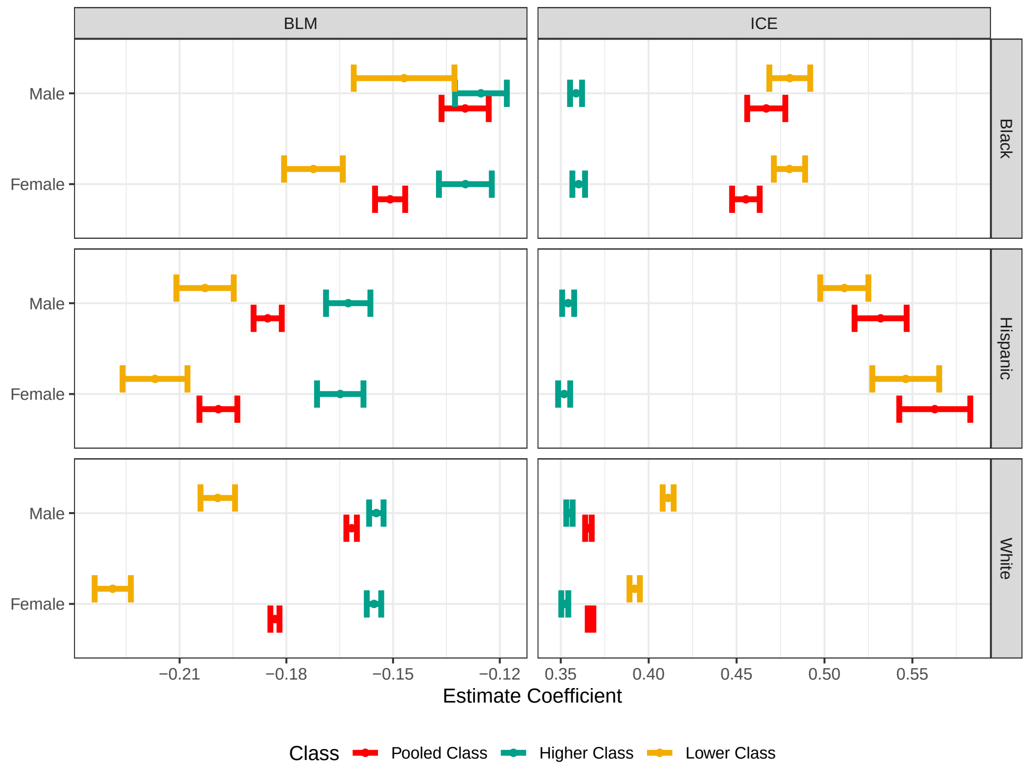


Figure 4.4: Random effects on age for models.

Substantively, our results comport with extant literature showing a relationship between identity and racially charged political attitudes. We also provide additional intersectional nuance. Using feeling thermometers from the 2020 ANES, we show that not taking class into account from an intersectional perspective over and understates effects across intersectional groups. The pooled class in red reflects a multilevel model random effect where the class is not taken into account. The teal and yellow show the higher and lower class grouping effects from the mixture model. In the case of BLM, we see that often the multilevel model on its own pools away class-based variation. For example, as age increases lower-class Black women are less likely to support Black lives matter as opposed to higher-class Black women, and the pooled effect would understate the degree to which age impacts lower-class Black women's sympathy towards BLM. Another stark example is lower-income white women, whose age effect is drastically understated in the pooled class context.

For ICE attitudes, our work shows consistently that the pooled measure of class overstates the degree to which age impacts higher-income individuals across racial and ethnic groups. For lower-income individuals in racial and ethnic minority groups, it is often the case that the pooled class effect for age adequately captures the relationship. However, the mixture model approach uncovers that age matters much less for higher-income individuals (especially Hispanic men and women) and the pooled method understates this phenomenon. Our method shows heterogeneity in the age effect across both BLM and ICE attitudes as shown by the gaps between the mixture model effects and the pooled class effects.

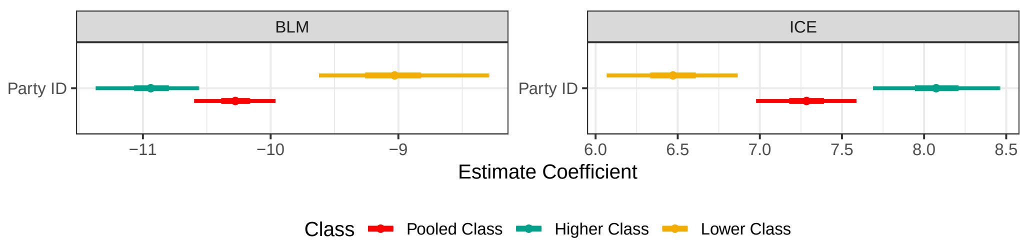


Figure 4.5: Estimated fixed effects.

Our method of measuring class also can uncover substantive relationships surrounding racialized feeling thermometers and political parties. Literature has shown that partisanship is linked to support for Black Lives Matter, with Democrats supporting and Republicans opposing by significant margins (Azevedo et al., 2022). Additionally, party is scaled from –3 to 3 with strong Democrats being –3. For BLM party identification matters less for lower class status individuals, or in other words, partisanship is a smaller driver of BLM or ICE attitudes for lower-class status individuals than it is for those in the higher-class status category. If a researcher were to just specify models using race and gender, the effect of party identification would be understated for higher class status people, and overstated for those with lower class status.

We can additionally show this by looking at the slopes for age on the ICE and BLM feeling thermometers. In Figure 6 we address race, gender, class, party identification, and age's impact on ICE attitudes. Broad trends for ICE attitudes that we see are among independents, it is often the case that the slope for age for the lower

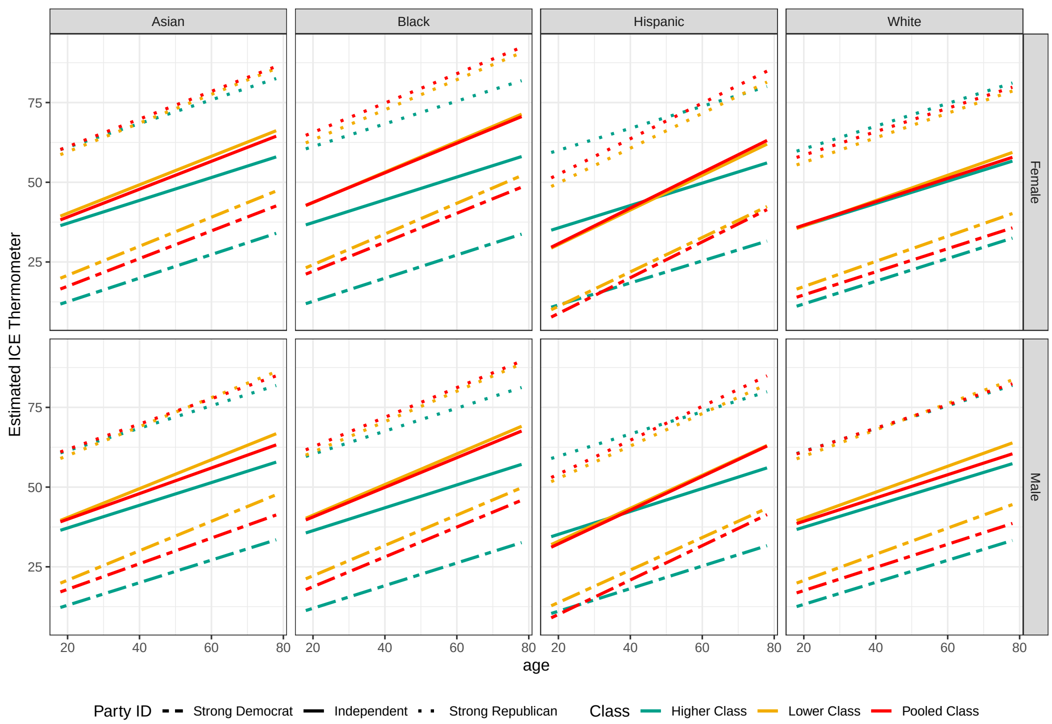


Figure 4.6: Expected ICE thermometer values for strong partisans and independents split by race and gender.

class status mixture model is similar to the pooled class model. However, higher class status independents more often have unique age trends on these racial attitudes. Looking at Hispanic men and women, we see that among strong Democrats increases in age are associated with increases in support for ICE; however, the rate of increase is much less for those with higher class status. This is also true for Black men and women, who have positive slopes for age and ICE support, but when an individual is in the higher class status group the rate at which their support for ICE increases over age is much slower than for the lower income folks. For these groups, the pooled version of class is overstating the degree to which positive attitudes towards ICE increase with age. Other interesting findings include that for White men who are strong Republicans, there is little to no class effect on age and ICE attitudes, as shown by the dotted lines being on top of each other. This means that the pooled method of articulating class garnered the same results as the splits by the mixture model. This result was similar for White women who identify as Independents who do not have class-based differences in age slopes, in other words, the relationship between age and support for ICE for this group is not distinct by class. White men and women strong Democrats both demonstrate class effects with lower class status groups being more sympathetic towards ICE as age increases, and higher class status being less sympathetic as age increases.

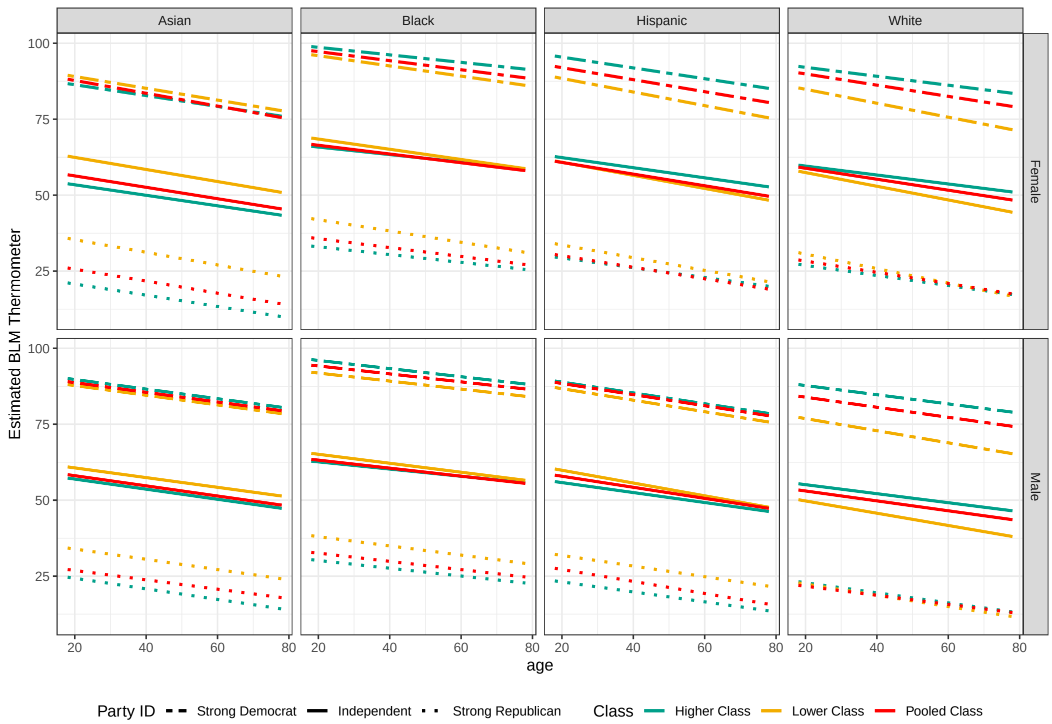


Figure 4.7: Expected BLM thermometer values for strong partisans and independents split by race and gender.

With regards to BLM attitudes, strong Democratic Black men and women both have decreasing support for BLM as age increases, but are overall very supportive. Higher-class status Black Americans are more sympathetic than their lower-class status counterparts (albeit not a huge difference) as evidenced by the pooled class model splitting the difference between the higher and lower status models. Black independent men and women do not demonstrate a clear class effect. Groups that have clear class effects on age and BLM attitudes are strong Democrat White men and women, Hispanic women, and strong Republican Asian women and men. Strongly Democratic white women and men show that higher class status individuals are overall more sympathetic towards BLM across ages as compared to the lower class status group. This again shows that a researcher could underestimate support for BLM for higher class status across ages, and overstate support for lower class status individuals.

4.5 Discussion and Conclusion

Class is widely recognized as a significant factor impacting individuals' political behavior and opinions. However, there has been a lack of consistency in existing research on the best way to approach estimating and utilizing class in quantitative models. Additionally, intersectionality poses that race, gender, and class need to be considered together when studying sociopolitical identities, but often only race and gender are addressed. These issues have led to researchers to present contradictory results on the impact of class in various situations or leave the important impact of class unexplained. In this paper, we recommend treating class as a context-dependent latent variable that can be recovered using mixture models. This mixture model can then be combined with the desired intersectional modeling tactic like multilevel models in our case. We defend this approach from a theoretical standpoint and using empirical evidence.

Methods for measuring class are generally split into two categories—socioeconomic status (SES) which is measured as some combination of material circumstances and subjective social status (SSS) which relies on individuals' self-identification. Researchers make informed decisions on what is the best measure for the outcome they are studying. In this paper, we introduce using the information provided by the outcome variable to help untangle the definition of class. Rather than relying on intuition, which has the potential to propagate biased thinking, our method has the ability to include all available contributors and uses the outcome information to determine the weighting of the aspects. This approach is able to seamlessly weave in measures of both SES and SSS into a single bespoke parameter.

In our two empirical examples, we look at thermometer ratings for ICE and BLM using the 2020 ANES. We find that the definitions of class in each case, while strongly correlated, are non-identical. In the context of ICE not having money in the stock market is a strong indication of belonging to the lower class status group, while this relationship is less strong in the context of BLM. In contrast, not having a college degree is a much stronger indication of being in the lower class status for BLM than for ICE. These results show that class is in fact context-dependent, and also provides insight into which aspects differentiate classes in the case of BLM and ICE.

The substantive results also comport with existing literature on racialized political phenomena. We find that overall individuals with higher class status have less change in their opinions as they age compared to their lower class status counterparts.

However, the opinions of those with higher class status are much more effected by party identification than equivalent individuals with lower class status. For the most part, the results of the pooled class model splits the results of the two class extremes. This means that traditional models are likely to skew the results towards the larger class group displayed in the data.

These preliminary results firmly support our approach to class. In reducing the subjectivity of class we are able to ameliorate the bias introduced through researcher intuition. Additionally, by accepting that class is not a clearly defined concept, we can exploit the myriad of ways its operationalized to come up with a holistic approach. We allow the definition to change with the outcome variables which allows for a whole new realm of study. This technique opens doors to work on how class is experienced by individuals as well as the effects of class on outcomes.

BIBLIOGRAPHY

Ajao, O., Hong, J., & Liu, W. (2015). A survey of location inference techniques on Twitter. *Journal of Information Science*, *41*(6), 855–864.

Alesina, A., & Rosenthal, H. (1995). *Partisan politics, divided government and the economy*. Cambridge University Press.

Algara, C., Hale, I., & Struthers, C. L. (2022). No balance, no problem: Evidence of partisan voting in the 2021 Georgia U.S. Senate runoffs. *American Politics Research*, *50*(4), 443–463.

Altafini, C., & Ceragioli, F. (2018). Signed bounded confidence models for opinion dynamics. *Automatica*, *93*, 114–125.

Alvarez, R. M. (1998). *Information and elections*. University of Michigan Press.

Alvarez, R. M., Nagler, J., & Bowler, S. (2000). Issues, economics, and the dynamics of multiparty elections: The British 1987 general election. *American Political Science Review*, *94*(1), 131–149.

Alvarez, R. M., & Schousen, M. M. (1993). Policy moderation or conflicting expectations?: Testing the intentional models of split-ticket voting. *American Politics Quarterly*, *21*(4), 410–438.

Ansolabehere, S., Snyder Jr, J. M., & Stewart III, C. (2001). Candidate positioning in U.S. House elections. *American Journal of Political Science*, 136–159.

Arora, M., & Stout, C. T. (2019). Letters for Black Lives: Co-ethnic mobilization and support for the Black Lives Matter movement. *Political Research Quarterly*, *72*(2), 389–402.

Azevedo, F., Marques, T., & Micheli, L. (2022). In pursuit of racial equality: Identifying the determinants of support for the Black Lives Matter movement with a systematic review and multiple meta-analyses [Publisher: Cambridge University Press]. *Perspectives on Politics*, *20*(4), 1305–1327. https://doi.org/10.1017/S1537592722001098

Bafumi, J., Erikson, R. S., & Wiezien, C. (2010). Balancing, generic polls and midterm congressional elections. *Journal of Politics*, *72*(3), 705–719.

Bailey, M. A. (2007). Comparable preference estimates across time and institutions for the court, congress, and presidency. *American Journal of Political Science*, *51*(3), 433–448.

Baron, D. P. (1994). A sequential choice theory perspective on legislative organization. *Legislative Studies Quarterly*, 267–296.

Barreto, M. A., Frasure-Yokley, L., Vargas, E. D., & Wong, J. (2018). Best practices in collecting online data with Asian, Black, Latino, and White respondents: Evidence from the 2016 Collaborative Multiracial Post-election Survey [Publisher: Taylor & Francis]. *Politics, Groups, and Identities*, *5503*, 1–10. https://doi.org/10.1080/21565503.2017.1419433

Bartels, L. M. (1996). Uninformed votes: Information effects in presidential elections. *American journal of political science*, 194–230.

Bartels, L. M. (2016). Unequal democracy: The political economy of the new gilded age–second edition. In *Unequal Democracy*. Princeton University Press. https://doi.org/10.1515/9781400883363

Bennett, S. (1991). Left behind: Exploring declining turnout among non-college young whites. Retrieved April 22, 2023, from https://uci.primo.exlibrisgroup.com

Berg, J. A. (2010). Race, class, gender, and social space: Using an intersectional approach to study immigration attitudes [Publisher: Midwest Sociological Society, Wiley]. *The Sociological Quarterly*, *51*(2), 278–302. Retrieved April 22, 2023, from https://www.jstor.org/stable/20697942

Bhattacharyya, A., Braverman, M., Chazelle, B., & Nguyen, H. L. (2013). On the convergence of the Hegselmann-Krause system. *ITCS '13: Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, 61–66.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Blumer, H. (1948). Public opinion and public opinion polling. *American Sociological Review*, *13*(5), 542–549. Retrieved November 17, 2022, from http://www.jstor.org/stable/2087146

Bonica, A. (2014). Mapping the ideological marketplace. *American Journal of Political Science*, *58*(2), 367–386.

Bourdieu, P. (1987). *Distinction: A social critique of the judgement of taste*. Harvard university press.

Boxell, L., & Steinert-Threlkeld, Z. (2022). Taxing dissent: The impact of a social media tax in Uganda. *World Development*, *158*(105950). https://doi.org/https://dx.doi.org/ssrn.3766440.2139/ssrn.3766440

boyd, d., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. *2010 43rd Hawaii International Conference on System Sciences*, 1–10.

Brock, A. (2012). From the Blackhand side: Twitter as a cultural conversation. *Journal of Broadcasting & Electronic Media*, *56*(4), 529–49.

Brooks, H. Z., & Porter, M. A. (2020). A model for the influence of media on the ideology of content in online social networks. *Phys. Rev. Research*, *2*, 023041. https://doi.org/10.1103/PhysRevResearch.2.023041

Buchanan, L., Bui, Q., & Patel, J. K. (2020). Black Lives Matter may be the largest movement in U.S. history. *The New York Times*. https://www.nytimes.com/interactive/2020/07/03/us/george-floyd-protests-crowd-size.html

Bueno De Mesquita, E. (2010). Regime change and revolutionary entrepreneurs. *American Political Science Review*, *104*(3), 446–466.

Bunyasi, T. L., & Smith, C. W. (2019). *Stay woke: A people's guide to making all Black Lives Matter*. NYU Press.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Burnham, W. D. (1982). *The current crisis of american politics*. Oxford University Press.

Burnham, W. D. (1987). The turnout problem. *Elections American Style*, *97*(98), 109.

Burns, A., & Eltham, B. (2009). Twitter free Iran: An evaluation of Twitter's role in public diplomacy and information operations in Iran's 2009 election crisis. *Communications Policy and Research Forum*.

Campbell, A. (1966). Surge and decline: A study of electoral change. In A. Campbell, P. E. Converse, W. E. Miller, & D. E. Stokes (Eds.), *Elections and the political order* (pp. 40–62). Wiley.

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). *The american voter: Unabridged edition*. University of Chicago Press.

Chan, T. W., & Goldthorpe, J. H. (2007). Class and status: The conceptual distinction and its empirical relevance. *American Sociological Review*, *72*(4), 512–532. https://doi.org/10.1177/000312240707200402

Chang, K.-C., Hobbs, W. R., Roberts, M. E., & Steinert-Threlkeld, Z. C. (2022). COVID-19 increased censorship circumvention and access to sensitive topics in China. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(4). https://doi.org/10.1073/pnas.2102818119

Chenoweth, E., Hamilton, B. H., Lee, H., Papageorge, N. W., Roll, S. P., & Zahn, M. V. (2022). Who protests, what do they protest, and why?

