
Analysis of Dimension Expansion in Spatial Modeling of American Voter Behavior

By

Zachary Parrott

A thesis submitted to the

Department of Government

&

in partial fulfillment of the requirements for completion of the
Government Honors Program in the College of Liberal Arts at
The University of Texas at Austin

May 2018

Advisors: Stephen Jessee and John McIver

Copyright statement

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognize that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's explicit consent.

© **Zachary Parrott, 2018**

Acknowledgements

I would like to express my sincerest gratitude to all the individuals that helped me throughout my entire research and writing process, especially my advisors. Thank you so much to Dr. Stephen Jessee and his continued guidance and patience. I feel truly grateful for the opportunity to get to work with someone that is so knowledgeable in this area of research; I owe a great deal of my understanding of this research to you, and I couldn't have made sense of this without you. To Dr. John McIver, thank you for your guiding hand throughout this process, especially when we needed that nudge the most. Your suggestions were incredibly influential to my work, but I appreciate the most that this concern stemmed from a genuine interest in our research. Finally I would like to thank the University of Texas for giving me the skills and materials that I couldn't have done this without.

Abstract

Analysis of Dimension Expansion in Spatial Modeling of American Voter Behavior

Zachary Parrott

The University of Texas at Austin, 2018

Understanding voter behavior has the potential to give us key insights and reflections on election outcomes and even serves useful in managing campaigns and predicting election outcomes. Yet political scientists have yet to create an accurate model of voting behavior. In fact, the most popular theories assert that there isn't much political sophistication to understand or interpret. Spatial Theory however, asserts that there is political sophistication but in the form of an underlying ideological framework, or an *Ideology*. Spatial Theory seeks to understand and predict voter behavior by relying on ideological similarities between candidates and voters.

But spatial theory isn't perfect either. Spatial models may be one of the most predictive models of voter behavior, but it is undeniably missing out on some predictive feature that affects voting behavior. The goal of this research is to attempt to find these missing features and to see if they can be adjusted for. If these features exists and can be accurately adjusted for, they may complete the model.

The most obvious of these potential features are common demographical features that have been known and empirically shown to have different voting behaviors as populations. I hypothesize that these demographic features affect a voter's world view and value systems. I believe that if I adjust for the effects of these features in my measurement of ideology, my model will get closer to more accurately reflecting voter behavior.

Table of Contents

I. Introduction	6
Overview	6
II. Foundation of Spatial Theory	8
2.1 Defining Ideology	10
2.2 Defense of Spatial Theory	12
2.3 Ideology and Outside Predictors	15
III. Building a New Spatial Model.....	18
3.1 Defining the Model and Variables	22
3.2 Process of Adjusting	26
IV. Insights from the Initial Model	29
4.1 Demographic Differences	31
V. Creating an Adjusted Model	37
VI. Initial Results	40
6.1 Source of Voter Variation In Election Outcomes	42
6.2 Implications for Spatial Theory	44
6.3 Racial Analyses.....	46
6.4 Insights on the Black Voter.....	48
6.5 Party Self-Identification Analyses	50