Chenoweth, E., & Pressman, J. (2017). Crowd Counting Consortium. Retrieved April 5, 2017, from https://sites.google.com/view/crowdcountingconsortium/home

Chong, D. (1991). *Collective action and the civil rights movement*. University of Chicago Press.

Chong, D., Rogers, R., & Tillery, A. B. (2004). Reviving group consciousness. *The Politics of Democratic Inclusion*, 45–74.

Clark-Parsons, R. (2022). *Networked feminism: How digital media makers transformed gender justice movements*. University of California Press.

Clifford, P., & Sudbury, A. (1973). A model for spatial conflict. *Biometrika*, *60*(3), 581–588. https://doi.org/10.1093/biomet/60.3.581

Clinton, J., Jackman, S., & Rivers, D. (2004). The statistical analysis of roll call data. *American Political Science Review*, *98*(2), 355–370.

Collins, P. H. (1999). *Black feminist thought: Knowledge conciousness and the politics of empowerment* (Revised 10th Anniv 2nd Edition). Routledge. hanco

Conover, P. J. (1988). The role of social groups in political thinking. *British Journal of Political Science*, *18*(1), 51–76.

Cox, J. M. (2017). The source of a movement: Making the case for social media as an informational source using Black Lives Matter. *Ethnic and Racial Studies*, *40*(11), 1847–54.

Crenshaw, K. W. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine. *Chicago Legal*, 139.

Crenshaw, K. W. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, *43*, 60.

Crowder, C. (2020). Following radical and mainstream African-American interest groups on social media: An intersectional analysis of Black organizational activism on Twitter. *National Review of Black Politics*, *1*(4), 474–95.

Crowder, C. (2021). When #BlackLivesMatter at the women's march: A study of the emotional influence of racial appeals on instagram. *Politics, Groups, and Identities*, 1–19.

Dawson Michael, C. (1994). *Behind the mule: Race and class in African-American politics*. Princeton, NJ: Princeton University Press.

Dhamoon, R. K. (2011). Considerations on mainstreaming intersectionality. *Political Research Quarterly*, *64*(1), 230–243. https://doi.org/10.1177/1065912910379227

Diemer, M. A., Mistry, R. S., Wadsworth, M. E., López, I., & Reimers, F. (2013). Best practices in conceptualizing and measuring social class in psychological research: Social class measurement. *Analyses of Social Issues and Public Policy*, *13*(1), 77–113. https://doi.org/10.1111/asap.12001

DiMaggio, P., Evans, J., & Bryson, B. (1996). Have American's social attitudes become more polarized? *American Journal of Sociology*, *102*(3), 690–755.

Dittmer, J. C. (2001). Consensus formation under bounded confidence. *Nonlinear Analysis: Theory, Methods & Applications*, *47*(7), 4615–4621.

Downs, A., et al. (1957). *An economic theory of democracy*. Harper & Row New York.

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of Political Economy*, *65*(2), 135–150. Retrieved February 7, 2023, from http://www.jstor.org/stable/1827369

Driscoll, J., & Steinert-Threlkeld, Z. C. (2020). Social media and Russian territorial irredentism: Some facts and a conjecture. *Post-Soviet Affairs*, *36*(2), 101–121.

Duggan, M., & Smith, A. (2013). *Social media update 2013* (tech. rep.). Pew Research Center. https://www.pewresearch.org/internet/2013/12/30/social-media-update-2013/

Dunivin, Z. O., Yan, H. Y., Ince, J., & Rojas, F. (2022). Black Lives Matter protests shift public discourse. *Proceedings of the National Academy of Sciences*, *119*(10), e2117320119.

Edsall, T. B. (2022). The left-right divide might help Democrats avoid a total wipeout. *New York Times*, *https://www.nytimes.com/2022/10/26/opinion/democrats-midterms-polarization.html*.

Evans, G., & Opacic, A. (2022). How social class influences political choices. In *The Cambridge Handbook of Political Psychology* (1st ed., pp. 382–398). Cambridge University Press. https://doi.org/10.1017/9781108779104.027

Evans, M. D. R., & Kelley, J. (2004). Subjective social location: Data from 21 nations. *International Journal of Public Opinion Research*, *16*(1), 3–38. https://doi.org/10.1093/ijpor/16.1.3

Fiorina, M. P. (1992). An era of divided government. *Political Science Quarterly*, *107*(3), 387–410. Retrieved February 7, 2023, from http://www.jstor.org/stable/2152437

Fiorina, M. P. (2003). *Divided government*. Longman.

Fireman, B., & Gamson, W. A. (1977). Utilitarian logic in the resource mobilization perspective.

Fisher, D. R. (2020). The diversity of the recent Black Lives Matter protests is a good sign for racial equity. *Brookings*.

Fortunato, S., Latora, V., Pluchino, A., & Rapisarda, A. (2005). Vector opinion dynamics in a bounded confidence consensus model. *International Journal of Modern Physics C*, *16*(10), 1535–1551. https://doi.org/10.1142/S0129183105008126

Frasure-Yokley, L. (2018). Choosing the velvet glove: Women voters, ambivalent sexism, and vote choice in 2016. *Journal of Race, Ethnicity, and Politics*, *3*(1), 3–25.

Freelon, D., McIlwain, C., & Clark, M. (2018). Quantifying the power and consequences of social media protest. *New Media & Society*, *20*(3), 990–1011.

Freelon, D., McIlwain, C. D., & Clark, M. D. (2016). Beyond the hashtags: #Ferguson, #BlackLivesMatter, and the online struggle for offline justice. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2747066

Gause, L. (2022). Costly protest and minority representation in the United States. *PS: Political Science & Politics*, *55*(2), 279–281.

Gerber, A. S., Green, D. P., & Larimer, C. W. (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review*, *102*(1), 33–48.

Giorgi, S., Guntuku, S. C., Himelein-Wachowiak, M., Kwarteng, A., Hwang, S., Rahman, M., & Curtis, B. (2022). Twitter corpus of the #BlackLivesMatter movement and counter protests: 2013 to 2021. *Proceedings of the International AAAI Conference on Web and Social Media*, *16*, 1228–35.

Goldthorpe, J. H., & McKnight, A. (2006). The economic basis of social class. *Mobility and inequality: Frontiers of research in sociology and economics*, 109–136.

Grabowski, A., & Kosiński, R. (2006). Ising-based model of opinion formation in a complex network of interpersonal interactions. *Physica A: Statistical Mechanics and its Applications*, *361*(2), 651–664.

Habermas, J. (1991). The public sphere. In C. Mukerji & M. Schudson (Eds.). The University of California Press.

Hancock, A.-M. (2007a). Intersectionality as a normative and empirical paradigm. *Politics & Gender*, *3*(2), p.248. https://doi.org/10.1017/S1743923X07000062

Hancock, A.-M. (2007b). When multiplication doesn't equal quick addition: Examining intersectionality as a research paradigm. *Perspectives on Politics*, *5*(1), 63–79. https://doi.org/10.1017/S1537592707070065

Hancock, A.-M. (2019). Empirical intersectionality: A tale of two approaches. In *The Palgrave Handbook of Intersectionality in Public Policy* (pp. 95–132). Palgrave Macmillan.

Hanretty, C., Lauderdale, B. E., & Vivyan, N. (2020). A choice-based measure of issue importance in the electorate. *American Journal of Political Science*, *64*(3), 519–535.

Hegselmann, R. (2020). Polarization and radicalization in the bounded confidence model: A computer-aided speculation. *Advances in the sociology of trust and cooperation: Theory, experiment, and field studies*, 197–226.

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, *5*(3).

Hickok, A., Kureh, Y., Brooks, H. Z., Feng, M., & Porter, M. A. (2022). A bounded-confidence model of opinion dynamics on hypergraphs. *SIAM Journal on Applied Dynamical Systems*, *21*(1), 1–32.

Historical inflation rates: 1914-2023: US inflation calculator. (2023). https://www.usinflationcalculator.com/inflation/historical-inflation-rates/

Hitt, M. P., Volden, C., & Wiseman, A. E. (2017). Spatial models of legislative effectiveness. *American Journal of Political Science*, *61*(3), 575–590.

Hsiao, Y. (2021). Evaluating the mobilization effect of online political network structures: A comparison between the Black Lives Matter network and ideal type network configurations. *Social Forces*, *99*(4), 1547–74. https://doi.org/10.1093/sf/soaa064

Huang, Z. (1997). A fast clustering algorithm to cluster very large categorical data sets in data mining. *Dmkd*, *3*(8), 34–39.

Huckfeldt, R. R. (1984). Political loyalties and social class ties: The mechanisms of contextual influence [Publisher: [Midwest Political Science Association, Wiley]]. *American Journal of Political Science*, *28*(2), 399–417. https://doi.org/10.2307/2110879

Ince, J., Rojas, F., & Davis, C. A. (2017). The social media response to Black Lives Matter: How Twitter users interact with Black Lives Matter through hashtag use. *Ethnic and Racial Studies*, *40*(11), 1814–30.

Jackman, M. R., & Jackman, R. W. (1973). An interpretation of the relation between objective and subjective social status. *American Sociological Review*, *38*(5), 569–582. https://doi.org/10.2307/2094408

Jackman, M. R., & Jackman, R. W. (1983). *Class awareness in the united states*. University of California Press.

Jackson, S. J., & Foucault Welles, B. (2016). #Ferguson is everywhere: Initiators in emerging counterpublic networks. *Information, Communication & Society*, *19*(3), 397–418.

Jacobson, G. C. (1987). The marginals never vanished: Incumbency and competition in elections to the U.S. House of Representatives, 1952-82. *American Journal of Political Science*, *31*(1), 126–141. Retrieved February 16, 2023, from http://www.jstor.org/stable/2111327

Jacobson, G. C. (2007). Referendum: The 2006 midterm congressional elections. *Political Science Quarterly*, *122*(1), 1–24.

Jacobson, G. C. (2019). Extreme referendum: Donald Trump and the 2018 midterm elections. *Political Science Quarterly*, *134*(1), 9–38.

Jacobson, G. C. (2022). The 2022 U.S. midterm election: A conventional referendum or somthing different? *University of California, San Diego*, *https://www.dropbox.com/s/95s7hip8bzg5x8n/Jacobson%202022%20Essay.pdf?dl=0*.

Jacobson, G. C., & Carson, J. L. (2019). *The politics of congressional elections* (Ninth). Rowman & Littlefield.

Jasper, J. M. (1997). *The art of moral protest: Culture, biography, and creativity in social movements*. University of Chicago Press.

Johnston, H., & Klandermans, B. (1995). *Social movements and culture: Social movements, protest, and contention, volume 4*. University of Minnesota Press.

Junn, J., & Masuoka, N. (2008). Asian american identity: Shared racial status and political context. *Perspectives on Politics*, *6*(4), 729–740. https://doi.org/10.1017/S1537592708081887

Kan, U., Feng, M., & Porter, M. A. (2021). An adaptive bounded-confidence model of opinion dynamics on networks. https://doi.org/10.48550/ARXIV.2112.05856

Kann, C., Hashash, S., Steinert-Threlkeld, Z., & Alvarez, R. (2023). Gathertweet: A python package for collecting social media data on online events. *Journal of Computer and Communications*, *11*(2), 172–193. https://doi.org/10.4236/jcc.2023.112012

Karami, A., Kadari, R. R., Panati, L., Nooli, S. P., Bheemreddy, H., & Bozorgi, P. (2021). Analysis of geotagging behavior: Do geotagged users represent the Twitter population? *ISPRS International Journal of Geo-Information*, *10*(6), 373.

Kern, H. L., & Hainmueller, J. (2006). Electoral balancing, divided government and 'midterm' loss in German elections. *The Journal of Legislative Studies*, *12*(2), 127–149. https://doi.org/10.1080/13572330600738845

Kiewiet, D. R. (1983). *Macroeconomics and micropolitics: The electoral effects of economic issues*. University of Chicago Press.

Kinder, D. R., & Kiewiet, D. R. (1979). Economic discontent and political behavior: The role of personal grievances and collective economic judgments in congressional voting. *American Journal of Political Science*, *23*(3), 495–527.

Klar, S., & Krupnikov, Y. (2016). *Independent politics*. Cambridge University Press.

Kramer, G. H. (1971). Short-term fluctuations in U.S. voting behavior, 1896–1964. *American Political Science Review*, *65*(1), 131–143.

Krosnick, J. A. (1988). The role of attitude importance in social evaluation: A study of policy preferences, presidential candidate evaluations, and voting behavior. *Journal of personality and social psychology*, *55*(2), 196.

Lacy, D., Niou, E. M., Paolino, P., & Rein, R. A. (2019). Measuring preferences for divided government: Some Americans want divided government and vote to create it. *Political Behavior*, *41*, 79–103.

Lamont, M. (2002). *The dignity of working men: Morality and the boundaries of race, class, and immigration*. Russell Sage Foundation Books at Harvard University Press.

Larson, J. M., Nagler, J., Ronen, J., & Tucker, J. A. (2019). Social networks and protest participation: Evidence from 130 million Twitter users. *American Journal of Political Science*, *63*(3), 690–705.

Leeper, T., & Robison, J. (2018). More important, but for what exactly? The insignificant role of subjective issue importance in vote decisions. *Political Behavior*.

Leighley, J. E., & Nagler, J. (1992). Socioeconomic class bias in turnout, 1964-1988: The voters remain the same. *The American Political Science Review*, *86*(3), 725–736. https://doi.org/10.2307/1964134

Leighley, J. E., & Nagler, J. (2007). Unions, voter turnout, and class bias in the U.S. electorate, 1964–2004 [Publisher: The University of Chicago Press]. *The Journal of Politics*, *69*(2), 430–441. https://doi.org/10.1111/j.1468-2508.2007.00541.x

Lewis, A. R. (2017). *The rights turn in conservative christian politics: How abortion transformed the culture wars*. Cambridge University Press.

Lin, C., & He, Y. (2009). Joint sentiment/topic model for sentiment analysis. *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, 375–84.

Lin, C., He, Y., Everson, R., & Ruger, S. (2011). Weakly supervised joint sentiment-topic detection from text. *IEEE Transactions on Knowledge and Data Engineering*, *24*(6), 1134–45.

Little, A. T. (2016). Communication technology and protest. *The Journal of Politics*, *78*(1), 152–66.

Long, S., & McCarthy, J. (2020). Two in three Americans support racial justice protests. *The Gallup Organization*. https://bit.ly/38kCevL

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.

MacQueen, J. (1967). Classification and analysis of multivariate observations. *5th Berkeley Symp. Math. Statist. Probability*, 281–297.

Madestam, A., Shoag, D., Veuger, S., & Yanagizawa-Drott, D. (2013). Do political protests matter? Evidence from the Tea Party movement. *The Quarterly Journal of Economics*, *128*(4), 1633–85.

Manekin, D., & Mitts, T. (2022). Effective for whom? Ethnic identity and nonviolent resistance. *American Political Science Review*, *116*(1), 161–180. https://doi.org/10.1017/S0003055421000940

Martins, A. C. R. (2008). Continuous opinions and discrete actions in opinion dynamics problems. *International Journal of Modern Physics C*, *19*(04), 617–624.

Mason, L. (2018). *Uncivil agreement: How politics became our identity*. University of Chicago Press.

Matthews, D. R., & Prothro, J. W. (1966). *Negroes and the new southern politics*. Harcourt, Brace & World.

McCall, L. (2005). The complexity of intersectionality. *Signs: Journal of Women in Culture and Society*, *30*(3), 1771–1800. https://doi.org/10.1086/426800

McClain, P. D., Johnson Carew, J. D., Walton Jr., E., & Watts, C. S. (2009). Group membership, group identity, and group consciousness: Measures of racial identity in American politics? *Annual Review of Political Science*, *12*, 471–85.

McDermott, M., Knowles, E. D., & Richeson, J. A. (2019). Class perceptions and attitudes toward immigration and race among working-class whites. *Analyses of Social Issues and Public Policy*, *19*(1), 349–380. https://doi.org/10.1111/asap.12188

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and applications to clustering* (Vol. 38). M. Dekker New York.

McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite mixture models. *Annual review of statistics and its application*, *6*, 355–378.

McNollgast. (1994). Politics and the courts: A positive theory of judicial doctrine and the rule of law. *Southern California Law Review*, *68*, 1631.

Mebane, W. (2000). Coordination, moderation, and institutional balancing in American Presidential and House elections. *American Political Science Review*, *94*(1), 37–58.

Metzger, M. M., Bonneau, R., Nagler, J., & Tucker, J. A. (2016). Tweeting identity? Ukrainian, Russian, and #Euromaidan. *Journal of Comparative Economics*, *44*(1), 16–40.

Miller, A. H., Gurin, P., Gurin, G., & Malanchuk, O. (1981). Group consciousness and political participation. *American Journal of Political Science*, 494–511.

Morales, J. S. (2021). Legislating during war: Conflict and politics in Colombia. *Journal of Public Economics*, *193*, 104325. https://doi.org/10.1016/j.jpubeco.2020.104325

Much, M. (2022). *Bayesian multilevel modeling for the intersections of race and gender* (preprint). Politics and International Relations. https://doi.org/10.33774/apsa-2022-rxv5f

Munger, K. (2016). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, 1–21. https://doi.org/10.1007/s11109-016-9373-5

Olsen, M. E. (1970). Social and political participation of Blacks. *American Sociological Review*, 682–97.

Olson, M. (1965). *The logic of collective action*. Harvard University Press.

Olteanu, A., Weber, I., & Gatica-Perez, D. (2015). Characterizing the demographics behind the #BlackLivesMatter movement. *arXiv preprint arXiv:1512.05671*.

Osteen, P. (2010). An introduction to using multidimensional item response theory to assess latent factor structures. *Journal of the Society for Social Work and Research*, *1*(2), 66–82.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Pearlman, W. (2018). Moral identity and protest cascades in Syria. *British Journal of Political Science*, *48*(4), 877–901.

Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, *2*(11), 559–572.

Petitjean, C., & Talpin, J. (2022). Tweets and doorknocks. Differentiation and cooperation between Black Lives Matter and community organizing. *Perspectives on Politics*, 1–15.

Pfaff, S. (1996). Collective identity and informal groups in revolutionary mobilization: East Germany in 1989. *Social Forces*, *75*(1), 91–117.

Pipal, C., Schoonvelde, M., & Schumacher, G. (2019). JST and RJST in political speech.

Polletta, F., & Jasper, J. M. (2001). Collective identity and social movements. *Annual Review of Sociology*, *27*(1), 283–305.

Poole, K. T., & Rosenthal, H. (1985). A spatial model for legislative roll call analysis. *American journal of political science*, 357–384.

Poole, K. T., & Rosenthal, H. (2000). *Congress: A political-economic history of roll call voting*. Oxford University Press on Demand.

Presidential job approval center. (2023). https://news.gallup.com/interactives/185273/presidential-job-approval-center.aspx

Putnam, L., Chenoweth, E., & Pressman, J. (2020). The Floyd protests are the broadest in U.S. history — and are spreading to white, small-town america. Retrieved June 5, 2020, from https://www.washingtonpost.com/politics/2020/06/06/floyd-protests-are-broadest-us-history-are-spreading-white-small-town-america/

Rahimi, B. (2011). The agonistic social media: Cyberspace in the formation of dissent and consolidation of state power in postelection Iran. *The Communication Review*, *14*(3), 158–78.

Raleigh, C. a. A. L. a. H. H. a. J. K. (2010). Introducing ACLED-armed conflict location and event data. *Journal of Peace Research*, *47*, 651–660.

Ray, R., Brown, M., Fraistat, N., & Summers, E. (2017). Ferguson and the death of Michael Brown on Twitter: #BlackLivesMatter, #TCOT, and the evolution of collective identities. *Ethnic and Racial Studies*, *40*(11), 1797–1813.

Reny, T. T., & Newman, B. J. (2021). The opinion-mobilizing effect of social protest against police violence: Evidence from the 2020 George Floyd protests. *American Political Science Review*, *115*(4), 1499–1507. https://doi.org/10.1017/S0003055421000460

Riker, W. H. (1980). Implications from the disequilibrium of majority rule for the study of institutions. *American Political Science Review*, *74*(2), 432–446.

Sanchez, G. R. (2006). The role of group consciousness in political participation among Latinos in the United States. *American Politics Research*, *34*(4), 427–450.

Sanchez, G. R., & Vargas, E. D. (2016). Taking a closer look at group identity: The link between theory and measurement of group consciousness and linked fate. *Political Research Quarterly*, *69*(1), 160–174.

Shadmehr, M., & Bernhardt, D. (2011). Collective action with uncertain payoffs: Coordination, public signals, and punishment dilemmas. *American Political Science Review*, *105*(4), 829–51.

Shayo, M. (2009). A model of social identity with an application to political economy: Nation, class, and redistribution. *American Political Science Review*, *103*(2), 147–74.

Shepsle, K. A. (1979). Institutional arrangements and equilibrium in multidimensional voting models. *American Journal of Political Science*, 27–59.

Shor, B., Berry, C., & McCarty, N. (2010). A bridge to somewhere: Mapping state and congressional ideology on a cross-institutional common space. *Legislative Studies Quarterly*, *35*(3), 417–448.

Shuman, E., Hasan-Aslih, S., van Zomeren, M., Saguy, T., & Halperin, E. (2022). Protest movements involving limited violence can sometimes be effective: Evidence from the 2020 Black Lives Matter protests. *Proceedings of the National Academy of Sciences*, *119*(14), e2118990119.