VII. Conclusion53

7.1 Looking Forward54

7.2 Reflections55

References.....60

Tables

Table 1| Ideological Effect of Race.....32

Table 2| Ideological Effect of Gender.....32

Table 3| Ideological Effect of Education Level.....32

Table 4| Ideological Effect of Family Income Level33

Table 5| Ideological Effect of Party Identification.....33

Figures

Figure 1| Generic Probabilistic Curve.....20

Figure 2| Survey Issue Set Question List.....58

Figure 3| Base Ideology Model Regression.....29

Figure 4| Base Ideology Probabilistic Curve.....30

Figure 5| Voter Distribution by Ideal Point.....31

Figure 6| Probabilistic Curves by Party ID.....34

Figure 7| Distribution of Republicans across Ideal Points.....35

Figure 8| Adjusted Model Regression.....38

Figure 9| Ideology vs Adjusted Probabilistic Curve.....39

Figure 10| Male Distribution by Ideal Point.....42

Figure 11| Female Distribution by Ideal Point.....43

Figure 12| Black vs Non-Black Probabilistic Curves.....48

I. Introduction

Voting in elections is the vehicle citizens use to relay their political preferences to their government and elected officials. There then must be some relationship that exists connecting these citizen's policy views and the selection made amongst candidates by constituents when voting or more simply defined as voter choice. Thus understanding what makes a voter choose amongst a set of candidates is a key to knowing what policy outcomes an electorate desires and expects. An approach that seeks to understand and explain voter choice is Spatial Theory. The more accurate a spatial model is in reflecting the electorate's voting choices, the more insights and assertions we will be able to attain from it. It is thus my goal to build a more accurate spatial model that is both more correct and comprehensive than previous attempts at building a spatial model. Previous attempts at building spatial models have sought to gather additional variables and factor them into their models along with ideological proximity. The idea in mind is that there are multiple variables affecting voter choice, and by observing all of the relevant variables we can get a more accurate representation of voter choice in a spatial model. My approach will build upon this sentiment and this thesis will focus on creating a spatial model that takes into account all the relevant variables

and applies them to a spatial model. My hope is that the small effect each variable plays in affecting voter choice will be accounted for, creating an accurate and comprehensive spatial model.

I set out on this research because understanding voter choice, while difficult, is incredibly important. Understanding voter choice has large implications for candidates and politicians. For candidates, understanding why individuals vote the way they do can be a huge help to their campaigning and mobilization strategies. Knowing what voters respond to and where they lie on the issues is invaluable to a campaign, and the accuracy of candidates at predicting this is what can decide elections. Additionally, this information remains valuable to office holders, as it signals to them why voters voted for or against them which provides insights into what the politician's constituents and voters want, which can act as a voter mandate for that official. As a result, understanding voter choice aides both voters and politicians, as citizens can use their votes to communicate what they want from politicians, and politicians can use this information to effectively campaign and serve in office. However, the most important takeaway from voter choice is that it grants us information on what matters to the public. A better and more complete understanding of voter choice would allow us to extrapolate the important issue dimension of today's society. For this reason constructing a more accurate spatial model could have huge payoffs in granting us a clearer understanding of the attitudes and important issue dimensions of our society.

II. Foundation of Spatial Theory

The central premise of spatial theory is that the policy views of both voters and candidates can be represented by a location on a left-right policy spectrum, where voters will choose to vote for the candidate whose location on this spectrum is closest to them. This theoretical approach posits that ideology can be defined through one's policy views and that voter choice is thus determined by ideological proximity of voters to candidates.

Ideology is the key feature in determining voter choice in spatial theory. Spatial theory assumes that voters possess an ideology, which will be defined as a “set of relationships that bind together a person's beliefs across a wide range of different political issue into some sort of systematic structure” (Jessee, 2012). Voters then use ideology as a guide to judge policy decisions and candidates and are a framework used to formulate political opinions.

Spatial theory maintains two key assumptions regarding voter's use of ideology as a framework to guide political opinion. At its core spatial theory relies on rational choice theory, this theoretical framework is characterized by the assertion individuals are rational actors. Rational individuals, taking into account relevant and available information, think rationally

about a set of alternatives they are given to choose from and will choose the alternative that will promote the self-determined optimal outcome. How individuals determine what is the optimal outcome is addressed by the second assumption regarding voters, this assumption is that individual's motivation for action is self-interested (Downs, 1957). Meaning that the optimal course of action for voters is one in which their interests and ideals are pursued and promoted. When both of these assumptions are brought together and applied to the electoral context of spatial theory we arrive at the assumption that "the voter recognizes his own self-interests, evaluates alternative policies or candidates on the basis of which will best serve the self-interest, and casts his vote for the policy or candidate most favorably evaluated" (Enelow, 1984).

It logically follows that in a spatial model there is some point perfectly in between two candidates that can be said to be the ideological middle, this is the ideological midpoint. The assertion in spatial theory that voter choice is determined by ideological proximity results in the presumption that once a voter crosses this ideological midpoint from one side to another their choice in candidate should flip to the one they are now ideologically closer to.

However, it is fairly straightforward to prove that spatial theory is not perfect, a perfect model being one where any and all voters past the ideological midpoint vote for the candidate who is ideologically closer to them. This is simply not the case. In every election there are many voters that vote for a candidate that is ideologically further from them. This is why a far more common practice among scholars is the use of a probabilistic spatial model, where the probability of voting for one candidate over the other approaches 100% the closer one gets to that candidate's position and is near 50% at the ideological midpoint (Jessee, 2012). However, this approach raises an issue. If voters are not voting for the candidate that is ideologically closer to them, then they cannot be said to be rational, for they have failed to properly evaluate options

and vote in what should be their best interests according to ideological proximity. However, rationality is necessary for evaluating and using ideological proximity and thus cannot be abandoned. Thus, the use of a probabilistic spatial model forces us to accept that ideological proximity is not the sole factor that defines voter choice, and that if these voters are indeed acting rationally, they must be evaluating other factors that affect their ideology but aren't being measured.

2.1 Defining Ideology

The methodology that is used to measure ideology is a key feature that will determine how voters and candidates are situated on a spatial model. In *Ideology and Spatial Voting in American Elections*, Stephen Jessee points to the key aspects that are relevant to determining which methodology of measuring ideology should be used. He considers that because spatial theory assumes that there is either some ideological continuum or space, we will need a way of measuring ideology that allows for it to be spatially structured along some left-right ideological spectrum. To do this, an objective foundation for what determines an ideological point must be built (Jessee, 2012). What I seek when I say objective foundation is a way of measuring ideology where each ideological location is a separate location that means the same thing to everyone and would only be replicated by possessing an exactly similar ideology. For this objective foundation we need voters to be placed on the model by a more operational ideology rather than voters choosing their own location through a symbolic ideology. Self-identified symbolic ideologies, such as party identification, represent an ideology that is subjective to the voter, where voters define themselves based off what the ideological labels mean to them. As a result we must turn away from the ANES seven-point distinction scale and party identification, as symbolic identifications such as “liberal” or “conservative” mean different things to different people.

I will require an operational ideology, one that can be said to be objective and consistent in definition across all voters. Issue or policy based measurements of ideology fit these requirements in many ways. The first being that this approach to measuring ideology allows for the potential to empirically establish that people hold political views that are related in rather than assume it through self-identification. This allows us to formally test whether different policy views are related and often held together. This approach allows survey respondents to define ideological dimensions that underlie their political beliefs, rather than presume their ideology based on identification. These ideological dimensions are ideological frameworks that can be thought of as the underlying logic and moral system that an individual uses to evaluate specific policy issues. Secondly, policy-based ideology measurements are much better operationalized as they rely on many survey questions, reducing the granularity of measured ideologies. If we are to think of ideology as existing on a continuum then we need a large and comprehensive set of possible ideologies. This wider range of possible ideological points provides us a more nuanced variation of ideologies for voters to occupy, allowing us to place voters closer to their actual ideological position and differentiate between different but close ideologies. And the third being that by only giving people positive and negative response options for policy proposals, such as “support” and “oppose”, greatly reduces the room for ambiguity in both questions and responses. This grants us a more objective measurement of ideology, as the room for subjective interpretations about the policy issues being asked has been greatly reduced. Jesse exemplifies this well by noting how “if two different respondents both say that they support immediately bringing all troops home from Iraq, it seems safe to say that they are expressing the same view, or at least something very close” (Jessee, 2012).

2.2 Defense of Spatial Theory

It should be noted that the longstanding dominant theoretical framework that most political scientists use to understand voter choice today, the Michigan Model, is starkly in contrast with spatial theory. The Michigan Model asserts that most Americans do not possess a coherent nor rational ideology, and they have very little political knowledge on issues and policies. This theoretical framework is presented originally in *The American Voter* by Angus Campbell, Warren E. Converse, Philip E. Miller and Donald E. Stokes, which was presented nearly 60 years ago, yet remains the dominant theoretical framework. It suggests that Americans possess very little ideological sophistication, meaning that many of their ideological views are unfounded by any deep thought. It operates under the assumption that voters are inattentive and uninterested in politics and thus are not rational in their calculations of who to vote for. Instead they view voter choice as a result of party self-identification and how one thinks of oneself (Campbell et al. 1960). This argument is well founded, logically strong, and is backed by a great amount of reaffirming evidence and studies, as there is a very high association between how one politically identifies and how one votes.