Siegel, A. A., & Badaan, V. (2020). #No2Sectarianism: Experimental approaches to reducing sectarian hate speech online. *American Political Science Review*, *114*(3), 837–855. https://doi.org/10.1017/S0003055420000283

Siegel, D. A. (2009). Social networks and collective action. *American Journal of Political Science*, *53*(1), 122–138.

Simien, E. M. (2007). Doing intersectionality research: From conceptual issues to practical examples. *Politics and Gender*, *3*(2), 264–271. https://doi.org/10.1017/S1743923X07000086

Sirbu, A., Pedreschi, D., Giannotti, F., & Kertesz, J. (2019). Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model. *PLOS One*, *14*(3), 1–20.

Sobolev, A., Joo, J., Chen, K., & Steinert-Threlkeld, Z. C. (2020). News and geolocated social media accurately measure protest size variation. *American Political Science Review*, *114*(4), 1343–1351. https://doi.org/10.1017/S0003055420000295

Sosnaud, B., Brady, D., & Frenk, S. (2013). Class in name only: Subjective class identity, objective class position, and vote choice in american presidential elections. *Social Problems*, *60*(1), 81–99. https://doi.org/10.1525/sp.2013.60.1.81

Speier, H. (1950). Historical development of public opinion. *American Journal of Sociology*, *55*(4), 376–388.

Steinert-Threlkeld, Z. C. (2017). Spontaneous collective action: Peripheral mobilization during the Arab spring. *American Political Science Review*, *111*(2), 379–403.

Stokes, A. K. (2003). Latino group consciousness and political participation. *American Politics Research*, *31*(4), 361–378.

Stuber, J. M. (2006). Talk of class: The discursive repertoires of white working- and upper-middle-class college students [Publisher: SAGE Publications Inc]. *Journal of Contemporary Ethnography*, *35*(3), 285–318. https://doi.org/10.1177/0891241605283569

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Biosocial Science*, *1*(S1), 173–191. https://doi.org/10.1017/S0021932000023336

Tate, K. (1994). *From protest to politics: The new Black voters in american elections*. Harvard University Press.

Taylor, S. J., Muchnik, L., Kumar, M., & Aral, S. (2022). Identity effects in social media. *Nature Human Behaviour*, *7*(January), 27–37. https://doi.org/10.1038/s41562-022-01459-8

Teske, N. (1997). *Political activists in America: The identity construction model of political participation*. Cambridge University Press.

Tilly, C. (1977). From mobilization to revolution.

Tong, X., Li, Y., Li, J., Bei, R., & Zhang, L. (2022). What are people talking about in #BlackLivesMatter and #StopAsianHate? Exploring and categorizing Twitter topics emerging in online social movements through the latent dirichlet allocation model. *arXiv e-prints*, arXiv–2205.

Tufte, E. R. (1975). Determinants of the outcomes of midterm congressional elections. *American Political Science Review*, *69*(3), 812–826.

Tullock, G. (1971). Public decisions as public goods. *Journal of Political Economy*, *79*(4), 913–18.

Valenzuela, S. (2013). Unpacking the use of social media for protest behavior: The roles of information, opinion expression, and activism. *American Behavioral Scientist*, *57*(7), 920–42.

Verba, S., & Nie, N. H. (1987). *Participation in America: Political democracy and social equality*. University of Chicago Press.

Wang, D., Piazza, A., & Soule, S. A. (2018). Boundary-spanning in social movements: Antecedents and outcomes. *Annual Review of Sociolgy*, *44*, 167–187.

Weber, M. (1968). *Economy and society: An outline of interpretative sociology*. California University Press.

Weihs, C., Ligges, U., Luebke, K., & Raabe, N. (2005). Klar analyzing German business cycles. In D. Baier, R. Decker, & L. Schmidt-Thieme (Eds.), *Data analysis and decision support* (pp. 335–343). Springer-Verlag.

Weldon, S. L. (2006). The structure of intersectionality: A comparative politics of gender. *Politics & Gender*, *2*(02). https://doi.org/10.1017/S1743923X06231040

Westwood, S. J., Grimmer, J., Tyler, M., & Nall, C. (2022). Current research overstates American support for political violence. *Proceedings of the National Academy of Sciences*, *119*(12), e2116870119.

Williamson, V., Trump, K.-S., & Einstein, K. L. (2018). Black Lives Matter: Evidence that police-caused deaths predict protest activity. *Perspectives on Politics*, *16*(2), 400–15.

Woolley, J. T. (2022). The 2022 midterm elections: What the historical data suggest. *The American Presidency Project, https://www.presidency.ucsb.edu/analyses/the-2022-midterm-elections-what-the-historical-data-suggest*.

Wright, E. O. (Ed.). (2005). *Approaches to class analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511488900

Yuval-Davis, N. (2015). Situated intersectionality and social inequality [Place: Paris Publisher: Presses de Sciences Po]. *Raisons politiques*, *58*(2), 91–100. https://doi.org/10.3917/rai.058.0091

Appendix A

COLLECTIVE IDENTITY IN COLLECTIVE ACTION

A.1 Data Collection

Our data collection process necessitated a number of stages. Full details of the process and the package used can be seen in Kann et al., 2023. The first was deciding which protests to target. First, we choose to narrow the decision to a three target cities—this enabled us to take advantage of clustering and the cities of Los Angeles, Houston, and Chicago were picked due to their large size. From May until September, the Crowd Counting Consortium found 126 protests in Los Angeles, 21 in Houston, and 115 in Chicago. These numbers were untenable with the current data processes. As a result, we choose to look at a few protests immediately after the murder of George Floyd and then about a protest every 2 weeks after this point. Our analysis process makes the exclusion of protests acceptable, since we are only looking at an individuals patterns around the time they protest and not at their rates of protest or other overall protest behaviour. In the rest of this section, we discuss the keyword choices, the works for which we search in order to identify protesters, as well as an overview of the data about the protests.

Keywords

The keywords listed in Table A.1 fall into three different categories: (1) calls to mobilize others to actively join protests, (2) names of the individuals who were victims of injustice, and (3) phrases that are commonly chanted during protests. These three categories are designed to capture the organizational period before the protests, the protest itself, and the topics most likely discussed during the protests. By capturing these aspects of the protests, we are able to collect data on the individuals who we consider to be active protesters.

Keywords
Black Lives Matter
BLM
George Floyd
Justice for Floyd
Walk with Us
Kneel with Us
March with Us
I Can't Breathe
March for Peace
Take a Knee
Breonna Taylor
No Justice No Peace
Say Their Names
Ahmaud Aubrey

Table A.1: Keywords for protesters.

Protest Overview

The tables below provide a list of each City-Date protest pair we investigated. In addition, they provide the estimated size from the Crowd Counting Consortium as well as the number of protesters we managed to label and their average tweeting behaviour throughout the summer. A few main relevant takeaways to note; Los Angeles seems to have a higher Twitter protester to CCC protester ratio, however Los Angeles Twitter protesters seem to tweet less than those in the other cities.

City	Protest Date	CCC Estimated Size	# Protesters Twitter	Avg. # Protester Tweets
Los Angeles	5-27	250	351	1593.29
Los Angeles	5-28	750	604	1582.40
Los Angeles	5-29	2000	516	1285.68
Los Angeles	6-06	3000	325	1473.25
Los Angeles	6-13	200	326	1785.37
Los Angeles	6-27	100	143	1509.47
Los Angeles	7-14	50	79	2386.67
Los Angeles	7-25	150	88	2049.43
Los Angeles	7-26	500	78	2256.35
Los Angeles	8-24	112	112	2173.41
Los Angeles	8-25	200	111	1769.14
Los Angeles	8-26	300	156	2100.17
Los Angeles	9-23	500	196	2033.42

Table A.2: Overview of raw Twitter data: Los Angeles.

City	Protest Date	CCC Estimated Size	# Protesters Twitter	Avg. # Protester Tweets
Houston	5-26	200	2	116.5
Houston	5-29	200	35	793.26
Houston	5-30	200	24	440.58
Houston	6-02	200	180	637.08
Houston	6-08	40	26	700.27
Houston	6-13	50	4	1026.5
Houston	7-04	2000	8	646.75

Table A.3: Overview of raw Twitter data: Houston.

City	Protest Date	CCC Estimated Size	# Protesters Twitter	Avg. # Protester Tweets
Chicago	5-29	300	13	2350.15
Chicago	5-30	1750	35	1209.91
Chicago	5-31	200	44	395.45
Chicago	6-05	2000	48	465.60
Chicago	6-06	25000	69	476.50
Chicago	6-08	200	22	534.60
Chicago	6-12	1400	13	802.46
Chicago	6-13	200	18	238.33
Chicago	6-14	2500	16	520.19
Chicago	6-17	200	12	218.58
Chicago	6-19	2000	28	161
Chicago	6-24	500	7	1323.57
Chicago	6-28	2000	5	211.8
Chicago	7-02	200	9	409.44
Chicago	7-17	1000	9	544.78
Chicago	7-20	100	3	307.67
Chicago	7-24	200	3	113
Chicago	7-25	300	5	1708.6
Chicago	8-08	100	3	266.67
Chicago	8-18	135	3	174
Chicago	8-29	200	5	280.4
Chicago	9-23	24	10	607.3
Chicago	9-24	500	11	4173.82
Chicago	9-26	200	7	5755.29

Table A.4: Overview of raw Twitter data: Chicago.

A.2 Topic Analysis

In the paper the details of the topic analysis are discussed briefly. In this section, the details of the process are outlined. We also provide the results from our validation analyses.

RJST

Topic analysis methods for text data are widely used in social science. The current state-of-the art in unsupervised topic modeling is latent Dirichlet Allocation (LDA) as presented in Blei et al., 2003. More recently, Lin and He, 2009 introduced Joint Sentiment Topic (JST) analysis which, with minimal guidance, is able to estimate the sentiment and then the topic of documents as well as Reverse Joint Sentiment Topic (RJST) analysis to first estimate topics and then sentiment. For our purposes, the ordering of topic and then sentiment makes more intuitive sense. The short length of tweets makes it difficult to cover more than one topic, unlike longer pieces such as movie or film reviews, and it is important for our study to distinguish when individuals are discussing BLM. While there has been significant work (Lin & He, 2009; Lin et al., 2011) which claims that RJST performs worse in sentiment classification, recent work by Pipal et al., 2019 has found that in some cases, RJST can have superior performance. We believe that our data necessitates the use of RJST.

RJST is based on LDA. In LDA each document is modeled as a distribution of topics which is in term a distribution over words. In both JST and RJST an extra, fourth, layer is added - just in a different place. For RJST, we insert a sentiment layer between the topics and the words. Thus, we have that each document is a distribution of topics which is a distribution of topic-specific sentiments which are each a probability distribution of words. It can be factored into three terms:

$$P(w, \ell, z) = P(w|\ell, z)P(\ell, z) = P(w|\ell, z)P(\ell|z)P(z). \tag{A.1}$$

Thus, we end up with the probability of any topic, the probability of a sentiment given a topic, and the probability of a word given a sentiment topic pair. Formally, the generative process can be summarized as:

1. For each document d, choose a topic distribution $\theta_d \sim Dir(\alpha)$
2. For each topic z in document d, choose a sentiment distribution $\pi_{d,\ell} \sim Dir(\gamma)$
3. For each word w_i in document d

- choose a topic $z_i \sim \theta_d$
- choose a sentiment $\ell_i \sim \pi_{d,z_i}$
- choose a word w_i from the multinomial distribution over words defined by ℓ_i and z_i (parameter $\phi_{z_i}^{\ell_i}$ which is the per-corpus joint sentiment-topic word distribution).

The hyperparameter α in this case is the prior for topic distribution. That is, it can be thought of as the prior distribution of topics before having seen any documents. Similarly, γ can be thought of as the prior count of sentiment-topic pairs before any documents are seen.

In order to estimate the model, we use the modified version of Phan's Gibbs LDA++ package written by Lin for R.[1] This is calibrated using the coherence score of the model and searching over the range of topics from 2 to 30 (which correlates with 6 to 90 sentTopic values). The various results can be seen in Figure A.1. For each term, a higher value indicates a better fit and the precise meaning of each term can be found in the documentation for the tex2vec R package.[2] These values lead us to choose a final choice of 5 topics. We left the number of sentiments as three following Lin and He, 2009. The most frequently used words in each senTopic can be seen in Figure A.2, the size represents the number of tweets the word appears in. In Table 2, we list the author-generated label for each topic-sentiment pair. For the remainder of the analysis, we focus on the 5 BLM related senTopics which cover: BLM General, BLM George Floyd/Breaonna Taylor, BLM Civil Rights, BLM Los Angeles News, and BLM Police Violence. This choice is validated in Appendix A.2.

Topic Choices

Figure A.1 shows the different coherence scores for each of the topic numbers chosen. For more information on the different metrics, check out https://rdrr.io/github/dselivanov/text2vec/man/coherence.html. These results were the main driver in our decision to choose 5 topics over a different number of topics.

[1] See http://gibbslda.sourceforge.net/ and https://github.com/linron84/JST/

[2] https://rdrr.io/github/dselivanov/text2vec/man/coherence.html

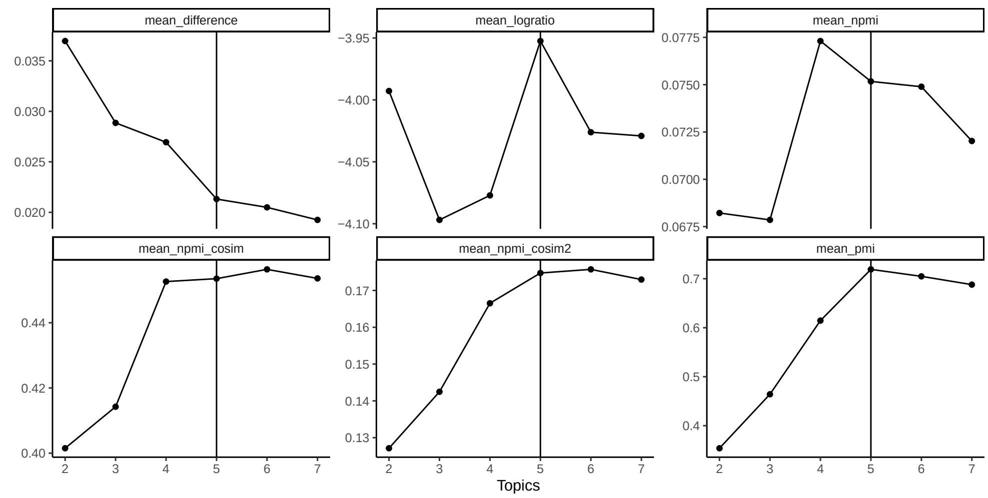


Figure A.1: Coherence metrics for various numbers of topics.

Topic Overview

The word clouds for each topic can be seen in the figure below, this, as well as a detailed analysis of tweets scoring high in each sentiment topic pair are what lead to our author generated labels presented in Table 2.

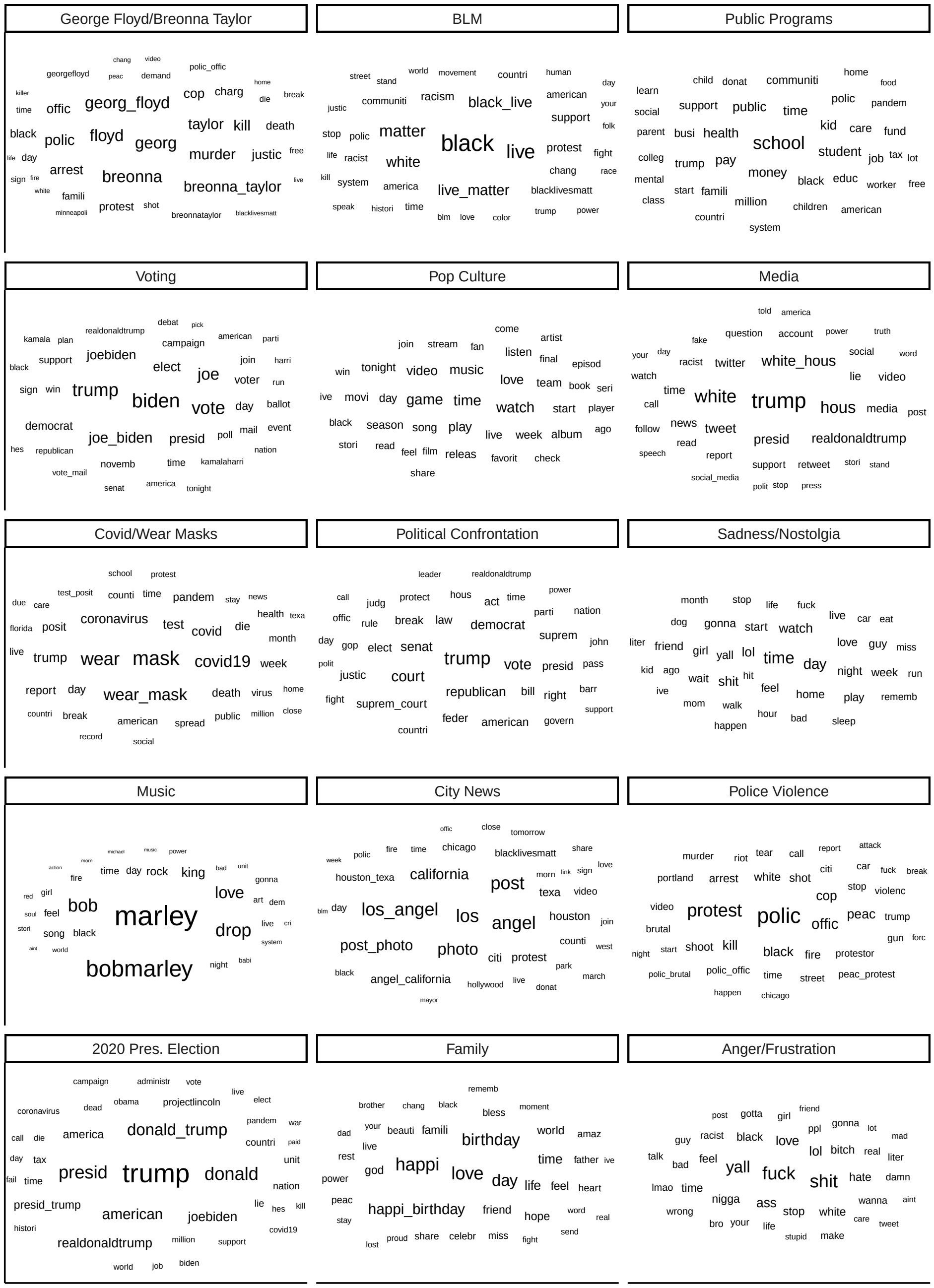


Figure A.2: sentTopic WordClouds.

Topic Validation

If we look at the distribution of all the senTopics individually over time, it is clear that our five BLM ones have the same structure while the others appear random. This can be seen in Figure A.3. Additionally, the patter mimics the Google Trends structure of “BLM” searches over the same time period. This is seen in Figure 1.

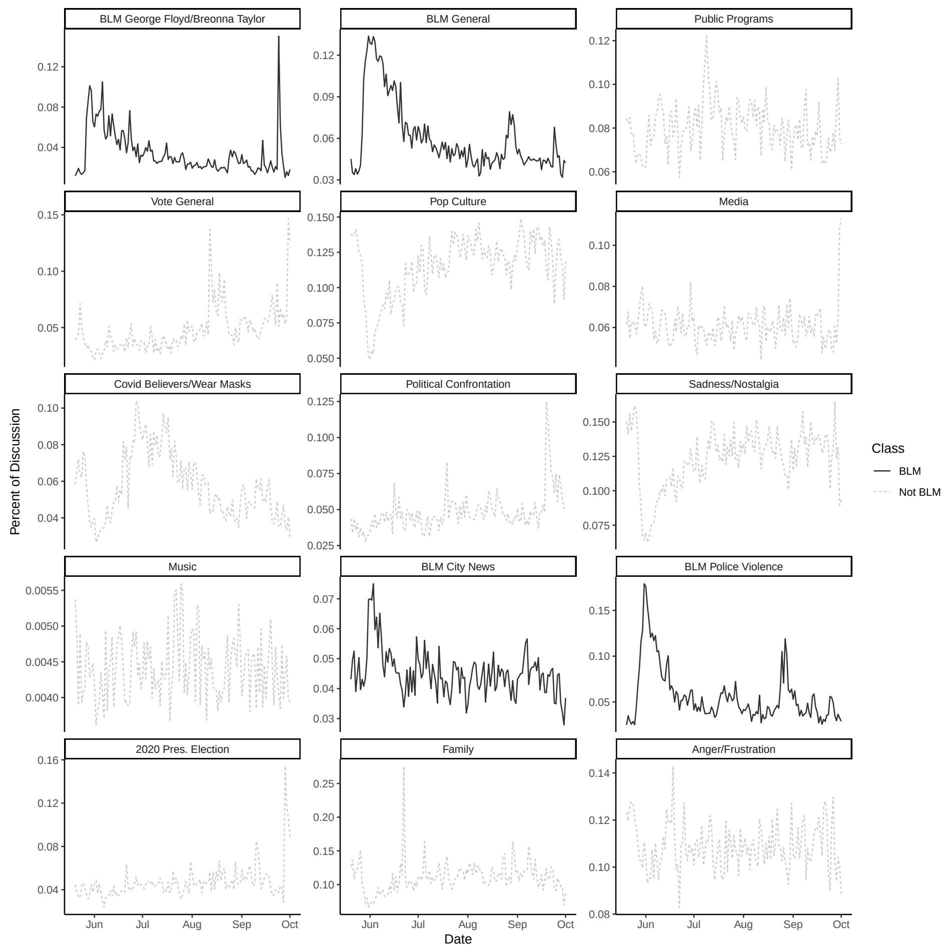


Figure A.3: Average distribution of senTopics over time.

We also label the tweets originally found when searching for protesters, and thus including at least one of our BLM relevant keywords, as *protest tweets* and the rest of the tweets an individual user publishes over the summer as *timeline tweets*. The box and whisker plot of the percent BLM topic for each city for these types of tweets can be seen in Figure A.4. In this way, we are using hand-labeled BLM tweets to check their consistency with the unsupervised topic modeling technique. The clear separation between the two groups further increases our confidence in the model.

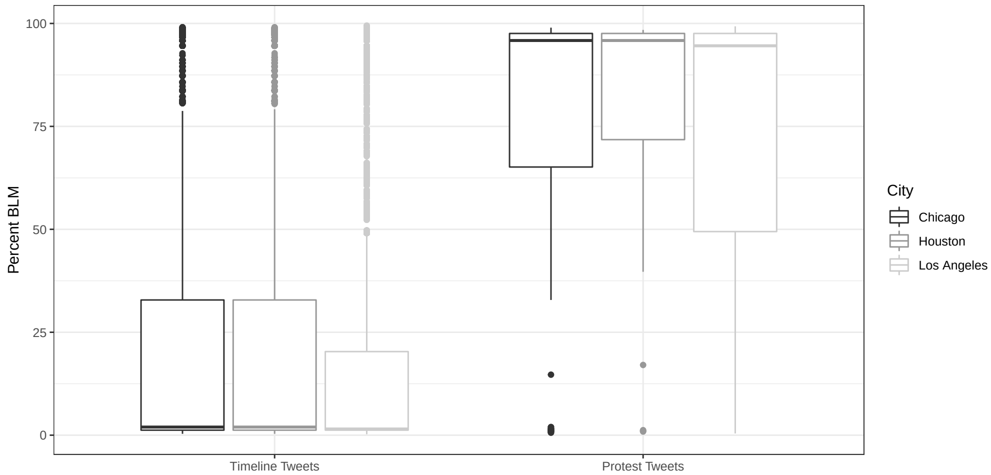


Figure A.4: Comparison of protest tweets and other tweets by protesters in an effort to validate BLM measure.

Finally, we selected the 400 tweets with the highest BLM rating, the 200 with the lowest BLM rating and then 200 closest to 50%. These tweets were then hand coded by four individuals on a scale of 0 to 1 for percent related to BLM. A boxplot for mean hand codings for each tweet can be seen in Figure A.5. It is clear from these responses that the unsupervised method is in line with the hand codings done by the four individuals. In addition in Table A.5 the correlation of the scores for each person as well as the RJST model can be seen.

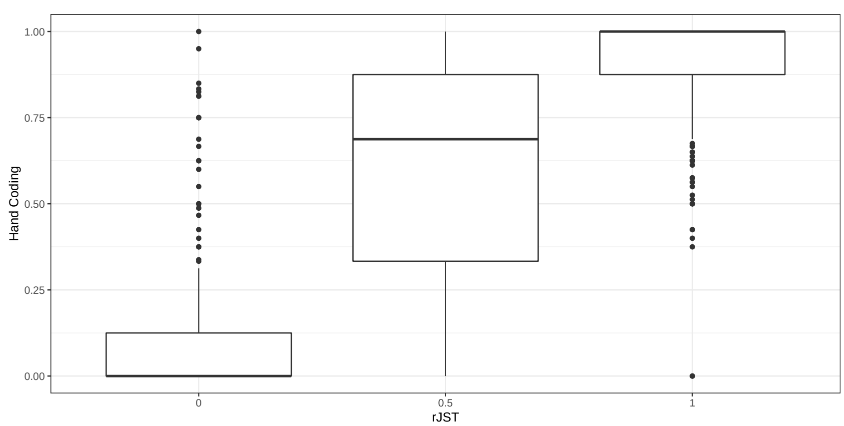


Figure A.5: Boxplots for the average score for each tweet based on the hand coders based on whether they were within the group closest to 0, 50, and 100 percent related to BLM according to the RJST model.

	BLM_topic	P1	P2	P3	P4	Avg
BLM_topic	1.00	0.78	0.76	0.61	0.76	0.80
P1	0.78	1.00	0.88	0.76	0.82	0.95
P2	0.76	0.88	1.00	0.72	0.75	0.92
P3	0.61	0.76	0.72	1.00	0.70	0.88
P4	0.76	0.82	0.75	0.70	1.00	0.90
Avg	0.80	0.95	0.92	0.88	0.90	1.00

Table A.5: Correlation matrix between the hand coded response of the four people and then RJST model. It is clear that the unsupervised model is as close to the individuals as they are to each other. This helps to validate our model and lends support to the conclusions drawn using it

A.3 Regression

Regression Information

Finally, we do some robustness checks on the the regressions run. First, in Figure A.6 we look at the daily tweeting of the individuals. We can see that more of our sample of Los Angeles tweeters tweet each day, yet, consistently, more than 50% of the sample is tweeting each day. Inspiring confidence that each day we are getting a large sample of tweets and not overweighting specific users. In addition, we can see that, for the most part, there are not distinct types of tweeters in terms of their frequencies—rather there is a continuum of frequencies. Regardless, for all the regressions to follow we also ran them restricting to users who tweet more than 25% and 50% of the days and see no major differences in the results.

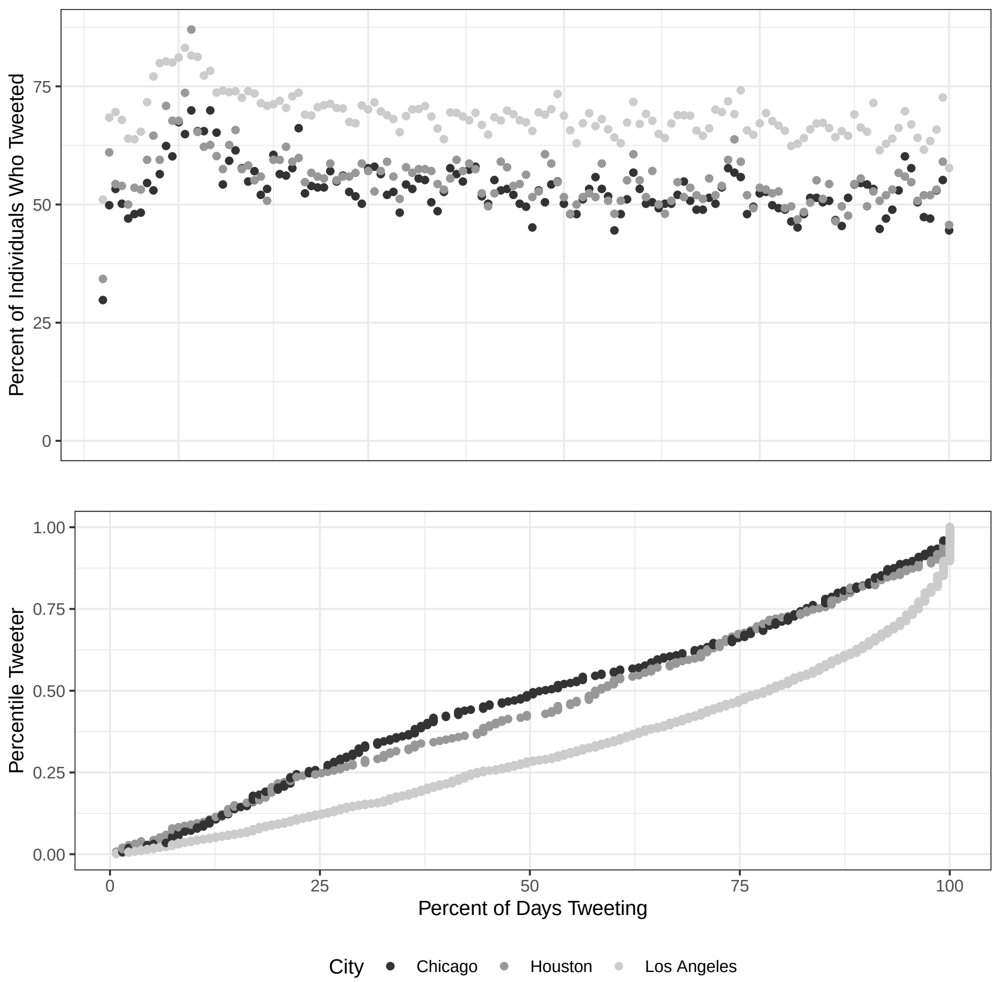


Figure A.6: Tweeting frequencies.

Regression Robustness Results

In this section, we list the regression results with interaction terms in order to justify the pooling of cities. In addition, we include placebo tests in order to show the robustness of our results. The lack of significant results when the protest day is falsely assigned to 10 days prior to or following the actual protest events validates even the small results we get in our main analysis. Finally, tests are done doing similar tests but for random groups of Topics. These results were null for one set and in the opposite direction than the BLM results for the other. This is a result of the fact that we are looking at the percent of time individuals spend discussing each time. If they discuss BLM more, this must be taken from the conversation they are having on other topics. Therefore, the significance of these results do not take away from the significance found in the paper.

Table A.6: OLS regression results with day and individual fixed effects and interaction terms.

	Dependent variable:			
	Interest	log(Interest)	Identity	log(Identity)
	(1)	(2)	(3)	(4)
Protest Day - 2	0.672	0.065	1.560*	0.009
	(0.673)	(0.040)	(0.894)	(0.045)
Protest Day - 1	1.499**	0.126***	−1.080	−0.070
	(0.656)	(0.039)	(0.871)	(0.044)
Protest Day	6.854***	0.450***	1.603**	−0.040
	(0.613)	(0.036)	(0.814)	(0.040)
Protest Day + 1	9.746***	0.617***	0.965	−0.075**
	(0.588)	(0.035)	(0.781)	(0.038)
Protest Day + 2	3.547***	0.205***	−0.775	−0.067*
	(0.619)	(0.037)	(0.823)	(0.040)
Protest Day + 3	2.621***	0.187***	−0.277	0.016
	(0.624)	(0.037)	(0.828)	(0.040)
Protest Day + 4	1.548**	0.119***	1.332	0.041
	(0.626)	(0.037)	(0.832)	(0.040)
pro_n2:stateIL	0.275	−0.085	−3.706	−0.259*
	(2.113)	(0.126)	(2.807)	(0.146)
pro_1:stateIL	0.060	−0.263**	−0.517	−0.043
	(1.779)	(0.106)	(2.364)	(0.123)
pro_1:stateTX	4.004*	−0.023	−1.895	−0.072
	(2.099)	(0.125)	(2.789)	(0.134)
pro_3:stateIL	−5.672***	−0.381***	−0.624	−0.172
	(2.108)	(0.126)	(2.799)	(0.141)
pro_3:stateTX	−9.243***	−0.598***	−3.606	−0.109
	(2.385)	(0.142)	(3.168)	(0.161)
Observations	165,301	165,301	165,301	105,555
R^2	0.003	0.004	0.0002	0.0003
Adjusted R^2	-0.015	-0.015	-0.019	-0.029
F Statistic	20.496***	21.748***	0.987	1.080

*p<0.1; **p<0.05; ***p<0.01

Note: Only the entries with significant results in at least one regression are listed, the rest are omitted.

Placebo Tests

Table A.7: OLS BLM regression with control comparison.

	Time 1	True	Time 2	Topic 1	Topic 2
Protest Day - 14	0.439				
	(0.988)				
Protest Day - 13	0.325				
	(0.971)				
Protest Day - 12	−0.533				
	(0.949)				
Protest Day - 11	−0.519				
	(0.930)				
Protest Day - 10	0.722				
	(0.940)				
Protest Day - 9	−0.839				
	(0.886)				
Protest Day - 8	−0.219				
	(0.801)				
Protest Day - 7	-1.360^{*}				
	(0.734)				
Protest Day - 6	−0.784				
	(0.723)				
Protest Day - 4		−0.360		1.199	−0.907
		(0.708)		(0.775)	(0.820)
Protest Day - 3		0.462		−0.980	0.027
		(0.729)		(0.798)	(0.844)
Protest Day - 2		0.928		0.252	−0.461
		(0.709)		(0.775)	(0.820)
Protest Day - 1		1.096		−0.319	−0.322
		(0.695)		(0.760)	(0.804)
Protest Day		7.748^{***}		-1.563^{**}	1.277^{*}
		(0.645)		(0.706)	(0.747)
Protest Day + 1		10.816^{***}		-3.258^{***}	−1.098
		(0.605)		(0.661)	(0.700)
Protest Day + 2		3.838^{***}		−0.527	0.303
		(0.656)		(0.718)	(0.760)
Protest Day + 3		1.807^{***}		1.028	−0.282
		(0.658)		(0.720)	(0.762)
Protest Day + 4		1.740^{***}		0.150	-1.810^{**}
		(0.660)		(0.722)	(0.764)
Protest Day + 6			0.281		
			(0.666)		
Protest Day + 7			0.740		
			(0.675)		
Protest Day + 8			0.467		
			(0.691)		
Protest Day + 9			−0.252		
			(0.730)		
Protest Day + 10			−0.425		
			(0.725)		
Protest Day + 11			−1.201		
			(0.752)		
Protest Day + 12			-1.289^{*}		
			(0.740)		
Protest Day + 13			0.561		
			(0.753)		
Protest Day + 14			1.750^{**}		
			(0.736)		
R^2	0.00004	0.003	0.0001	0.0002	0.0001
Adjusted R^2	-0.019	-0.016	-0.019	-0.019	-0.019
F Statistic	0.757	51.056^{***}	1.554	3.892^{***}	1.490

Note: N = 162,697, df = 9;159,669 *p<0.1; **p<0.05; ***p<0.01

Table A.8: Log OLS BLM regression with control comparison.

	Time 1	True	Time 2	Topic 1	Topic 2
Protest Day - 14	0.011				
	(0.059)				
Protest Day - 13	−0.015				
	(0.058)				
Protest Day - 12	−0.002				
	(0.056)				
Protest Day - 11	−0.036				
	(0.055)				
Protest Day - 10	0.068				
	(0.056)				
Protest Day - 9	−0.063				
	(0.053)				
Protest Day - 8	−0.023				
	(0.048)				
Protest Day - 7	-0.076^{*}				
	(0.044)				
Protest Day - 6	−0.020				
	(0.043)				
Protest Day - 4		−0.003		0.083^{*}	−0.007
		(0.042)		(0.045)	(0.045)
Protest Day - 3		0.024		−0.055	0.026
		(0.043)		(0.046)	(0.047)
Protest Day - 2		0.074^{*}		0.058	0.026
		(0.042)		(0.045)	(0.045)
Protest Day - 1		0.094^{**}		0.012	0.017
		(0.041)		(0.044)	(0.045)
Protest Day		0.480^{***}		−0.010	0.152^{***}
		(0.038)		(0.041)	(0.041)
Protest Day + 1		0.627^{***}		-0.104^{***}	0.049
		(0.036)		(0.038)	(0.039)
Protest Day + 2		0.221^{***}		0.039	0.077^{*}
		(0.039)		(0.042)	(0.042)
Protest Day + 3		0.135^{***}		0.102^{**}	0.041
		(0.039)		(0.042)	(0.042)
Protest Day + 4		0.151^{***}		0.053	−0.007
		(0.039)		(0.042)	(0.042)
Protest Day + 6			0.060		
			(0.040)		
Protest Day + 7			0.085^{**}		
			(0.040)		
Protest Day + 8			0.060		
			(0.041)		
Protest Day + 9			−0.015		
			(0.043)		
Protest Day + 10			−0.015		
			(0.043)		
Protest Day + 11			−0.048		
			(0.045)		
Protest Day + 12			-0.078^{*}		
			(0.044)		
Protest Day + 13			0.070		
			(0.045)		
Protest Day + 14			0.059		
			(0.044)		
R^2	0.00004	0.003	0.0001	0.0001	0.0001
Adjusted R^2	-0.019	-0.016	-0.019	-0.019	-0.019
F Statistic	0.738	51.458^{***}	1.909^{**}	2.522^{***}	2.023^{**}

Note: N = 162,697, df = 9; 159,669 $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table A.9: Percent plural OLS regression with control comparison.

	Time 1	True	Time 2	Topic 1	Topic 2
Protest Day - 14	1.904				
	(1.306)				
Protest Day - 13	1.068				
	(1.284)				
Protest Day - 12	0.358				
	(1.254)				
Protest Day - 11	0.229				
	(1.229)				
Protest Day - 10	1.058				
	(1.242)				
Protest Day - 9	−0.256				
	(1.171)				
Protest Day - 8	0.574				
	(1.059)				
Protest Day - 7	0.229				
	(0.970)				
Protest Day - 6	−0.215				
	(0.956)				
Protest Day - 4		−0.851		−0.851	−0.851
		(0.937)		(0.937)	(0.937)
Protest Day - 3		0.878		0.878	0.878
		(0.965)		(0.965)	(0.965)
Protest Day - 2		1.652*		1.652*	1.652*
		(0.938)		(0.938)	(0.938)
Protest Day - 1		−1.495		−1.495	−1.495
		(0.919)		(0.919)	(0.919)
Protest Day		1.227		1.227	1.227
		(0.854)		(0.854)	(0.854)
Protest Day + 1		1.173		1.173	1.173
		(0.800)		(0.800)	(0.800)
Protest Day + 2		−1.146		−1.146	−1.146
		(0.869)		(0.869)	(0.869)
Protest Day + 3		−1.042		−1.042	−1.042
		(0.871)		(0.871)	(0.871)
Protest Day + 4		0.749		0.749	0.749
		(0.874)		(0.874)	(0.874)
Protest Day + 6			0.309		
			(0.880)		
Protest Day + 7			−0.919		
			(0.892)		
Protest Day + 8			−1.179		
			(0.913)		
Protest Day + 9			−0.871		
			(0.965)		
Protest Day + 10			1.254		
			(0.958)		
Protest Day + 11			0.908		
			(0.994)		
Protest Day + 12			−0.909		
			(0.978)		
Protest Day + 13			0.269		
			(0.995)		
Protest Day + 14			−1.295		
			(0.973)		
R^2	0.00003	0.0001	0.0001	0.0001	0.0001
Adjusted R^2	-0.019	-0.019	-0.019	-0.019	-0.019
F Statistic	0.445	1.838*	1.006	1.838*	1.838*

Note: N = 162,697, df = 9; 159,669 — *p<0.1; **p<0.05; ***p<0.01

Table A.10: Log percent plural OLS regression with control comparison.

	Time 1	True	Time 2	Topic 1	Topic 2
Protest Day - 14	0.305				
	(0.186)				
Protest Day - 13	0.004				
	(0.183)				
Protest Day - 12	0.254				
	(0.179)				
Protest Day - 11	−0.091				
	(0.175)				
Protest Day - 10	0.070				
	(0.177)				
Protest Day - 9	0.142				
	(0.167)				
Protest Day - 8	0.054				
	(0.151)				
Protest Day - 7	0.038				
	(0.138)				
Protest Day - 6	−0.001				
	(0.136)				
Protest Day - 4		−0.117		−0.117	−0.117
		(0.134)		(0.134)	(0.134)
Protest Day - 3		0.152		0.152	0.152
		(0.138)		(0.138)	(0.138)
Protest Day - 2		0.326^{**}		0.326^{**}	0.326^{**}
		(0.134)		(0.134)	(0.134)
Protest Day - 1		−0.032		−0.032	−0.032
		(0.131)		(0.131)	(0.131)
Protest Day		0.462^{***}		0.462^{***}	0.462^{***}
		(0.122)		(0.122)	(0.122)
Protest Day + 1		0.595^{***}		0.595^{***}	0.595^{***}
		(0.114)		(0.114)	(0.114)
Protest Day + 2		0.076		0.076	0.076
		(0.124)		(0.124)	(0.124)
Protest Day + 3		0.085		0.085	0.085
		(0.124)		(0.124)	(0.124)
Protest Day + 4		0.194		0.194	0.194
		(0.125)		(0.125)	(0.125)
Protest Day + 6			0.049		
			(0.126)		
Protest Day + 7			0.062		
			(0.127)		
Protest Day + 8			−0.124		
			(0.130)		
Protest Day + 9			−0.069		
			(0.138)		
Protest Day + 10			0.138		
			(0.137)		
Protest Day + 11			0.148		
			(0.142)		
Protest Day + 12			0.012		
			(0.140)		
Protest Day + 13			0.210		
			(0.142)		
Protest Day + 14			−0.018		
			(0.139)		
R^2	0.00004	0.0003	0.00004	0.0003	0.0003
Adjusted R^2	-0.019	-0.019	-0.019	-0.019	-0.019
F Statistic	0.657	5.402^{***}	0.650	5.402^{***}	5.402^{***}

Note: N = 162,697, df = 9; 159,669 *p<0.1; **p<0.05; ***p<0.01

Appendix B

THE 2022 U.S. MIDTERM ELECTION

B.1 Survey Question Wording

Vote Choice

House vote: In the November 2022 election for U.S. Congress in the district where you live, which candidate did you vote for?