The work of Anthony Downs offers safety for spatial theory from this criticism, but his defense from this line of thinking came 3 years prior to *The American Voter* in his article that is among the earliest to introduce spatial theory: *An Economic Theory of Political Action in a Democracy*. Downs asserts that the above line of thinking is misguided as he states: “Apathy among citizens toward elections, [and] ignorance of the issues ... can all be explained logically as efficient reactions to imperfect information in a large democracy. Any normative theory that regards them as signs of unintelligent behavior in politics has failed to face the fact that information is costly in the real world” (Downs, 1957). Downs argues that it is a waste of time

and energy for most citizens to acquire political information. Downs' argument extends the rational and self-interested qualities of individuals to an even further level than just using them in voting calculations. His main point is that acquiring political information and making an informed decision is costly, whereas the value of any one vote is so insignificant to an election that there is essentially nothing to gain by voting for the "correct" candidate, and thus it is not worth it to waste time trying to acquire the information needed to make an informed decision. So, when Downs defines voters as rational he does not mean it in the traditional logical sense, instead he uses the economic definition of rationality which is more synonymous with being efficient.

Downs poses a method around this inefficiency that allows voters to vote for the "correct" candidate without expending time: ideology. He posits that "many a voter finds party ideologies useful because they remove the necessity for relating every issue to his own conception of the good society" (Downs, 1957). Downs is asserting here that ideologies allow for individuals to evaluate and choose parties or candidates based on underlying ideological dimensions without needing to gather the relevant political information to know what the exact policies are. In general similar ideologies will promote similar policy outcomes, so using ideology as the tool for deciding how to vote based on shared issue dimensions is efficient because it doesn't cost anything while still leading voters to vote for their "correct" candidate, which provides a marginal (even if near zero) value.

Thus Downs offers a strong counter argument. Under his theory, the Michigan Model has misinterpreted the voters' rational efficiency of foregoing on political information as a lack of sophistication and deep thought. Additionally, Downs offers a finding by Hotelling that finds that "The frequency distribution of voters along the [left-to-right] scale is ... fixed in any one

society” (Downs, 1957). This quality of society would mean that, so long as the parties’ ideology remains consistent, there would be a great deal of consistency in voter choice over time. This again is something the Michigan Model has falsely interpreted; it has interpreted this relative consistency of voting to a similar party as using party lines to determine voting because of voter ignorance on policies. The model has assumed that this lack of political sophistication means the voter must just be simply voting along party lines, but the Michigan Model has failed to consider that “if the voter discovers a correlation between each party's ideology and its policies, he can rationally vote by comparing ideologies rather than policies” (Downs, 1957)

Further arguments in the case against the Michigan Model, Stephen Ansolabehere, Johnathan Rodden, and James M Snyder in their article: “Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting.” they show that the use of policy based measurements to measure ideology has been shown to be as reliable and consistent as party identification. Their research finds that when one introduces issue scales composed of many measures instead of a single item survey that ideology can be shown to be just as stable and have a similar association to that of party identification (Ansolabehere et al. 2008). They find that when larger amounts of questions are asked there is more stability and consistency, which they suggest are due to this method removing subjectivity and by using the law of large numbers where there is much less room for variance in a person's answers over time if they are asked many survey items rather than few. In addition asking many questions makes it more likely that the policy opinion structure that may be the most important to an individual is accounted for. Not only does this create a reliable and consistent policy ideology measure, but they also found that this measure was as effective at predicting and explaining the vote as party identification was. They found that opposed to the bulk of literature that using only a single item,

when they use issue scales they “find that issue voting may rival party in explaining the vote. The combined effects of issue preferences are about as large as party identification in the multivariate analyses predicting the vote” (Ansolabehere et al. 2008).

2.3 Ideology and Outside Predictors

I have previously established that ideology can be used as a consistent and reliable measure of ideology, and shown that it can be as effective as party identification. With it established that there isn't a sole dominant way of understanding and predicting voter choice or it at least established that there are multiple measures that can explain it to an equally large extent, we can conclude that multiple factors play a role in voter choice. We have also recognized that most spatial models are not absolutely perfect and show a probabilistic curve at the ideological midpoint. It should be noted that a probabilistic curve displays the likelihood a voter is to choose one candidate over the other at each ideal point. This curve alone cannot tell how many votes one candidate will receive. It must be used along with the raw distribution of voters by ideal point to reflect election outcomes. Two populations that have the same probabilistic curve can produce completely opposite election outcomes if one population is mostly distributed along high ideal points and the other mostly distributed along low ideal points.

With this in mind and assuming the key assumptions of spatial theory, I assert that when an individual votes for the candidate whose ideological location is further away than that of the other candidate then there must be some other factors or variables that are playing a role in that individuals voter choice. I am not alone in this assertion. Previous authors and researchers have explored accounting for other variable that they believe impact vote choice and adjusting their models for these variables impact.

One of the variables that is most explored sharing an impact on vote choice is the other dominant predictor of voter choice: party identification. In both *Ideology and Spatial Voting in American Elections* by Stephen Jessee and “Estimating the Parameters of a Spatial Model of Elections: An Empirical Test Based on the 1980 National Election Study” by James Enelow and Melvin Hinich, the association between ideology, party identification and voter choice are analyzed. These works of literature take the first step in what I hope my research will reach towards. Adding a party identification dimension to the spatial model resulted in a more accurate spatial model in both of these studies. When party identification was factored in to help predict the behavior, there were fewer voters past the ideological midpoint that voted for the candidate they were spatially further from. This is a key insight for my research, as it reaffirms my previous deductions and research. Enelow and Hinich offer an interesting new take on the effect of party identification on voter choice in spatial models. They come to a “plausible interpretation ... that party identification is a policy related variable that represents for the voter a ‘running tally’ of his relative assessment of the two major parties” (Enelow, Hinich, 1985). Additionally, Jessee finds at the end of his research that “The conclusion drawn by a majority of American political behavior scholars has been that partisanship represents a central variable in political thought generally as well as voter behavior more specifically and therefore should be included (controlled for) in virtually any statistical models analyzing vote choice or other political variables” (Jessee, 2012). Although, Jessee does voice concerns that party identification may actually be too synonymous with voter choice, which may make it’s diminish its use for us in determining why individuals favor one candidate over the other. Yet there are still gaps in all these models as it still results in a probabilistic model and isn’t entirely accurate. Even after this

additional correction there remain a number of voters that inexplicably vote for the further candidate ideologically.

Other writing explores other methods as well but with different factors or variables. One such work is “Voter turnout and candidate strategies in American elections” by James Adams and Samuel Merrill. They factors in social factors like demographics, education level and income level. They produce similar results: that these factors when accounted for make the probabilistic spatial model more perfect and results in a steeper probabilistic curve along the ideal point (Adams, Merrill, 2003). While they show a more perfect spatial model there are still voters that go unexplained. This is in part because these factors are largely considered independent of each other in this work as their research was more concerned with how the candidates responded to these qualities. However they are still able to produce a key finding through “extensive behavioral research—that voters are influenced by considerations such as race, class, and partisanship, which are not entirely tied to the candidates’ positions in the current campaign” (Adams, Merrill, 2003). My thought is that a combination of these factors may complete the picture; all the pieces to perfect spatial theory are there they just haven't all been considered at once.

III. Building a New Spatial Model

With my foundation established and relevant previous works of research explored, I turn to the key characteristics my model I hope to create. I introduce these characteristics with the assertion that they are in some way included in an ideological space. I believe that possessing particular characteristics are equivalent to a significant ideological spatial movement. I argue that certain key characteristics are important to how voters interact with policy issues and ideology in a way that affects their voter choice. As a result a particular set of policy issue opinions means different things for voters who differ on key characteristics, which will invariably affect their vote choice in a spatial model.

My goal from the outset is to create a more accurate spatial model, and I hypothesize that after taking many different variables and feature into account when I create my model and adjust for them accordingly that I will create a near perfect spatial model that will be able to serve as a powerful tool in understanding previous election outcomes. Of course, the idea that there exists some perfect model that could accurately reflect every voter's choice is unlikely to be true. Every election will have some amount of voter error. Voter error being an error in the physical voting

process where the voter votes for a candidate they did not want to vote for. This happens in some small and negligible amount by things like voting machine errors, a miss click or mismark on a ballot, or a voter not understanding the voting system but these errors are unlikely to have any significant impact on the model's predictive power. Voter's making errors in judging their own ideology should be mitigated or even eliminated by the use of policy-based ideology.

When I say a "perfect" model, I mean one that eliminates what I will call spatial errors. Spatial errors I will define as those voters who, on some current spatial model, vote against the candidate that is closer in proximity to them on the model. Voter error is caused by voters, whereas spatial errors are caused by a spatial model's incorrect placement or faulty prediction of a voter onto an ideological space. Spatial errors are the result of spatial models not accurately taking into account all of the relevant dimension that affect voter choice, and as a result are missing the dimensions that are key in determining that voter's choice.