- Democratic candidate
- Republican candidate
- Neither
- Not sure
- Didn't vote

Economics

Financial Situation: We are interested in how people are getting along financially these days. Would you say that you and your family living here are better off or worse off financially than you were a year ago?

- Better off
- The same
- Worse off

Economic Situation: Now thinking about the economy. Would you say that over the past year the nation's economy has gotten better, stayed the same, or gotten worse?

- Gotten better
- Stayed the same
- Gotten worse

Issues

Most Important: How important, if at all, were each of the following issues for you as you thought about whom you would vote for in the congressional election in

your area in November 2022?
Rows:

- Immigration
- Abortion
- Foreign Policy
- Economic Inequality
- The COVID-19 outbreak
- Violent crime
- Health care
- The economy
- Racial and ethnic inequality
- Climate change
- Inflation
- Gun policy
- Supreme Court appointments

Columns:

- Very important
- Somewhat important
- Not too important
- Not important at all

Better Job: Which political party would do a better job with:
Rows:

- Preventing terrorism
- Mitigating climate change
- Abortion policy
- Law enforcement and criminal justice reform
- Preventing further spread of Covid-19
- Reducing the federal budget deficit
- Growing the economy

- Providing affordable healthcare
- American foreign policy
- Inflation

Columns:

- Democrat
- Republican
- Not sure

Demographics

Race: What racial or ethnic group best describes you?

- White
- Black or African American
- Hispanic or Latino
- Asian American
- American Indian/Native American
- Arab, Middle Eastern, or North African
- Native Hawaiian
- Not Hawaiian, but other Pacific Islander

Religion: What is your religious preference? Is it Protestant, Catholic, Jewish, Muslim, some other religion, or no religion?

- Protestant
- Catholic
- Jewish
- Muslim
- Some other religion
- No religion

Education: What is the highest level of education you have completed?

- No HS
- High school graduate

- Some college
- 2-year
- 4-year
- Post-grad

Age: Respondent age by category

- Under 30
- 30-44
- 54-64
- 65+

Region: Calculated from respondent's state of residence

- Northeast
- Midwest
- South
- West

Gender: Which gender identity do you most identify with?

- Woman
- Man
- Non-Binary/Fluid
- Prefer not to say

B.2 Crosstabulations

Characteristic	Democrat	Republican	Neither	Not sure	Didn't vote
Party ID					
Democrat	93%	1.4%	1.2%	0.5%	3.5%
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Republican	4.9%	89%	1.9%	0.8%	3.0%
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)

Independent	37% (0.02)	42% (0.02)	9.0% (0.01)	2.1% (0.01)	10% (0.01)
Other	36% (0.06)	47% (0.06)	7.1% (0.04)	3.8% (0.02)	5.9% (0.02)
Not sure	34% (0.10)	13% (0.06)	10% (0.05)	14% (0.08)	28% (0.09)
% (SE(%))					

Table B.1: Weighted congressional party vote choice by party ID.

Characteristic	**Democrat**	**Republican**	**Neither**	**Not sure**	**Didn't vote**
Party ID					
Democrat	740 (93%)	14 (1.8%)	11 (1.4%)	4 (0.5%)	25 (3.1%)
Republican	24 (3.9%)	565 (91%)	10 (1.6%)	5 (0.8%)	19 (3.0%)
Independent	215 (37%)	234 (41%)	60 (10%)	11 (1.9%)	54 (9.4%)
Other	29 (36%)	38 (48%)	4 (5.0%)	3 (3.8%)	6 (7.5%)
Not sure	10 (26%)	6 (16%)	5 (13%)	4 (11%)	13 (34%)
n(unweighted) (% unweighted)					

Table B.2: Unweighted congressional party vote choice by party ID.

Characteristic	**Democratic candidate**	**Republican candidate**
Gender		
Woman	54% (0.02)	46% (0.02)
Man	44% (0.02)	56% (0.02)
Non-Binary/Fluid	96% (0.04)	4.3% (0.04)
Prefer not to say	34% (0.16)	66% (0.16)
Educational Attainment		
No HS	44% (0.08)	56% (0.08)
High school graduate	44% (0.03)	56% (0.03)
Some college	50% (0.03)	50% (0.03)
2-year	51% (0.04)	49% (0.04)
4-year	50% (0.02)	50% (0.02)

Post-grad	63% (0.03)	37% (0.03)
Region		
Northeast	57% (0.03)	43% (0.03)
Midwest	46% (0.03)	54% (0.03)
South	45% (0.02)	55% (0.02)
West	57% (0.03)	43% (0.03)
Race		
White	43% (0.01)	57% (0.01)
Black or African American	84% (0.03)	16% (0.03)
Hispanic or Latino	59% (0.05)	41% (0.05)
Asian American	61% (0.09)	39% (0.09)
American Indian/Native American	37% (0.09)	63% (0.09)
Arab, Middle Eastern, or North African	77% (0.19)	23% (0.19)
Native Hawaiian	100% (0.00)	0% (0.00)
Not Hawaiian, but other Pacific Islander	65% (0.19)	35% (0.19)
Religion		
Protestant	35% (0.02)	65% (0.02)
Catholic	43% (0.03)	57% (0.03)
Jewish	67% (0.06)	33% (0.06)
Muslim	81% (0.11)	19% (0.11)
Some other religion	50% (0.03)	50% (0.03)
No religion	71% (0.02)	29% (0.02)
Age		
Under 30	63% (0.04)	37% (0.04)
30-44	54% (0.03)	46% (0.03)
45-64	43% (0.02)	57% (0.02)
65+	50% (0.02)	50% (0.02)
% (SE(%))		

Table B.3: Weighted demographics by party ID.

Characteristic	**Democratic candidate**	**Republican candidate**
Gender		
Woman	584 (58%)	423 (42%)
Man	412 (49%)	427 (51%)

Non-Binary/Fluid	19 (95%)	1 (5.0%)
Prefer not to say	3 (33%)	6 (67%)
Educational Attainment		
No HS	23 (48%)	25 (52%)
High school graduate	191 (48%)	211 (52%)
Some college	225 (54%)	191 (46%)
2-year	118 (55%)	95 (45%)
4-year	262 (53%)	231 (47%)
Post-grad	199 (66%)	104 (34%)
Region		
Northeast	241 (61%)	153 (39%)
Midwest	208 (55%)	167 (45%)
South	311 (46%)	362 (54%)
West	258 (60%)	175 (40%)
Race		
White	696 (48%)	741 (52%)
Black or African American	185 (86%)	29 (14%)
Hispanic or Latino	91 (65%)	50 (35%)
Asian American	20 (62%)	12 (38%)
American Indian/Native American	11 (34%)	21 (66%)
Arab, Middle Eastern, or North African	6 (86%)	1 (14%)
Native Hawaiian	6 (100%)	0 (0%)
Not Hawaiian, but other Pacific Islander	3 (50%)	3 (50%)
Religion		
Protestant	237 (40%)	360 (60%)
Catholic	185 (45%)	224 (55%)
Jewish	44 (68%)	21 (32%)
Muslim	10 (77%)	3 (23%)
Some other religion	150 (55%)	122 (45%)
No religion	392 (76%)	127 (24%)
Age		
Under 30	145 (67%)	73 (33%)
30-44	195 (58%)	140 (42%)
45-64	348 (48%)	370 (52%)
65+	330 (55%)	274 (45%)

n (unweighted) (% (unweighted))

Table B.4: Unweighted demographics by party ID.

Characteristic	**Democratic candidate**	**Republican candidate**
Immigration		
Very important	29% (0.02)	71% (0.02)
Somewhat important	65% (0.02)	35% (0.02)
Not too important	85% (0.02)	15% (0.02)
Not important at all	79% (0.05)	21% (0.05)
Abortion		
Very important	67% (0.02)	33% (0.02)
Somewhat important	51% (0.03)	49% (0.03)
Not too important	24% (0.03)	76% (0.03)
Not important at all	12% (0.02)	88% (0.02)
Foreign Policy		
Very important	42% (0.02)	58% (0.02)
Somewhat important	55% (0.02)	45% (0.02)
Not too important	62% (0.03)	38% (0.03)
Not important at all	43% (0.07)	57% (0.07)
Economic Inequality		
Very important	77% (0.02)	23% (0.02)
Somewhat important	56% (0.02)	44% (0.02)
Not too important	20% (0.03)	80% (0.03)
Not important at all	6.3% (0.01)	94% (0.01)
COVID-19		
Very important	73% (0.02)	27% (0.02)
Somewhat important	65% (0.02)	35% (0.02)
Not too important	37% (0.03)	63% (0.03)
Not important at all	11% (0.02)	89% (0.02)
Violent Crime		
Very important	36% (0.02)	64% (0.02)
Somewhat important	61% (0.02)	39% (0.02)
Not too important	81% (0.03)	19% (0.03)

Not important at all	85% (0.06)	15% (0.06)
Health Care		
Very important	68% (0.02)	32% (0.02)
Somewhat important	38% (0.02)	62% (0.02)
Not too important	19% (0.03)	81% (0.03)
Not important at all	13% (0.04)	87% (0.04)
The Economy		
Very important	38% (0.01)	62% (0.01)
Somewhat important	79% (0.02)	21% (0.02)
Not too important	90% (0.04)	10% (0.04)
Not important at all	52% (0.11)	48% (0.11)
Racial and Ethnic Inequality		
Very important	81% (0.02)	19% (0.02)
Somewhat important	59% (0.02)	41% (0.02)
Not too important	26% (0.03)	74% (0.03)
Not important at all	7.0% (0.01)	93% (0.01)
Climate Change		
Very important	84% (0.02)	16% (0.02)
Somewhat important	60% (0.03)	40% (0.03)
Not too important	23% (0.03)	77% (0.03)
Not important at all	4.7% (0.01)	95% (0.01)
Inflation		
Very important	35% (0.01)	65% (0.01)
Somewhat important	80% (0.02)	20% (0.02)
Not too important	91% (0.03)	9.2% (0.03)
Not important at all	67% (0.10)	33% (0.10)
Gun Policy		
Very important	61% (0.02)	39% (0.02)
Somewhat important	47% (0.02)	53% (0.02)
Not too important	34% (0.03)	66% (0.03)
Not important at all	18% (0.03)	82% (0.03)
Supreme Court Appointments		
Very important	60% (0.02)	40% (0.02)
Somewhat important	44% (0.02)	56% (0.02)
Not too important	34% (0.03)	66% (0.03)
Not important at all	20% (0.04)	80% (0.04)

% (SE(%))

Table B.5: Weighted most important issue by party ID.

Characteristic	**Democratic candidate**	**Republican candidate**
Immigration		
Very important	286 (31%)	635 (69%)
Somewhat important	402 (70%)	174 (30%)
Not too important	259 (88%)	35 (12%)
Not important at all	70 (84%)	13 (16%)
Abortion		
Very important	737 (72%)	289 (28%)
Somewhat important	184 (53%)	160 (47%)
Not too important	62 (25%)	182 (75%)
Not important at all	35 (13%)	226 (87%)
Foreign Policy		
Very important	336 (46%)	398 (54%)
Somewhat important	481 (59%)	339 (41%)
Not too important	172 (66%)	89 (34%)
Not important at all	29 (48%)	31 (52%)
Economic Inequality		
Very important	603 (81%)	140 (19%)
Somewhat important	328 (61%)	214 (39%)
Not too important	64 (22%)	230 (78%)
Not important at all	23 (7.8%)	273 (92%)
COVID-19		
Very important	438 (78%)	124 (22%)
Somewhat important	378 (68%)	175 (32%)
Not too important	146 (40%)	220 (60%)
Not important at all	54 (14%)	338 (86%)
Violent Crime		
Very important	407 (39%)	640 (61%)
Somewhat important	346 (67%)	174 (33%)
Not too important	203 (85%)	36 (15%)

Not important at all	62 (90%)	7 (10%)
Health Care		
Very important	722 (72%)	275 (28%)
Somewhat important	246 (40%)	362 (60%)
Not too important	38 (19%)	157 (81%)
Not important at all	12 (16%)	63 (84%)
The Economy		
Very important	531 (41%)	757 (59%)
Somewhat important	389 (82%)	83 (18%)
Not too important	83 (91%)	8 (8.8%)
Not important at all	15 (62%)	9 (38%)
Racial and Ethnic Inequality		
Very important	588 (85%)	106 (15%)
Somewhat important	306 (62%)	188 (38%)
Not too important	93 (28%)	245 (72%)
Not important at all	31 (8.9%)	318 (91%)
Climate Change		
Very important	665 (88%)	95 (12%)
Somewhat important	267 (65%)	144 (35%)
Not too important	67 (25%)	206 (75%)
Not important at all	19 (4.4%)	412 (96%)
Inflation		
Very important	466 (38%)	767 (62%)
Somewhat important	355 (84%)	69 (16%)
Not too important	169 (93%)	13 (7.1%)
Not important at all	27 (77%)	8 (23%)
Gun Policy		
Very important	657 (66%)	339 (34%)
Somewhat important	243 (51%)	237 (49%)
Not too important	84 (35%)	154 (65%)
Not important at all	34 (21%)	126 (79%)
Supreme Court Appointments		
Very important	666 (65%)	365 (35%)
Somewhat important	250 (47%)	282 (53%)
Not too important	80 (37%)	136 (63%)
Not important at all	22 (23%)	74 (77%)

n (unweighted) (% (unweighted))

Table B.6: Unweighted most important issue by party ID.

Characteristic	**Democratic candidate**	**Republican candidate**
Preventing Terrorism		
Democrat	91% (0.01)	8.5% (0.01)
Republican	13% (0.01)	87% (0.01)
Not sure	75% (0.03)	25% (0.03)
Mitigating Climate Change		
Democrat	84% (0.01)	16% (0.01)
Republican	14% (0.02)	86% (0.02)
Not sure	22% (0.02)	78% (0.02)
Abortion Policy		
Democrat	87% (0.01)	13% (0.01)
Republican	10.0% (0.01)	90% (0.01)
Not sure	29% (0.03)	71% (0.03)
Criminal justice reform		
Democrat	93% (0.01)	7.3% (0.01)
Republican	12% (0.01)	88% (0.01)
Not sure	65% (0.03)	35% (0.03)
COVID-19		
Democrat	90% (0.01)	9.6% (0.01)
Republican	11% (0.02)	89% (0.02)
Not sure	28% (0.02)	72% (0.02)
Deficit		
Democrat	93% (0.01)	6.5% (0.01)
Republican	14% (0.01)	86% (0.01)
Not sure	63% (0.03)	37% (0.03)
The Economy		
Democrat	92% (0.01)	7.7% (0.01)
Republican	11% (0.01)	89% (0.01)
Not sure	75% (0.03)	25% (0.03)
Health Care		

Democrat	87% (0.01)	13% (0.01)
Republican	11% (0.02)	89% (0.02)
Not sure	22% (0.02)	78% (0.02)
Foreign Policy		
Democrat	94% (0.01)	6.2% (0.01)
Republican	11% (0.01)	89% (0.01)
Not sure	63% (0.03)	37% (0.03)
Inflation		
Democrat	91% (0.01)	9.0% (0.01)
Republican	12% (0.01)	88% (0.01)
Not sure	82% (0.02)	18% (0.02)
% (SE(%))		

Table B.7: Weighted ability by party ID.

Characteristic	**Democratic candidate**	**Republican candidate**
Preventing Terrorism		
Democrat	632 (94%)	43 (6.4%)
Republican	115 (13%)	737 (87%)
Not sure	271 (78%)	77 (22%)
Mitigating Climate Change		
Democrat	851 (86%)	135 (14%)
Republican	66 (14%)	402 (86%)
Not sure	101 (24%)	320 (76%)
Abortion Policy		
Democrat	872 (89%)	112 (11%)
Republican	73 (11%)	578 (89%)
Not sure	73 (30%)	167 (70%)
Criminal justice reform		
Democrat	734 (94%)	47 (6.0%)
Republican	113 (13%)	734 (87%)
Not sure	171 (69%)	76 (31%)
COVID-19		
Democrat	823 (92%)	69 (7.7%)
Republican	57 (11%)	483 (89%)

Not sure	138 (31%)	305 (69%)
Deficit		
Democrat	652 (95%)	32 (4.7%)
Republican	129 (16%)	697 (84%)
Not sure	237 (65%)	128 (35%)
The Economy		
Democrat	749 (94%)	46 (5.8%)
Republican	108 (12%)	761 (88%)
Not sure	161 (76%)	50 (24%)
Health Care		
Democrat	885 (89%)	108 (11%)
Republican	58 (10%)	510 (90%)
Not sure	75 (24%)	239 (76%)
Foreign Policy		
Democrat	757 (95%)	41 (5.1%)
Republican	100 (12%)	734 (88%)
Not sure	161 (66%)	82 (34%)
Inflation		
Democrat	667 (93%)	50 (7.0%)
Republican	120 (14%)	759 (86%)
Not sure	231 (83%)	48 (17%)
n (unweighted) (% (unweighted))		

Table B.8: Unweighted ability by party ID.

Characteristic	**Democratic candidate**	**Republican candidate**
Economic Situation		
Gotten better	82% (0.03)	18% (0.03)
Stayed the same	80% (0.02)	20% (0.02)
Gotten worse	34% (0.01)	66% (0.01)
Financial Situation		
Better off	70% (0.03)	30% (0.03)
The same	65% (0.02)	35% (0.02)
Worse off	30% (0.02)	70% (0.02)
% (SE(%))		

Table B.9: Weighted views on financial situation by party ID.

Characteristic	**Democratic candidate**	**Republican candidate**
Economic Situation		
Gotten better	212 (89%)	27 (11%)
Stayed the same	336 (84%)	65 (16%)
Gotten worse	470 (38%)	765 (62%)
Financial Situation		
Better off	193 (76%)	62 (24%)
The same	545 (69%)	240 (31%)
Worse off	280 (34%)	555 (66%)
n (unweighted) (% (unweighted))		

Table B.10: Unweighted views on financial situation by party.

B.3 Regression Result Figures

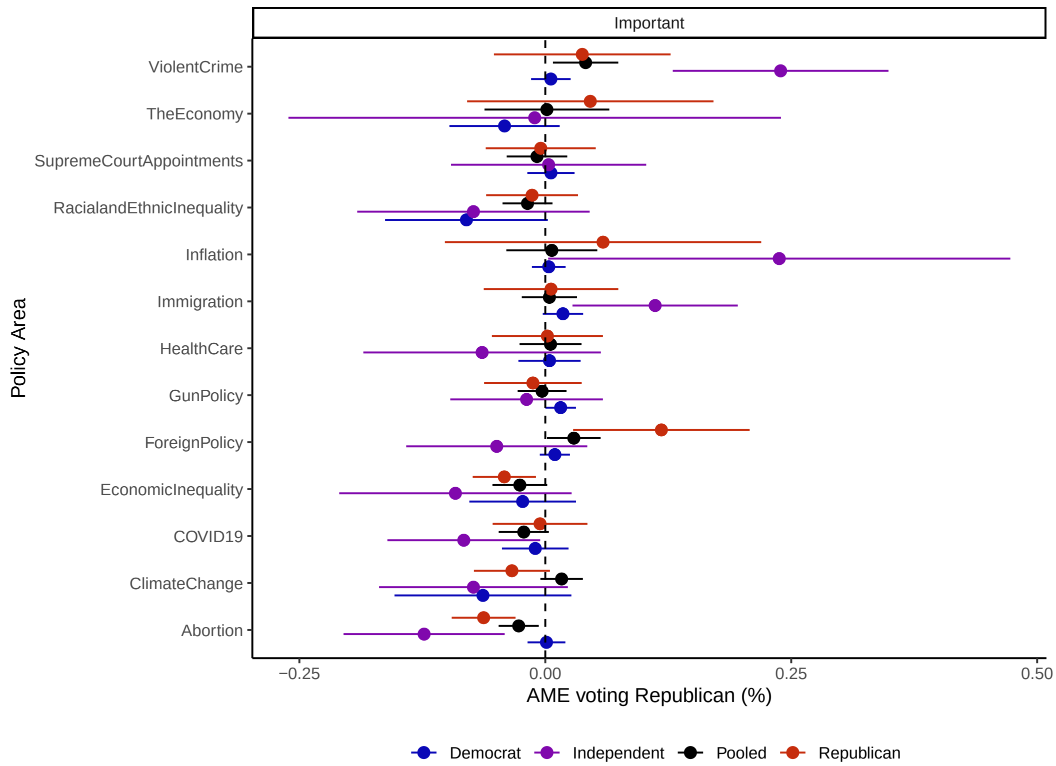


Figure B.1: Average marginal effect (with 95% confidence intervals) of viewing policy areas as important on the probability of voting for the Republican congressional candidate. The plots show the results for the party-based model for each party as well as the pooled model results.