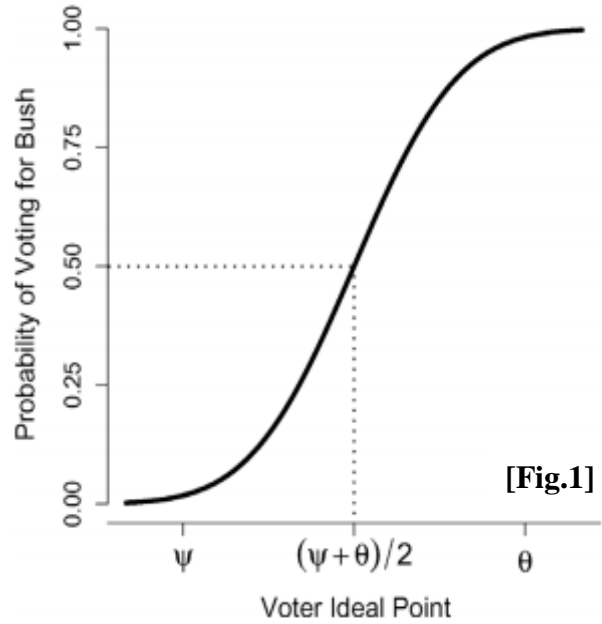
My thesis will focus around reducing these spatial errors and as a result I will seek to account for as many factors that affect voter choice as possible. Of course fitting all of these factors into a spatial model can be very difficult. The traditional simple spatial model is typically only one dimensional, that dimension being ideology. Probabilistic models typically consist of two dimensions, the first being ideology and the second being the probability that a voter chooses a certain candidate. However this dimension reflects a dependent value and thus there is only one relevant dimension in determining voter behaviors so these models are known as monodimensional, when there are multiple independent variable dimensions the models are called multidimensional. Probabilistic models have a probability curve that resembles that of cube root function (see fig. 1). These curves additionally vary in the grade of their slope depending on how high the association is between the voter's ideological proximity and the

likelihood of voting for that candidate. A more accurate model curve would have a much higher grade, where one's likelihood of voting one way increase or decrease rapidly as one moves away from the ideal point.

My model will be probabilistic. Meaning once all my variables have been adjusted for I will be able to create a curve that reflects the new probability of voting for a specific

candidate based on the proportion of voters who have been adjusted to that ideological location that vote for the specified candidate. A perfect model would be a vertical line along the midpoint but this is very unlikely to happen as there will be some spatial error as it is unlikely that I will be able to capture every single possible dimension. Additionally, I return to the fact that a perfect prediction of human behavior is impossible to attain, so any model that I create will have at least some error in it through voter error meaning at the very least there will always be a very slight tilt in the probabilistic curve.

Additionally my spatial model will remain monodimensional. Even though I take into account multiple features, I am using them to construct a single more accurate ideology variable, so a monodimensional model will still work. And while it is enticing to add dimensions as I add in more variables, this greatly complicates the model and carries large consequences. These consequences are well described by the McKelvey chaos theorem. The theorem asserts that if policy preferences are defined over a multidimensional area then majority rule is less powerful. In a multidimensional space there is not necessarily a convergence to the middle which makes it



“theoretically possible to design voting procedures which, starting from any given point will end up at any other point in the space of alternatives” (McKelvey, 1975). McKelvey finds that in multidimensional models, majority rule is not transitive when there is no Condorcet point. A Condorcet point is the point that would win a two-option vote against all of the other points. So when majority rule is not transitive it becomes possible for a voting system to vote in favor of any possibility. This creates significant issues for my model as this would create possible scenarios where the model loses all predictive power as the majority vote could end up anywhere. Additionally, a monodimensional model will restrict me to only being able to observe and make predictions for 2 candidate races. This is due to the fact that if there were three candidates situated on a single ideological dimension it would become the dominant strategy for the two exterior candidates to collapse along this dimension onto the middle candidate. Because there are no candidates to take the voters more ideologically extreme from them and thus no repercussions of moving towards the ideological center, they can both continue to capture more of the middle candidates voters as they become the ideologically closer candidate to more and more of the middle candidates voters until that candidates has been squeezed out of electoral relevancy.