B.4 Regression Result Tables

	Average Maginal Effects:			
	Probability Voting for Republican for Congress			
	Full	$p < 0.1$	$p < 0.05$	$p < 0.01$
Economics				
Follow Goverment				
Don't know	0.106**	0.124***	0.118**	
	(0.041)	(0.048)	(0.053)	
Hardly at all	0.025*	0.027	0.027*	
	(0.014)	(0.017)	(0.016)	

Most of the time	−0.012	−0.015	−0.015	
	(0.012)	(0.011)	(0.011)	
Only now and then	0.001	0.002	0.004	
	(0.013)	(0.014)	(0.014)	
EconomicSituation				
Gotten better	0.027	0.025	0.023	
	(0.026)	(0.019)	(0.021)	
Gotten worse	0.031**	0.033**	0.032**	
	(0.015)	(0.014)	(0.014)	
Financial Situation				
Better off	0.002			
	(0.019)			
Worse off	0.002			
	(0.011)			
Party Abilities				
Health Care				
Democrat	−0.058***	−0.064***	−0.059***	−0.146***
	(0.016)	(0.019)	(0.018)	(0.025)
Republican	−0.053***	−0.057***	−0.052***	−0.034
	(0.014)	(0.017)	(0.016)	(0.022)
Inflation				
Democrat	0.105***	0.076**	0.075**	0.049
	(0.039)	(0.031)	(0.029)	(0.034)
Republican	0.094**	0.097***	0.098***	0.180***
	(0.044)	(0.031)	(0.029)	(0.033)
Abortion				
Democrat	−0.044**	−0.066***	−0.062***	
	(0.018)	(0.019)	(0.019)	
Republican	0.029*	0.035	0.038	
	(0.016)	(0.022)	(0.024)	
Covid				
Democrat	−0.047**	−0.063**	−0.064***	
	(0.024)	(0.026)	(0.023)	
Republican	−0.004	−0.006	−0.012	
	(0.014)	(0.016)	(0.015)	

Crime				
Democrat	0.035**	0.027	0.023	
	(0.016)	(0.019)	(0.020)	
Republican	0.018	0.027*	0.026*	
	(0.016)	(0.014)	(0.014)	
Budget				
Democrat	−0.025			
	(0.022)			
Republican	−0.001			
	(0.015)			
Foreign Policy				
Democrat	−0.001			
	(0.020)			
Republican	0.004			
	(0.017)			
Climate				
Democrat	−0.015			
	(0.014)			
Republican	−0.017			
	(0.012)			
Grow Economy				
Democrat	−0.028			
	(0.025)			
Republican	−0.00004			
	(0.017)			
Law Enforcement				
Democrat	−0.021			
	(0.023)			
Republican	0.018			
	(0.017)			
Important Issues				
Abortion	−0.032***	−0.036***	−0.038***	−0.068***
	(0.011)	(0.012)	(0.012)	(0.014)
Foreign Policy	0.035***	0.032***	0.025**	0.022
	(0.013)	(0.011)	(0.012)	(0.017)

Economic Inequality	−0.026**	−0.046***	−0.046***	
	(0.013)	(0.013)	(0.014)	
Violent Crime	0.041**	0.041***	0.039**	
	(0.016)	(0.015)	(0.016)	
Climate Change	0.018*	0.015		
	(0.011)	(0.012)		
COVID19	−0.023*	−0.025*		
	(0.014)	(0.014)		
Gun Policy	−0.005			
	(0.013)			
Health Care	0.006			
	(0.015)			
Immigration	0.005			
	(0.013)			
Inflation	0.003			
	(0.022)			
Racial and Ethnic Inequality	−0.016			
	(0.012)			
Supreme Court Appointments	−0.003			
	(0.015)			
The Economy	0.004			
	(0.029)			
Demographics				
PartyID				
Democrat	−0.239***	−0.258***	−0.261***	−0.320***
	(0.041)	(0.040)	(0.039)	(0.039)
Not sure	−0.080	−0.120**	−0.122**	−0.117**
	(0.063)	(0.061)	(0.060)	(0.059)
Other	−0.014	−0.013	−0.006	−0.011
	(0.029)	(0.030)	(0.030)	(0.033)
Republican	0.158***	0.174***	0.171***	0.246***
	(0.032)	(0.032)	(0.033)	(0.036)
Race				

American Indian/Native American	−0.053*	−0.050*	−0.046*	−0.056
	(0.030)	(0.027)	(0.028)	(0.034)
Arab, Middle Eastern, or North African	−0.103*	−0.103**	−0.119**	−0.107**
	(0.058)	(0.051)	(0.056)	(0.051)
Asian American	0.046	0.044*	0.040	0.022
	(0.029)	(0.027)	(0.026)	(0.024)
Black or African American	−0.044**	−0.053***	−0.056***	−0.053**
	(0.020)	(0.019)	(0.020)	(0.026)
Hispanic or Latino	−0.001	−0.001	−0.005	0.007
	(0.014)	(0.016)	(0.016)	(0.021)
Native Hawaiian	−0.501***	−0.503***	−0.505***	−0.505***
	(0.006)	(0.006)	(0.006)	(0.006)
Not Hawaiian, but other Pacific Islander	0.056	0.058	0.055	−0.047
	(0.104)	(0.091)	(0.084)	(0.113)
Region				
Midwest	0.0003	0.0004	−0.001	0.003
	(0.016)	(0.016)	(0.015)	(0.016)
South	0.013	0.013	0.015	0.025
	(0.013)	(0.013)	(0.013)	(0.015)
West	0.002	−0.001	0.001	0.008
	(0.014)	(0.013)	(0.013)	(0.017)
Religion				
Catholic	0.003	0.002	0.003	−0.005
	(0.011)	(0.013)	(0.013)	(0.017)
Jewish	−0.053*	−0.048*	−0.053*	−0.069**
	(0.031)	(0.026)	(0.029)	(0.035)
Muslim	0.055*	0.059**	0.058**	0.017
	(0.028)	(0.028)	(0.028)	(0.036)
No religion	−0.013	−0.015	−0.015	−0.040***
	(0.012)	(0.012)	(0.012)	(0.015)
Some other religion	−0.009	−0.005	−0.004	−0.005
	(0.016)	(0.017)	(0.017)	(0.019)

Age				
30-44	0.018	0.017	0.018	0.007
	(0.012)	(0.011)	(0.011)	(0.014)
45-64	0.009	0.007	0.009	0.018
	(0.012)	(0.011)	(0.012)	(0.012)
Under 30	−0.023	−0.031*	−0.030*	−0.024
	(0.017)	(0.017)	(0.018)	(0.022)
Education				
College Graduate	−0.011	−0.013	−0.011	−0.020*
	(0.010)	(0.010)	(0.010)	(0.010)
Post Grad	−0.020	−0.027	−0.025	−0.038*
	(0.020)	(0.020)	(0.021)	(0.021)
Gender				
Non-Binary/Fluid	−0.022	−0.038	−0.050	−0.029
	(0.071)	(0.072)	(0.082)	(0.063)
Prefer not to say	0.058**	0.059**	0.056**	0.095*
	(0.024)	(0.025)	(0.025)	(0.049)
Woman	−0.006	−0.006	−0.006	−0.010
	(0.009)	(0.009)	(0.009)	(0.010)
Observations	1,871	1,871	1,871	1,871
Note:				*p<0.1; **p<0.05; ***p<0.01

Table B.11: Comparison of pooled models.

	Dependent variable:			
	Congressional Vote			
	Pooled	Republican	Democrat	Independent
Economics				
EconomicSituation				
Gotten better	0.023	0.019	−0.009***	−0.502***
	(0.021)	(0.021)	(0.0001)	(0.025)
Gotten worse	0.032**	0.019	0.031***	0.050*

	(0.014)	(0.019)	(0.001)	(0.029)
FollowGoverment				
Don't know	0.118**	0.031***	0.102***	−0.516***
	(0.053)	(0.008)	(0.003)	(0.018)
Hardly at all	0.027*	0.016	0.003***	0.120**
	(0.016)	(0.012)	(0.001)	(0.048)
Most of the time	−0.015	−0.026	−0.027***	0.008
	(0.011)	(0.017)	(0.001)	(0.022)
Only now and then	0.004	−0.013	−0.035***	0.059
	(0.014)	(0.017)	(0.001)	(0.051)
Party Abilities				
Abortion				
Democrat	−0.062***	−0.064	−0.065***	−0.049
	(0.019)	(0.048)	(0.002)	(0.047)
Republican	0.038	0.017	0.013***	0.131***
	(0.024)	(0.030)	(0.002)	(0.049)
Covid				
Democrat	−0.064***	−0.062***	−0.066***	−0.146***
	(0.023)	(0.023)	(0.002)	(0.050)
Republican	−0.012	−0.003	−0.080***	−0.075*
	(0.015)	(0.015)	(0.002)	(0.043)
Crime				
Democrat	0.023	0.034	−0.000	0.118***
	(0.020)	(0.049)	(0.000)	(0.037)
Republican	0.026*	0.036	−0.022***	0.066**
	(0.014)	(0.048)	(0.00000)	(0.032)
HealthCare				
Democrat	−0.059***	−0.046*	0.0005**	−0.064
	(0.018)	(0.023)	(0.0002)	(0.047)
Republican	−0.052***	−0.077***	0.176***	−0.012
	(0.016)	(0.018)	(0.004)	(0.032)
Inflation				
Democrat	0.075**	0.189	0.012***	−0.002
	(0.029)	(0.183)	(0.003)	(0.048)
Republican	0.098***	0.210	0.023***	0.135***
	(0.029)	(0.184)	(0.001)	(0.046)

Important Issues				
Abortion	−0.038***	−0.020	−0.00004	−0.110***
	(0.012)	(0.015)	(0.0001)	(0.038)
EconomicInequality	−0.046***	−0.041***	−0.083***	−0.062*
	(0.014)	(0.012)	(0.002)	(0.033)
ForeignPolicy	0.025**	0.053*	0.017***	−0.003
	(0.012)	(0.031)	(0.00000)	(0.027)
ViolentCrime	0.039**	0.041	−0.014**	0.115**
	(0.016)	(0.032)	(0.006)	(0.048)
Demographics				
PartyID				
Democrat	−0.261***			
	(0.039)			
Not sure	−0.122**			
	(0.060)			
Other	−0.006			
	(0.030)			
Republican	0.171***			
	(0.033)			
Age				
30-44	0.018	0.015	0.009***	0.053*
	(0.011)	(0.023)	(0.003)	(0.029)
45-64	0.009	−0.009	−0.011***	0.039
	(0.012)	(0.023)	(0.002)	(0.029)
Under 30	−0.030*	−0.046	−0.021***	−0.005
	(0.018)	(0.034)	(0.001)	(0.038)
Education				
College Graduate	−0.011	−0.018	0.032***	−0.001
	(0.010)	(0.021)	(0.001)	(0.023)
Post Grad	−0.025	−0.053	0.051***	−0.031
	(0.021)	(0.033)	(0.002)	(0.030)
Gender				
Non-Binary/Fluid	−0.050		0.161***	−0.040
	(0.082)		(0.006)	(0.070)
Prefer not to say	0.056**		0.277***	0.086
	(0.025)		(0.008)	(0.059)

Woman	−0.006	0.008	0.023***	−0.059**
	(0.009)	(0.014)	(0.001)	(0.024)
Race				
American Indian/Native American	−0.046*	−0.042	−0.028***	−0.104*
	(0.028)	(0.030)	(0.001)	(0.062)
Arab, Middle Eastern, or North African	−0.119**	0.037***	−0.001	−0.543***
	(0.056)	(0.005)	(0.001)	(0.014)
Asian American	0.040	0.037***	0.041***	0.178***
	(0.026)	(0.005)	(0.007)	(0.054)
Black or African American	−0.056***	−0.031	−0.019***	−0.105*
	(0.020)	(0.028)	(0.002)	(0.057)
Hispanic or Latino	−0.005	−0.027	−0.023***	0.017
	(0.016)	(0.032)	(0.001)	(0.028)
Native Hawaiian	−0.505***		−0.010***	−0.536***
	(0.006)		(0.004)	(0.014)
Not Hawaiian, but other Pacific Islander	0.055	−0.084	0.714***	−0.097
	(0.084)	(0.053)	(0.007)	(0.070)
Region				
Midwest	−0.001	−0.003	−0.014***	0.021
	(0.015)	(0.028)	(0.002)	(0.038)
South	0.015	0.012	−0.026***	0.052
	(0.013)	(0.021)	(0.002)	(0.037)
West	0.001	−0.009	0.035***	0.025
	(0.013)	(0.023)	(0.001)	(0.039)
Religion				
Catholic	0.003	−0.010	−0.002**	0.082**
	(0.013)	(0.014)	(0.001)	(0.041)
Jewish	−0.053*	−0.056	−0.005***	−0.086
	(0.029)	(0.062)	(0.002)	(0.061)
Muslim	0.058**	0.013	0.089***	0.402***
	(0.028)	(0.020)	(0.005)	(0.022)
No religion	−0.015	−0.006	−0.018***	0.006

	(0.012)	(0.022)	(0.001)	(0.026)
Some other religion	−0.004	−0.002	0.030***	0.002
	(0.017)	(0.021)	(0.005)	(0.039)
Observations	1,871	588	751	449
Note:				*p<0.1; **p<0.05; ***p<0.01

Table B.12: Final regression results.

B.5 Pooling Tests

Variable	Wald Statistic
Abortion	118.85
COVID19	1701.37
ClimateChange	1349.41
Economic Situation: Gotten better	3139.38
Economic Situation: Gotten worse	2934.48
EconomicInequality	788.16
Financial Situation: Better off	1232.64
Financial Situation: Worse off	1686.93
ForeignPolicy	2773.38
GunPolicy	897.08
HealthCare	1989.29
Immigration	1963.12
Inflation	4000.10
RacialandEthnicInequality	1460.28
SupremeCourtAppointments	1140.09
TheEconomy	835.61
ViolentCrime	908.58

Table B.13: We report Wald statistics testing whether party-level interactions are jointly 0 in a pooled model. Large test statistics suggest we can reject the null that coefficients desegregated in the pooled are jointly 0. We instead report the Average Marginal Effects from logit specifications dis-aggregated by party ID.

Appendix C

A REPULSIVE BOUNDED-CONFIDENCE MODEL

C.1 Additional Proofs

Proposition 1. *Let $G = (V, E)$ be a network with 2 nodes , so that $V = \{0, 1\}$. Then the dynamical process described by Equation 3.4 converges, and at time of convergence T,*

$$|x_0(T) - x_1(T)| \leq \max\{c, |x_0(0) - x_1(0)|\}.$$

Proof. Suppose that there are no edges, or $E = \emptyset$. Then the model converges at time $T = 1$, and

$$|x_0(T) - x_1(T)| = |x_0(0) - x_1(0)|.$$

Now suppose that $E = \{(x_0, x_1)\}$, and $|x_0(0) - x_1(0)| \geq c$. Then the update rule will result in no changes, the model converges at $T = 1$

$$|x_0(T) - x_1(T)| = |x_0(0) - x_1(0)|.$$

Now suppose that $|x_0(0) - x_1(0)| < c$. If $A_{01} = -1$, the two nodes repel each other. Then

$$\begin{aligned}
x_1(1) &= x_1(0) + \frac{(x_1(0) - x_1(0)) + c - (x_1(0) - x_0(1))}{2} \\
x_0(1) &= x_0(0) + \frac{(x_0(0) - x_0(0)) - (c - (x_1(0) - x_0(0))}{2} \\
x_1(1) - x_0(1) &= x_1(0) - x_0(0) + \frac{2(c - (x_1(0) - x_0(0)))}{2} \\
&= c
\end{aligned}$$

and $x_1(1) - x_0(1) >= c$, so that after this time, these two nodes will no longer affect each other, and cannot push each other further, so the model has converged, and $\max_{i,j} |x_0(T) - x_1(T)| \leq c$.

If $A_{01} = 1$, the two nodes attract each other, and the model is equivalent to standard Hegselmann–Krause, so that we have convergence to a single point and

$$|x_0(T) - x_1(T)| = 0.$$

This covers all possible cases, and the proposition is proven. □

Lemma 1. *Suppose $i \in V$ a node. Define the following sets:*

$$V_i^+(t) = \left\{ j \in V : A_{ij} = 1 \text{ and } |x_j(t) - x_i(t)| < c \right\}$$
$$U_i(t) = \left\{ j \in V : A_{ij} = -1 \text{ and } \left[(0 < x_j(t) - x_i(t) < c) \text{ or } \left(x_j(t) = x_i(t) \text{ and } j > i \right) \right] \right\}$$
$$L_i(t) = \left\{ j \in V : A_{ij} = -1 \text{ and } \left[(0 < x_i(t) - x_j(t) < c) \text{ or } \left(x_j(t) = x_i(t) \text{ and } i > j \right) \right] \right\}.$$

Then

$$x_i(t+1) = \frac{\sum_{j \in V_i^+(t)} x_j(t) + \sum_{j \in U_i(t)} (x_j(t) - c) + \sum_{j \in L_i} (x_j(t) + c)}{|V_i^+(t)| + |U_i(t) + |L_i(t)|}. \tag{3.5}$$

Proof. From Equation 3.4, we rearrange

$$\begin{aligned}
x_i(t+1) &= x_i(t) + \frac{\sum_{j \in V} A_{ij} M_{ij}(t) \mathbf{1}_{|x_j(t) - x_i(t)| < c}}{\sum_{j \in V} |A_{ij}| \mathbf{1}_{|x_j(t) - x_i(t)| < c}} \\
&= x_i(t) + \frac{\sum_{j \in V_i^+(t) \cup U_i(t) \cup L_i(t)} A_{ij} M_{ij}(t)}{\sum_{j \in V_i^+(t) \cup U_i(t) \cup L_i(t)} |A_{ij}|} \\
&= x_i(t) + \frac{\sum_{j \in V_i^+(t)} (x_j(t) - x_i(t))}{|V_i^+(t)| + |U_i(t)| + |L_i(t)|} \\
&\quad + \frac{\sum_{j \in U_i(t)} (-1)(1)(c - (x_j(t) - x_i(t)))}{|V_i^+(t)| + |U_i(t)| + |L_i(t)|} \\
&\quad + \frac{\sum_{j \in L_i(t)} (-1)(-1)(c - (x_i(t) - x_j(t)))}{|V_i^+(t)| + |U_i(t)| + |L_i(t)|} \\
&= \frac{\sum_{j \in V_i^+(t)} x_j(t) + \sum_{j \in U_i(t)} (x_j(t) - c) + \sum_{j \in L_i} (x_j(t) + c)}{|V_i^+(t)| + |U_i(t) + |L_i(t)|}.
\end{aligned}$$

□

Lemma 2. *Let $i \in V$ at time t, and let $W(t) \subset V$ be a set of nodes such that $W(t)$ is completely contained in $V_i^+(t) \cup U_i(t) \cup L_i(t)$. Define*

$$\overline{W}(t) = \frac{\sum_{j \in W(t)} x_j(t)}{|W(t)|}$$

to be the average of $x_j(t)$ for all $j \in W(t)$. Then we can rewrite Equation 3.5 *as*

$$x_i(t+1) = \frac{\sum_{j \in \left(V_i^+(t) \cup U_i(t) \cup L_i(t) \right) \setminus W(t)} x_j(t) + \sum_{j \in W(t)} \overline{W}(t) + (|L_i(t)| - |U_i(t)|)\, c}{|V_i^+(t)| + |U_i(t) + |L_i(t)|}.$$

Proof. We rearrange Equation 3.5 as follows:

$$\begin{aligned} x_i(t+1) &= \frac{\sum_{j\in V_i^+(t)} x_j(t) + \sum_{j\in U_i(t)}(x_j(t)-c) + \sum_{j\in L_i}(x_j(t)+c)}{|V_i^+(t)| + |U_i(t) + |L_i(t)|} \\ &= \frac{\sum_{j\in\left(V_i^+(t)\cup U_i(t)\cup L_i(t)\right)} x_j(t) + (|L_i(t)| - |U_i(t)|)\, c}{|V_i^+(t)| + |U_i(t) + |L_i(t)|} \\ &= \frac{\sum_{j\in\left(V_i^+(t)\cup U_i(t)\cup L_i(t)\right)\setminus W(t)} x_j(t) + \sum_{j\in W(t)} \overline{W}(t) + (|L_i(t)| - |U_i(t)|)\, c}{|V_i^+(t)| + |U_i(t) + |L_i(t)|}. \end{aligned}$$

□

Lemma 3. *Let $G = (V, E)$ be a network with n nodes and m edges with confidence bound c. Suppose that every edge in G is repulsive. At time t, suppose $x_i(t) > x_j(t)$ for all other nodes j, so that i is the node with the highest opinion value at time t. Then $x_i(t+1) > x_j(t+1)$ for all j.*

Proof. Note that $V_k^+(t) = \{x_k\}$ for all k, t, since every edge in G is repulsive. For convenience we define the following sets:

$$\begin{aligned} U_{ij}(t) &= U_j(t) \bigcap U_i(t) \\ U'_{ij}(t) &= U_i(t) \setminus U_{ij}(t) \\ L_{ij}(t) &= L_i(t) \bigcap L_j(t) \\ L'_{ji}(t) &= L_j(t) \setminus L_{ij}(t) \\ W_{ij}(t) &= L_i(t) \bigcap U_j(t). \end{aligned}$$

Unpacking this notation, $U_{ij}(t)$ consists of all nodes that repel both i and j downward, while $L_{ij}(t)$ consists of all nodes that repel both i and j upward. $U'_{ij}(t)$ consists of nodes which repel i downward, but not j (note that if $x_i(t) < x_j(t)$, this is automatically empty), while $L'_{ji}(t)$ consists of nodes which repel j upward, but not i (again, if $x_i(t) < x_j(t)$, this is empty). Finally, $W_{ij}(t)$ consists of nodes which repel i upward and j downward (empty if $x_j(t) > x_i(t)$).