My thesis will be focusing strictly on U.S. presidential elections. I find the U.S. presidential election to be the ideal election for spatial models to be built off of, especially one with my goal. The U.S. is one of the most diverse countries in the world racially, educationally, culturally, financially, and just about every other feature of diversity. This is ideal for my purposes as I seek to adjust and capture the way in which these differentiations affect ideology and vote choice, and having many different variables and features to choose from allows me flexibility in choosing what variables I will use to build as accurate a spatial model as possible.

Furthermore, the primary election year I will focus on is the 2016 election. I chose this year as I want my model to be as modern as possible so to increase my predictive power of future elections.

However, I have had concerns about how unique and hostile this election was. I do have concerns that the rhetoric and candidates played a larger role than normal in normal election years which will mitigate the predictive power of many of my variables and features. Conversely, I also have concerns that this election may have been too polarizing along demographic lines and this will result in me over correcting for the roles these feature play in determining voter choice and I will overstate the impact of these features. While it is a concern that I will lose predictive power if this election was an extreme anomaly, the model will just then instead serve as a very interesting and powerful evaluative tool to understanding the intricacies and finer points of the 2016 election. When my model has been completed and adjusted, I will apply it to previous elections for comparisons. If it matches well and produces a similar result where the probabilistic curve has a steep grade, then I will be able to use the model predictively and hopefully will be able to apply it to future elections. However, if the election is an anomaly in voter choice then the probabilistic curve will have a shallower grade.

3.1 Defining the Model and Variables

The first and most central defining characteristic of the spatial model I hope to build will be how I define and measure ideology. As I have shown earlier in this thesis, there are significant benefits in using policy issues to measure ideology and this measurement can be proven consistent and reliable with large issue scales. Because of the advantage in having an objective and operationalized measurement that is as consistent and reliable as any other predictor of vote choice, this is the measurement I will use. I will be using the 2016 data set from the Cooperative

Congressional Election Study (CCES). This data set contains 64,000 adult respondents who took the survey in two phases. The first wave was administered in a pre-election phase from September to October, where they answered two thirds of the questions. This part of the survey gathered general political attitudes, various demographic factors, assessment of roll call voting choices, and political information. The post-election second wave was administered in November and the respondents answered the remaining third of the questions which covered items related to the 2016 election.

The CCES is perfect for my purposes, as it provides a large scale and extremely detailed and comprehensive survey. Because of this I was able to have an issue scale consisting of 37 issue policy questions. According to Ansolabehere et al. in “Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting.” this is more than twice the required issue scales needed to approach the validity, predictability, and consistency of party identification at determining voter choice, so it is safe to say that it will be comparable in its ability to predict voter choice.

When selecting which questions to include in the data set from the CCES survey, every question that was selected had a support/oppose answer type. This was done in an attempt to minimize polling error, as there is evidence to suggest that poll respondents are poor at deciphering the nuance between answer choices as the number of possible answer choices increases (Ansolabehere, 2008). Another criterion for question selection was to only select questions where there was a clear dichotomy between the ideologically liberal and conservative answer. This was done to so that the questions could be reflected on a monodimensional liberal-conservative scale and to ensure they were significant in reflecting where voters land on this scale. This led to a significant reduction in the number of usable questions; however the

remaining amount is still more than enough to produce a valid and consistent method of prediction. The questions cover a wide range of political topics and issues from LGBT rights to Immigration, however the selected issue dimensions can be found in figure 2, which has been included at the end of my thesis.

Once all the respondents' answers were coded for each survey question, ideology could be constructed using a factor analysis. A factor analysis is a statistical method used to find the variability among related variables. This allows us to find a variable that is not directly observable by finding how variables that are observable are interdependent and connect. This newly found variable is known as a latent variable. So to find the latent variable of ideology, the respondents' answers were aggregated and compared to find patterns and connections in policy issue opinions using R. My measure for ideology was constructed using this method, but was performed previously by Dr. Jessee, he provided me with an already operational ideology variable.

On top of this, the policy-based ideology of the two presidential candidates, Hillary Clinton and Donald Trump, was measured. The ideal point, which is the point where they are ideologically positioned and where they would ideally like to see policy moved to, for Clinton and Trump are $-.25$ and $.25$ respectively. Their ideal point measures have been scaled to create an equidistant even field on which we can organize our voters based on their ideology, and the factor analysis for the ideological position of all other voters was coerced to fit in this scale. As a result, I have a very basic spatial model from this, with the simple features already set. This includes the ideal points, placement of the two candidates, and the respondents, all of which have already been mapped out according to their policy-based ideologies.