Now, suppose $x_i(t) > x_j(t)$ for all $j \in V$. Then we can write

$$\begin{aligned} V_i^+(t) &= \{i\} \\ U_i(t) &= \emptyset \\ L_i(t) &= L_{ij}(t) \bigcup W_{ij}(t) \bigcup \{j\} \end{aligned}$$

and

$$
\begin{aligned}
V_j^+(t) &= \{j\} \\
U_i(t) &= \{i\} \bigcup W_{ij}(t) \\
L_i(t) &= L_{ij}(t) \bigcup L'_{ji}(t).
\end{aligned}
$$

Then we observe the following from the knowledge that nodes only effect each other if they are within confidence of each other.

$$
\begin{aligned}
x_j(t) + c &> x_i(t) \\
\overline{L_{ij}}(t) + c &> x_i(t) \\
\overline{W_{ij}}(t) + c &> x_i(t) \\
x_i(t) &> \overline{L'_{ji}} + c
\end{aligned}
$$

Then applying Lemma 1 and Lemma 2:

$$
\begin{aligned}
&x_i(t+1) \\
&= \frac{x_i(t) + (x_j(t) + c) + |L_{ij}(t)|(\overline{L_{ij}}(t) + c) + |W_{ij}(t)|(\overline{W_{ij}}(t) + c)}{2 + |L_{ij}(t)| + |W_{ij}(t)|} \\
&> \frac{x_i(t) + (x_j(t) + c) + |L_{ij}(t)|(\overline{L_{ij}}(t) + c) + |W_{ij}(t)|(\overline{W_{ij}}(t) + c) + |L'_{ji}(t)|(\overline{L'_{ji}}(t) + c)}{2 + |L_{ij}(t)| + |W_{ij}(t)| + |L'_{ji}(t)|} \qquad \text{(C.1)} \\
&> \frac{(x_i(t) - c) + x_j(t) + |L_{ij}(t)|(\overline{L_{ij}}(t) + c) + |W_{ij}(t)|(\overline{W_{ij}}(t) - c) + |L'_{ji}(t)|(\overline{L'_{ji}}(t) + c)}{2 + |L_{ij}(t)| + |W_{ij}(t)| + |L'_{ji}(t)|} \\
&= x_j(t+1)
\end{aligned}
$$

where the inequality in Equation C.1 follows because $x_i(t+1)$ is a weighted average, and $\overline{L'_{ji}}(t)$ is less than all of the other values being averaged in the previous line. The next inequality follows straightforwardly by replacing each value in the average with a smaller or equal value.

So if $x_i(t)$ has the highest value opinion at time t, it will always have the highest value opinion. □

Corollary 1. *Let $G = (V, E)$ be a network with n nodes and m edges with confidence bound c. Suppose that every edge in G is repulsive. At time t, let $M = \{i : x_i(t) \geq x_j(t) \forall j \in V\}$. Then $x_{\max_M i}(t+1) > x_j(t+1) \forall j \in V$.*

Proof. From the definitions of $U_i(t), L_i(t)$, we can observe that the member of M with highest index will have the largest corresponding set $L_i(t)$ and the smallest corresponding $U_i(t)$, so that at time $t+1$, that member of M will have the highest-valued opinion of all members of M. By the same logic as in the proof of Lemma 3, that opinion will also be the highest-valued opinion overall. □

Corollary 2. *Let $G=(V,E)$ be the complete network with n nodes with confidence bound c. Suppose that every edge in G is repulsive. At time t, let $M=\{i : x_i(t) \leq x_j(t) \forall j \in V\}$. Then $x_{\min_M i}(t+1) < x_j(t+1) \forall j \in V$.*

Proof. Proves that $x_i(t) < x_j(t)$ for all $j \in V$, then $x_i(t+1) < x_j(t+1)$ for all $j \in V$, by segmenting $U_i(t), L_i(t), U_j(t), L_j(t)$ into the appropriate subsets and reversing inequalities as needed as in lemma 3. Then the same logic as in corollary 1 proves the statement. □

Lemma 4. *Let $G=(V,E)$ be a network with n nodes and m edges with confidence bound c. Suppose that every edge in G is repulsive. At time t, suppose $x_i(t) > x_j(t)$ for all other nodes $j \in V$, so that i is the node with the highest-valued opinion at time t. Suppose that there is some node j such that $x_i(t) - x_j(t) < c$, and that j has the highest-valued opinion of all such nodes. Then*

$$\frac{2c}{2+|L_{ij}(t)|+|L'_{ji}(t)|} \leq x_i(t+1) - x_j(t+1) \leq \frac{(|L'_{ji}(t)|+2)c}{2+|L_{ij}(t)|+|L'_{ji}(t)|}.$$

Proof. By assumption, since j has the highest-valued opinion of all nodes within confidence of i, $W_{ij}(t)$ is empty. To prove the lower bound,

$$\begin{aligned}
x_i(t+1) &= \frac{x_i(t) + (x_j(t)+c) + |L_{ij}(t)|(\overline{L_{ij}}(t)+c)}{2+|L_{ij}(t)|} \\
&\geq \frac{x_i(t) + (x_j(t)+c) + |L_{ij}(t)|(\overline{L_{ij}}(t)+c) + |L'_{ji}(t)|(\overline{L'_{ji}}(t)+c)}{2+|L_{ij}(t)|+|L'_{ji}(t)|} \\
x_j(t+1) &= \frac{(x_i(t)-c) + x_j(t) + |L_{ij}(t)|(\overline{L_{ij}}(t)+c) + |L'_{ji}(t)|(\overline{L'_{ji}}(t)+c)}{2+|L_{ij}(t)|+|L'_{ji}(t)|} \\
x_i(t+1) - x_j(t+1) &\geq \frac{2c}{2+|L_{ij}(t)|+|L'_{ji}(t)|}.
\end{aligned}$$

To prove the upper bound,

$$
\begin{aligned}
x_i(t+1) &= \frac{x_i(t) + (x_j(t)+c) + |L_{ij}(t)|(\overline{L_{ij}}(t)+c)+}{2+|L_{ij}(t)|} \\
&\leq \frac{x_i(t) + (x_j(t)+c) + |L_{ij}(t)|(\overline{L_{ij}}(t)+c) + |L'_{ji}(t)|(x_j(t)+c)}{2+|L_{ij}(t)|+|L'_{ji}(t)|} \\
x_j(t+1) &= \frac{(x_i(t)-c) + x_j(t) + |L_{ij}(t)|(\overline{L_{ij}}(t)+c) + |L'_{ji}(t)|(\overline{L'_{ji}}(t)+c)}{2+|L_{ij}(t)|+|L'_{ji}(t)|} \\
x_i(t+1) - x_j(t+1) &\leq \frac{c + c + |L'_{ji}(t)|(x_j(t) - \overline{L'_{ji}}(t))}{2+|L_{ij}(t)|+|L'_{ji}(t)|} \\
&\leq \frac{(|L'_{ji}(t)|+2)c}{2+|L_{ij}(t)|+|L'_{ji}(t)|}.
\end{aligned}
$$

Notice that if $L'_{ji}(t)$ is empty, both inequalities become equalities, so that

$$x_i(t+1) - x_j(t+1) = \frac{2c}{2+|L_{ij}(t)|}$$

Notice also that if both $L_{ij}(t)$ and $L'_{ji}(t)$ are empty, that the distance between $x_i(t+1) - x_j(t+1)$ is precisely c. □

Corollary 3. *Let $G = (V, E)$ be a network with n nodes and m edges with confidence bound c. Suppose that every edge in G is repulsive. At time t, suppose $x_i(t) < x_j(t)$ for all other nodes $j \in V$, so that i is the node with the lowest-valued opinion at time t. Suppose that there is some node j such that $x_j(t) - x_i(t) < c$, and that j has the lowest-valued opinion of all such nodes. Then*

$$\frac{2c}{2+|U_{ij}(t)|+|U'_{ij}(t)|} \leq x_i(t+1) - x_j(t+1) \leq \frac{(|U'_{ij}(t)|+2)c}{2+|U_{ij}(t)|+|U'_{ij}(t)|}.$$

Lemma 5. *Let $G = (V, E)$ be the complete network with n nodes and confidence bound c. Suppose that every edge in G is repulsive. At time t, suppose that i and j are nodes such that $(i, j) \in E$, $x_i(t) > x_j(t)$, and $x_i(t) - x_j(t) < c$, and there exist no nodes k connected to i or j such that $x_i(t) > x_k(t) > x_j(t)$. Then*

$$|x_i(t+1) - x_j(t+1)| \leq c.$$

Proof. By the assumption that no nodes have values between $x_i(t)$ and $x_j(t)$, we have that $W_{ij}(t) = W_{ji}(t) = \emptyset$. Then to prove one direction of the bound,

$$
\begin{aligned}
x_i(t+1) &= \frac{|U'_{ij}(t)|(\overline{U'_{ij}}(t)-c)+|U_{ij}(t)|(\overline{U_{ij}}(t)-c)+x_i(t)+(x_j(t)+c)+|L_{ij}(t)|(\overline{L_{ij}}(t)+c)}{2+|U'_{ij}(t)|+|U_{ij}(t)|+|L_{ij}(t)|} \\
&\le \frac{|U'_{ij}(t)|(\overline{U'_{ij}}(t)-c)+|U_{ij}(t)|(\overline{U_{ij}}(t)-c)+x_i(t)+(x_j(t)+c)}{2+|U'_{ij}(t)|+|U_{ij}(t)|+|L_{ij}(t)|+|L'_{ji}(t)|}+ \\
&\quad \frac{|L_{ij}(t)|(\overline{L_{ij}}(t)+c)+|L'_{ji}(t)|(x_j(t)+c)}{2+|U'_{ij}(t)|+|U_{ij}(t)|+|L_{ij}(t)|+|L'_{ji}(t)|} \\
x_j(t+1) &= \frac{|U_{ij}(t)|(\overline{U_{ij}}(t)-c)+(x_i(t)-c)+x_j(t)+|L_{ij}(t)|(\overline{L_{ij}}(t)+c)+|L'_{ji}(t)|(\overline{L'_{ji}}(t)+c)}{2+|U_{ij}(t)|+|L_{ij}(t)|+|L'_{ji}(t)|} \\
&\ge \frac{|U'_{ij}(t)|(x_i(t)-c)+|U_{ij}(t)|(\overline{U_{ij}}(t)-c)+(x_i(t)-c)+x_j(t)}{2+|U'_{ij}(t)|+|U_{ij}(t)|+|L_{ij}(t)|+|L'_{ji}(t)|}+ \\
&\quad \frac{|L_{ij}(t)|(\overline{L_{ij}}(t)+c)+|L'_{ji}(t)|(\overline{L'_{ji}}(t)+c)}{2+|U'_{ij}(t)|+|U_{ij}(t)|+|L_{ij}(t)|+|L'_{ji}(t)|}.
\end{aligned}
$$

Combining both equations,

$$
\begin{aligned}
x_i(t+1)-x_j(t+1) &\le \frac{|U'_{ij}(t)|\left(\overline{U'_{ij}}(t)-x_i(t)\right)+c+c+|L'_{ji}(t)|\left(x_j(t)-\overline{L'_{ji}}(t)\right)}{2+|U'_{ij}(t)|+|U_{ij}(t)|+|L_{ij}(t)|+|L'_{ji}(t)|} \\
&\le \frac{\left(2+|U'_{ij}(t)|+|L'_{ji}(t)|\right)c}{2+|U'_{ij}(t)|+|U_{ij}(t)|+|L_{ij}(t)|+|L'_{ji}(t)|} \\
&\le c.
\end{aligned}
$$

To prove the other direction,

$$
\begin{aligned}
x_j(t+1) &= \frac{|U_{ij}(t)|(\overline{U_{ij}}(t)-c)+(x_i(t)-c)+x_j(t)+|L_{ij}(t)|(\overline{L_{ij}}(t)+c)+|L'_{ji}(t)|(\overline{L'_{ji}}(t)+c)}{2+|U_{ij}(t)|+|L_{ij}(t)|+|L'_{ji}(t)|} \\
&\le \frac{|U'_{ij}(t)|(\overline{L_{ij}}(t)+c)+|U_{ij}(t)|(\overline{U_{ij}}(t)-c)+(x_i(t)-c)+x_j(t)}{2+|U'_{ij}(t)|+|U_{ij}(t)|+|L_{ij}(t)|+|L'_{ji}(t)|}+ \\
&\quad \frac{|L_{ij}(t)|(\overline{L_{ij}}(t)+c)+|L'_{ji}(t)|(\overline{L'_{ji}}(t)+c)}{2+|U'_{ij}(t)|+|U_{ij}(t)|+|L_{ij}(t)|+|L'_{ji}(t)|} \\
x_i(t+1) &= \frac{|U'_{ij}(t)|(\overline{U'_{ij}}(t)-c)+|U_{ij}(t)|(\overline{U_{ij}}(t)-c)+x_i(t)+(x_j(t)+c)+|L_{ij}(t)|(\overline{L_{ij}}(t)+c)}{2+|U'_{ij}(t)|+|U_{ij}(t)|+|L_{ij}(t)|} \\
&\ge \frac{|U'_{ij}(t)|(\overline{U'_{ij}}(t)-c)+|U_{ij}(t)|(\overline{U_{ij}}(t)-c)+x_i(t)+(x_j(t)+c)}{2+|U'_{ij}(t)|+|U_{ij}(t)|+|L_{ij}(t)|+|L'_{ji}(t)|}+ \\
&\quad \frac{|L_{ij}(t)|(\overline{L_{ij}}(t)+c)+|L'_{ji}(t)|(\overline{U_{ij}}(t)-c)}{2+|U'_{ij}(t)|+|U_{ij}(t)|+|L_{ij}(t)|+|L'_{ji}(t)|}.
\end{aligned}
$$

Combining both inequalities yields

$$
x_j(t+1)-x_i(t+1) \le \frac{|U'_{ij}(t)|\left(\overline{L_{ij}}(t)-\overline{U'_{ij}}(t)+2c\right)+c+c+|L'_{ji}(t)|\left(\overline{L'_{ji}}(t)-\overline{U_{ij}}(t)+2c\right)}{2+|U'_{ij}(t)|+|U_{ij}(t)|+|L_{ij}(t)|+|L'_{ji}(t)|}.
$$

Note, however, that

$$\begin{aligned} 2c &= (x_i(t) + c) - (x_i(t) - c) \\ &> \overline{U'_{ij}}(t) - \overline{L_{ij}}(t) \\ &> (x_j(t) + c) - x_j(t) = c \end{aligned}$$

and similarly $c < \overline{U_{ij}}(t) - \overline{L'_{ji}}(t) < 2c$ so that we have

$$\begin{aligned} x_j(t+1) - x_i(t+1) &\leq \frac{|U'_{ij}(t)| \left(\overline{L_{ij}}(t) - \overline{U'_{ij}}(t) + 2c\right) + c + c + |L'_{ji}(t)| \left(\overline{L'_{ji}}(t) - \overline{U_{ij}}(t) + 2c\right)}{2 + |U'_{ij}(t)| + |U_{ij}(t)| + |L_{ij}(t)| + |L'_{ji}(t)|} \\ &\leq \frac{\left(2 + |U'_{ij}(t)| + |L'_{ji}(t)|\right) c}{2 + |U'_{ij}(t)| + |U_{ij}(t)| + |L_{ij}(t)| + |L'_{ji}(t)|} \\ &\leq c \end{aligned}$$

and the proof is finished. $\square$

Appendix D

INCORPORATING LATENT CLASS IDENTITIES

D.1 ANES Details

All of the variables came from either (1) direct questions asked in the ANES, (2) summary questions reported by the ANES, or (3) summaries based on questions in the ANES. We eliminated respondents who had incomplete answers to any of the variables needed for either analysis. For each subject we list out how many respondents had each type of inapplicable response. In total, we are left with 5,662 of the original 8,280 respondents or 68%.

Race—determined by ANES combination outcome variable:

V201549x:

-9. Refused

-8. Don't know

1. White, non-Hispanic

2. Black, non-Hispanic

3. Hispanic

4. Asian or Native Hawaiian/other Pacific Islander, non-Hispanic alone

5. Native American/Alaska Native or other race, non-Hispanic alone

6. Multiple races, non-Hispanic

Individuals who responded with Refused (-9) or Don't know (-8) were excluded from the analysis. This includes a total of 96 (1.16%) and 6 (0.07%) respondents, respectively.

Gender—explicitly asked in the survey:

V201600: What is your sex?

-9. Refused

1. Male
2. Female

Individuals who responded with Refused (-9) were excluded from the analysis. They made up a total of 67 (0.8%) respondents.

Age—determined by ANES combination outcome variable:

V201507x:

-9. Refused

80. 80 or older

Individuals who responded with Refused (-9) were excluded from the analysis. They made up a total of 348 (4.2%) respondents.

Party ID—determined by ANES combination outcome variable:

V201231x:

-9. Refused

-8. Don't know

1. Strong Democrat
2. Not very strong Democrat
3. Independent-Democrat
4. Independent
5. Independent-Republican
6. Not very strong Republican
7. Strong Republican

Individuals who responded with Refused (-9) or Don't know (-8) were excluded from the analysis. This includes a total of 31 (0.4%) and 4 (0.05%) respondents respectively.

College—generated from a question

V201510: What is the highest level of school you have completed or the highest degree you have received?

-9. Refused

-8. Don't know

1. Less than high school credential

2. High school graduate - High school diploma or equivalent (e.g., GED)

3. Some college but no degree

4. Associate degree in college–occupational/vocational

5. Associate degree in college–academic

6. Bachelor's degree (e.g., BA, AB, BS)

7. Master's degree (e.g., MA, MS, MEng, MEd, MSW, MBA)

8. Professional school degree (e.g., MD, DDS, DVM, LLB, JD)/Doctoral degree (e.g. PHD, EDD)

95. Other SPECIFY

Individuals who answered 1-5 were labeled as not having attended college (4502, 54.4%) and individuals who answered 6-8 were labeled as having attended college (3647, 44%). Individuals who responded with Refused (-9), Don't know (-8), or Other (95) were excluded from the analysis. This includes a total of 33 (0.4%), 1 (0.01%), and 97 (1.2%) respondents, respectively.

Income—determined by ANES combination outcome variable:

V201617x: Please choose the answer that includes the income of all members of your family during the past 12 months before taxes.

-9. Refused

-5. Interview breakoff (sufficient partial IW)

1. Under $9,999

2. $10,000-14,999

3. $15,000-19,999	10. $50,000-59,999	17. $100,000-109,999
4. $20,000-24,999	11. $60,000-64,999	18. $110,000-124,999
5. $25,000-29,999	12. $65,000-69,999	19. $125,000-149,999
6. $30,000-34,999	13. $70,000-74,999	20. $150,000-174,999
7. $35,000-39,999	14. $75,000-79,999	21. $175,000-249,999
8. $40,000-44,999	15. $80,000-89,999	22. $250,000 or more
9. $45,000-49,999	16. $90,000-99,999	

Individuals who responded with Refused (-9) or Interview breakoff (-5) were excluded from the analysis. This includes a total of 584 (7%) and 32 (0.4%) respondents, respectively.

Student loans—explicitly asked in the survey:

V202562: Do you currently owe money on student loans, or not?

- -9. Refused
- -7. No post-election data, deleted due to incomplete interview
- -6. No post-election interview
- -5. Interview breakoff (sufficient partial IW)
- 1. Yes
- 2. No

Individuals who responded with Refused (-9), No post-election data (-7), No post-election interview(-6) or Interview breakoff (-5) were excluded from the analysis. This includes a total of 16 (0.2%), 77 (0.9%), 754 (9%), and 103 (1.3%) respondents, respectively.

Employment Status—determined by ANES combination outcome variable:

V201534x:

- -2. Refused/Don't know/Inapplicable

1. R working now (if also retired, disabled, homemaker or student, working 20 or more hrs/wk)

2. R temporarily laid off

4. R unemployed

5. R retired (if also working, working <20 hrs/wk)

6. R permanently disabled (if also working, working <20 hrs/wk)

7. R homemaker (if also working, working <20 hrs/wk/incl nonworkg rs both homemaker and student)

8. R student (if also working, working <20 hrs/wk)

Individuals who responded with Refused/Don't know/Inapplicable (-2) were excluded from the analysis. This includes a total of 57 (0.7%) of respondents.

Socioeconomic Class—explicitly asked in the survey:

V202352: How would you describe your social class? Are you in the lower class, the working class, the middle class, or the upper class?

-9. Refused

-8. Don't know

-7. No post-election data, deleted due to incomplete interview

-6. No post-election interview

-5. Interview breakoff (sufficient partial IW)

1. Lower class

2. Working class

3. Middle class

4. Upper class//

Individuals who responded with Refused (-9), Don't know (-8), No post-election data (-7), No post-election interview(-6) or (-5) Interview breakoff were excluded from the analysis. This includes a total of 25 (0.3%), 2 (0.02%), 77 (0.9%), 754

(9%) and 53 (0.6%) of respondents, respectively.

Occupation—explicitly asked in the survey:

V201529: Which one of the following best describes your employment?

-9. Refused

-1. Inapplicable

1. For-profit company or organization

2. Non-profit organization (including tax-exempt and charitable organizations)

3. Local government (for example: city or county school district)

4. State government (including state colleges/universities)

5. Active duty U.S. Armed Forces or Commissioned Corps

6. Federal government civilian employee

7. Owner of non-incorporated business, professional practice, or farm

8. Owner of incorporated business, professional practice, or farm

9. Worked without pay in a for-profit family business or farm for 15 hours or more per week

Individuals who responded with Refused (-9) or Inapplicable (-1) were excluded from the analysis. This includes a total of 181 (2%) and 234 (2.8%) of respondents, respectively.

Union Affiliation—explicitly asked in the survey:

V201544: Do you or anyone else in this household belong to a labor union or to an employee association similar to a union?

-9. Refused

-8. Don't know

1. Yes

2. No

Individuals who responded with Refused (-9) or Don't know (-8) were excluded from the analysis. This includes a total of 39 (0.5%) and 4 (0.05%) of respondents, respectively.

Stock market investor—explicitly asked in the survey:

V201606: Do you personally, or jointly with a spouse, have any money invested in the stock market right now—either in an individual stock or in a mutual fund?

-9. Refused

-8. Don't know

-5. Interview breakoff (sufficient partial IW)

1. Yes

2. No

Individuals who responded with Refused (-9), Don't know (-8) or Interview breakoff (-5) were excluded from the analysis. This includes a total of 179 (2%), 641(7.7%) and 11 (0.1%) of respondents, respectively.

Outcome variables—we looked at two outcome variables. We differentiate them throughout the paper by refering to them as the two "subject matters." The options are ICE and BLM. Both come from explicit questions in the survey with the same outcome structure. The questions are:

ICE—*V202182*: How would you rate: The Immigration and Customs Enforcement (ICE) agency

BLM—*V202174*: How would you rate: Black Lives Matter movement

The outcome options are a scale from 0-100 or:

-9. Refused

-7. No post-election data, deleted due to incomplete interview

-6. No post-election interview

-5. Interview breakoff (sufficient partial IW)

-4. Technical error

998. Don't know

999. Don't recognize

All respondents who responded outside of 0-100 for either scale were removed from both analysis. The table below notes the number and percentage of each type of response for each subject.

	ICE		BLM	
Responses	Number	%	Number	%
-4	1	0.01	1	0.01
-5	16	0.19	14	0.17
-6	754	9.11	754	9.11
-7	77	0.93	77	0.93
-9	82	0.99	86	1.04
998	4	0.05	2	0.02
999	281	3.39	2	0.02
0-100	7065	85.33	7344	88.70

Table D.1: Number and percent of respondents who responded in each of the eliminated category or were kept in the analysis (0-100).

D.2 Stan Implementation

We run all analysis in Stan using the BRMS frontend in R (Bürkner, 2017, 2018). In order to generate the prior for the class specification we first run K-Modes on the class relevant covariates using the ‘klaR‘ package in R (Weihs et al., 2005). The income level is changed to the numerical value of the lower end of the bin. Income_norm refers to that value divided by 10,000. This is done in order to place it on a more appropriate scale when compared to the rest of the covariates. Given the dataframe “anes_2020,“ the original split is done as:

```
anes_2020 <- anes_2020 %>%
  mutate(group_class = kmodes(as.matrix(anes_2020 %>%
  dplyr::select(college,
                income,
                loans,
                employ,
```

```
            Class,
            occupation,
            union,
            investor)),
          2)$cluster - 1)
```

The class priors are then found running a logisitic model on the group_class variable and the same covariates:

```
class_fit <- brm(bf(group_class ~ 1 +
                                   college +
                                   income_norm +
                                   loans +
                                   employ +
                                   Class +
                                   occupation +
                                   union +
                                   investor),
                family = bernoulli(link = "logit"),
                anes_2020 ,
                control = list(adapt_delta = 0.92,
                                  max_treedepth = 12),
                warmup = 4000,
                iter = 5000,
                seed = 1234,
                chains = 4,
                cores = 4)
```

In Figure D.1 the histogram for estimated group probabilities from the logits for respondents of each class from the original clustering can be seen. It is clear that the logit does a good job of predicting the class generated in the clustering. The estimated output for each coefficient can be seen in Table D.2 under the columns Mean and Estimated SD.

We then run the un-mixed regression for each output variable split into groups when the estimated class is less than 50% or above 50%. For example, for the ICE model and the class represented by 0 we would have the command:

```
ICE_group_0 <- brm(bf(therm_ice ~ 1 +
                                   age +
```

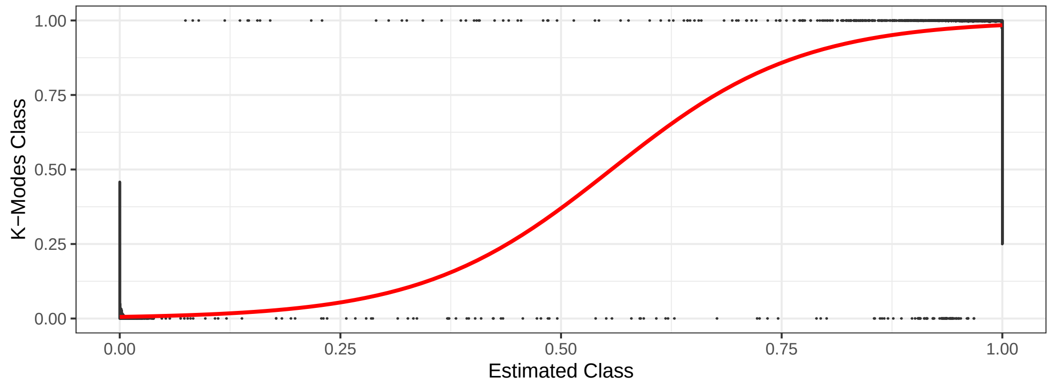


Figure D.1: Comparison of clustering from k-modes to the estimated group probability from the logit.

```
                    pid7 +
                    (1 + age||intersectional)),
          anes_2020 %>% filter(Estimate <= 0.5),
          seed = 1234,
          control = list(adapt_delta = 0.95,
                         max_treedepth = 12),
          warmup = 3000,
          iter = 4000,
          prior = c(set_prior("normal(0,50)",
                              class = "sd"),
                    set_prior("normal(0,50)")),
          chains = 4,
          cores = 4)
```

All of the solving parameters are consistent for each of the four regressions. The estimated means and standard deviations can be seen in Table D.3. Once these values have been found, we use the estimates to set the prior and increase the prior standard deviations in order to give the model more freedom. We set all priors to be normal. The R code for each of the final regressions can be seen below.

Parameter	Mean	Estimated SD
Intercept	1.2	0.4
College		
Yes	9.0	0.5
Income Norm	0.2	0.0
Student Loans		
No	-0.1	0.3
Employment Status		
Temporarilylaidoff	-0.1	0.5
unemployed	0.7	0.9
retired	0.0	0.3
permanentlydisabled	1.0	0.7
homemaker	0.8	0.6
student	1.6	1.2
Class		
WorkingClass	-6.5	0.5
MiddleClass	2.3	0.3
UpperClass	-1.1	0.5
Occupation		
Non-profit	-0.4	0.3
LocalGov	0.3	0.4
StateGov	0.2	0.5
ArmedForces	-1.2	0.9
FedGovCivilian	1.0	0.8
OwnernonMincorporated	0.1	0.4
Ownerincorporated	0.9	0.5
Workforfamily	0.9	0.9
Union Affiliation		
No	-0.1	0.3
Investor in Stock Market		
No	-9.3	0.5

Table D.2: Estimated mean and standard deviation for logit of k-mode class on class covariates as well as the standard deviation for the priors given to the final response.

	ICE				BLM			
	Estimate $\geq$ 50%		Estimate $\leq$ 50%		Estimate $\geq$ 50%		Estimate $\leq$ 50%	
	mean	sd	mean	sd	mean	sd	mean	sd
Fixed Effects								
Intercept	30.33	1.81	29.01	2.56	61.27	2.56	64.49	2.68
age	0.38	0.04	0.45	0.04	-0.16	0.04	-0.19	0.04
pid7	7.67	0.18	6.57	0.24	-10.74	0.18	-9.47	0.26
sigma	22.92	0.28	25.36	0.36	23.05	0.29	27.26	0.40
Random Effects (sd)								
Intercept	2.42	1.57	5.40	2.14	5.75	2.23	5.52	1.98
age	0.04	0.03	0.06	0.04	0.06	0.05	0.05	0.04

Table D.3: Estimates from models when artificially split into two groups.

```
priors_class <- c(
  prior(normal(  1.19,   2), Intercept                   , dpar = theta1),
  prior(normal(  9.01,   2), b, coef =  collegeYes        , dpar = theta1),
  prior(normal(  0.21,    0.12), b, coef =  income_norm   , dpar = theta1),
  prior(normal( -0.05,   1), b, coef =  loansNo           , dpar = theta1),
  prior(normal( -0.12,2.4), b, coef =  employTemporarilylaidoff,
  dpar = theta1),
  prior(normal(  0.70,3.6), b, coef =  employunemployed   , dpar = theta1),
  prior(normal(  0.02,1.2), b, coef =  employretired      , dpar = theta1),
  prior(normal(  0.99,2.8), b, coef =  employpermanentlydisabled,
  dpar = theta1),
  prior(normal(  0.80,   2), b, coef =  employhomemaker    , dpar = theta1),
  prior(normal(  1.57,   4), b, coef =  employstudent      , dpar = theta1),
  prior(normal( -6.55,   2), b, coef =  ClassWorkingClass  , dpar = theta1),
  prior(normal(  2.33,1.2), b, coef =  ClassMiddleClass   , dpar = theta1),
  prior(normal( -1.10,   2), b, coef =  ClassUpperClass    , dpar = theta1),
  prior(normal( -0.39,1.2), b, coef =  occupationNonMprofit, dpar = theta1),
  prior(normal(  0.29,1.6), b, coef =  occupationLocalGov, dpar = theta1),
  prior(normal(  0.16,   2), b, coef =  occupationStateGov, dpar = theta1),
  prior(normal( -1.20,   4), b, coef =  occupationArmedForces, dpar = theta1),
  prior(normal(  1.04,2.8), b, coef =  occupationFedGovCivilian,
  dpar = theta1),
  prior(normal(  0.09,1.6), b, coef =  occupationOwnernonMincorporated,
  dpar = theta1),
  prior(normal(  0.87,   2), b, coef =  occupationOwnerincorporated,
  dpar = theta1),
  prior(normal(  0.93,3.6), b, coef =  occupationWorkforfamily,
  dpar = theta1),
  prior(normal( -0.12,1.2), b, coef =  unionNo, dpar = theta1),
  prior(normal( -9.37,   2), b, coef =  investorNo, dpar = theta1))

priors_ICE <- c(
  priors_class,
  prior(normal(  2.46,    3), sd , group = intersectional, coef = Intercept,
  dpar = mu1),
  prior(normal(  0.04,   0.06), sd , group = intersectional, coef = age,
  dpar = mu1),
  prior(normal( 30.35,   4), Intercept, dpar = mu1),
  prior(normal(  0.38,   0.08), b , coef = age, dpar = mu1),
  prior(normal(  7.68,   0.4), b , coef = pid7, dpar = mu1),
  prior(normal( 22.92,   0.6),  sigma1),
  prior(normal(  5.49,   5 ), sd ,group = intersectional,
  coef = Intercept, dpar = mu2),
  prior(normal(  0.06,   0.08), sd , group = intersectional,
```

```
  coef = age, dpar = mu2),
  prior(normal( 29.15,  5 ), Intercept, dpar = mu2),
  prior(normal(  0.45,   0.1), b , coef = age, dpar = mu2),
  prior(normal(  6.58,0.5), b , coef = pid7, dpar = mu2),
  prior(normal( 25.36,0.7),  sigma2))

priors_BLM <- c(
  prios_class,
  prior(normal(  5.73,  5), sd ,group = intersectional,
  coef = Intercept, dpar = mu1),
  prior(normal(  0.06,  0.08 ), sd , group = intersectional,
  coef = age, dpar = mu1),
  prior(normal( 61.24,   5), Intercept    , dpar = mu1),
  prior(normal(  -0.16, 0.08), b , coef = age, dpar = mu1),
  prior(normal( -10.74,   0.4), b , coef = pid7, dpar = mu1),
  prior(normal( 23.05,   0.6),  sigma1),
  prior(normal(  5.73, 4), sd ,group = intersectional,
  coef = Intercept, dpar = mu2),
  prior(normal(  0.05,     0.1), sd , group = intersectional,
  coef = age, dpar = mu2),
  prior(normal( 64.45,     6), Intercept, dpar = mu2),
  prior(normal(  -0.19,   0.1), b , coef = age, dpar = mu2),
  prior(normal( -9.48,    0.5), b , coef = pid7, dpar = mu2),
  prior(normal(  27.26,   0.8 ),  sigma2))
```

We finally are able to run the two main regressions:

```
out_ICE <- brm(bf( therm_ice ~ 1 +
                               age +
                               pid7 +
                               (1 + age||intersectional),
               theta1 ~ 1 +
               college +
               income_norm +
               loans +
               employ  +
               Class +
               occupation +
               union +
               investor),
           family = mixture(gaussian(link = "identity"),
                            gaussian(link = "identity")),
```

```
        anes_2020,
        seed = 1234,
        control = list(adapt_delta = 0.95,
                       max_treedepth = 12),
        warmup = 4000,
        iter = 5000,
        chains = 4,
        cores = 4,
        prior = priors_ICE)

out_BLM <- brm(bf( therm_BLM ~ 1 +
age +
pid7 +
(1 + age||intersectional),
                theta1 ~ 1 +
                college +
                income_norm +
                loans +
                employ  +
                Class +
                occupation +
                union +
                investor),
            family = mixture(gaussian(link = "identity"),
                             gaussian(link = "identity")),
            anes_2020,
            seed = 1234,
            control = list(adapt_delta = 0.95,
                           max_treedepth = 12),
            warmup = 4000,
            iter = 5000,
            chains = 4,
            cores = 4,
            prior = priors_BLM)
```

D.3 Results

	ICE		BLM	
Parameter	mean	sd	mean	sd
Intercept	-3.7	1.8	-1.9	1.6
Normalized Income	0.2	0.1	0.1	0.1
College				
Yes	8.2	1.6	7.5	2.4
Student Loans				
No	-0.6	0.9	-0.4	0.9
Union				
No	1.5	1.0	1.6	0.9
Investor				
No	-8.3	1.7	-3.2	1.7
Employment Status				
Temporarilylaidoff	1.2	2.1	2.5	1.8
unemployed	2.1	3.7	1.2	3.0
retired	1.1	0.9	2.2	0.8
Permanently Disabled	0.9	2.9	-1.6	1.7
homemaker	1.9	1.7	0.4	1.6
student	2.4	3.6	1.2	4.0
Class				
Working Class	-5.6	1.7	-3.7	2.4
Middle Class	1.6	1.1	1.3	1.0
Upper Class	0.2	1.8	-0.2	1.6
Occupation				
Non-profit	-0.6	1.0	-0.3	0.8
LocalGov	-1.2	1.3	0.1	1.1
StateGov	0.9	1.7	1.4	1.8
ArmedForces	-4.0	5.2	-2.9	3.9
FedGovCivilian	0.3	2.4	-2.8	2.5
Owner non-incorporated	-0.1	1.3	0.8	1.3
Owner incorporated	-0.3	1.7	0.0	1.5
Work for family	1.4	3.5	2.6	3.2

Table D.4: Class-determining coefficient estimates from mixture models.

	Mixture Model				Pooled Model	
	Higher Class		Lower Class			
Parameter	mean	sd	mean	sd	mean	sd
Fixed Effects						
Intercept	28.6	2.1	26.7	2.4	28.5	2.5
age	0.4	0.0	0.4	0.0	0.4	0.0
Party ID	8.1	0.2	6.5	0.2	7.3	0.2
sd Intercept	2.1	1.6	6.2	2.1	6.4	2.6
sd Age	0.0	0.0	0.1	0.0	0.1	0.1
sigma1	22.0	0.3	25.3	0.3	24.0	0.2
Random Effects						
Asian Female	1.1	2.2	4.5	4.6	1.9	4.2
Asian Male	1.4	2.4	4.7	4.5	3.5	4.4
Black Female	1.3	2.4	7.2	3.7	6.1	3.7
Black Male	0.5	2.1	5.1	3.9	2.9	3.8
Hispanic Female	-0.0	1.9	-7.2	3.9	-9.1	4.3
Hispanic Male	-0.4	2.0	-3.7	3.6	-7.0	4.2
Multiple.Race Female	-1.5	2.5	-2.1	3.8	-4.7	4.0
Multiple.Race Male	0.7	2.2	3.3	4.4	2.5	4.3
Native.American Female	0.9	2.6	2.0	4.5	0.6	4.6
Native.American Male	0.8	2.5	2.2	4.4	-0.8	4.8
White Female	0.7	1.6	2.0	2.9	0.7	2.9
White Male	1.4	1.8	5.3	3.1	3.4	3.0
Age						
Asian Female	0.0	0.0	0.0	0.1	-0.0	0.1
Asian Male	0.0	0.0	0.0	0.1	-0.0	0.1
Black Female	0.0	0.0	0.0	0.1	0.0	0.1
Black Male	0.0	0.0	0.0	0.1	0.0	0.1
Hispanic Female	0.0	0.0	0.1	0.1	0.1	0.1
Hispanic Male	0.0	0.0	0.1	0.1	0.1	0.1
Multiple.Race Female	-0.0	0.0	-0.0	0.1	-0.0	0.1
Multiple.Race Male	0.0	0.0	-0.0	0.1	-0.1	0.1
Native.American Female	0.0	0.0	0.0	0.1	0.0	0.1
Native.American Male	0.0	0.0	0.0	0.1	0.1	0.1
White Female	0.0	0.0	-0.0	0.0	-0.1	0.1

White Male	0.0	0.0	-0.0	0.0	-0.1	0.1
lprior	-69.6		4.6		-15.9	0.0
log-posterior	-26118.0		8.4		-26078.9	6.5

Table D.5: Regression results for the Mixture and Pooled models for ICE thermometer rating.

	Mixture Model				Pooled Model	
	Higher Class		Lower Class			
Parameter	mean	sd	mean	sd	mean	sd
Fixed Effects						
Intercept	61.2	2.3	66.4	2.8	62.6	2.0
Age	-0.2	0.0	-0.2	0.0	-0.2	0.0
Party ID	-10.9	0.2	-9.0	0.3	-10.3	0.2
sd Intercept	6.0	2.1	6.6	1.9	4.9	1.6
sd Age	0.1	0.0	0.1	0.0	0.0	0.0
sigma	22.4	0.4	27.8	0.5	25.0	0.2
Random Effects						
Intercept						
Asian Female	-4.1	3.4	0.1	4.8	-2.2	3.2
Asian Male	-1.2	3.2	-2.1	4.9	-1.3	3.0
Black Female	7.2	3.7	5.9	3.4	7.1	2.8
Black Male	4.1	3.7	1.7	4.0	3.3	3.1
Hispanic Female	4.3	3.2	-0.9	3.5	2.2	2.7
Hispanic Male	-2.1	3.1	-2.6	3.5	-1.2	2.6
Multiple.Race Female	3.5	3.6	-0.8	4.1	2.3	2.9
Multiple.Race Male	-2.9	3.6	-4.1	4.6	-3.0	3.0
Native.American Female	-1.2	4.7	3.1	5.2	1.7	3.6
Native.American Male	-9.3	5.4	-4.0	3.2	-2.5	3.5
White Female	1.4	2.4	2.4	4.7	0.2	2.1
White Male	-3.0	2.5	-12.6	3.4	-6.4	2.2
Age						
Asian Female	-0.0	0.1	0.0	0.1	-0.0	0.0
Asian Male	-0.0	0.0	0.0	0.1	0.0	0.0
Black Female	0.0	0.1	0.0	0.1	0.0	0.0
Black Male	0.0	0.1	0.0	0.1	0.0	0.1

Hispanic Female	-0.0	0.0	0.0	0.1	0.0	0.0
Hispanic Male	-0.0	0.0	0.0	0.1	0.0	0.0
Multiple.Race Female	0.0	0.1	0.0	0.1	0.0	0.0
Multiple.Race Male	0.0	0.0	0.0	0.1	0.0	0.0
Native.American Female	-0.0	0.1	0.0	0.1	0.0	0.0
Native.American Male	-0.1	0.1	0.0	0.1	0.0	0.0
White Female	0.0	0.0	0.0	0.1	0.0	0.0
White Male	0.0	0.0	0.0	0.0	0.0	0.0
lprior	-63.4		5.9		-17.1	0.0
log-posterior	-26301.0		8.4		-26316.8	5.2

Table D.6: Regression results for the Mixture and Pooled models for BLM thermometer rating.