Analysis of Dimension Expansion in Spatial Modeling of American Voter Behavior

I have decided that I will select 5 key demographic and personal features that I will be adjusting for. I picked these 5 because I find that they are crucial to how one identifies, interacts with, and interprets political action and policy issues, which not only affects and changes one's ideology, but what also defines one's ideology. The first and most obvious is party identification. I selected this because of its strong association with voter choice, however, I also chose it as I feel that Enelow and Hinich were onto something when they mentioned a "plausible interpretation ... [of] party identification is a policy related variable that represents for the voter a 'running tally' of his relative assessment of the two major parties" (Enelow, Hinich, 1985) as it fits in with how I want to define these demographic and personal features, as a key part of one's ideology and how one interacts with that ideology. Furthermore, party identification will be unique of the features I adjust as it is the only feature which not only affects one's policy ideology, but can be affected by it as well. The next feature I will account and adjust for is income level, as I feel this has a very clear and resolute impact on how one interacts with and interprets economic policy issues. I will also account for and adjust for Education level. This too has and a slight impact on economic policy issues interpretations. However I found this one to be one of the most interesting features to account for because education level vastly affect how you interpret and interact with anything. The fourth feature I will adjust for is race. This one I feel is one of the most obvious on its impact in regards to the interpretation and interaction with policy issues. There is an undeniable and distinct racial tension that characterizes a great deal of American history and political life. This feature will be important as the interactions with and interpretations of policy issues will no doubt be affected by differing racial cultures, different levels of trust in the American political system, and differing institutionalized treatment. And finally, I will be accounting for gendered differences in which, like race, the interactions with

and interpretations of policy issues will no doubt be characterized by different value systems and a different treatment by society and the political system.

To ensure all features and responses are weighed equally for all respondents, any voter who did not have a complete data set for the chosen responses and features was removed from the data set. This was done to ensure there were no over corrections on particular features because a respondent didn't include other features that would have been more significant in their impact. Additionally, only respondents who voted for Hillary Clinton or Donald Trump were included. While there are plenty of insights on how ideology might bring one to vote independent to be found, the focus of my thesis is on how demographic features affects voting behavior on a liberal-conservative ideological scale which only has room for two candidates. This leaves us with a survey sample of approximately 41,000 respondents.

3.2 Process of Adjusting

With these definitions and features in place, along with issue scales providing the measurements for voter placement, I can build a new spatial model. This will start with finding the overall probabilistic curve. The base spatial model will be based off of policy ideology alone. Essentially, the goal is to establish what the true probability of any individual voter to vote for the candidate ideologically closer to them given their demographic characteristics. This will happen by determining what an individual's probability of voting for a candidate is given their ideological space for their specific demographic collection's probabilistic curve. I will then adjust their position on my model to reflect this true voting behavior.

While policy ideology tells us what policies some voters support, it does little to tell us which issues dimensions are important to them and carry more weight, which is why I will

isolate their effect and adjust the ideological positions along these demographic and personal lines. I am operating under the premise that these features I have isolated are the key dimensions which affect the relative directional importance of specific policies to that group. I use directional here because I will be moving individuals in the direction that their demographic group tends to lean. What I will do to adjust groups is isolate all the demographic groups and observe the correlatory effects that having some demographic feature has on voter behavior. This will be found by running a probit regression using both ideology and each demographic feature as a predictor variable. I will compare the impacts that each of these predictors has on the probability of voting for either candidate against the effect that ideology provided. Upon finding that impact, I will attempt to replicate it and apply it to all members of that demographic to adjust.

To observe the accurate impact of all of these features I will run all of them as predictors at once for the regression my model will be built off of. This is to ensure any confounding effects aren't double measured and to ensure I am only measuring the effect that that individual demographic feature produces. The equivalent ideological effect that possessing these demographic features have can be represented by dividing its correlatory coefficient by the coefficient of ideology. This will happen for every demographic feature to capture and observe the results, which in turn can give us the equation for adjusting each voter. Each voter is adjusted by a sum of all of their demographic features impacts divided by the coefficient of ideology. This represents what each effect's equivalent ideological movement would be. In a way this is converting all those demographic features into ideological movement.

While it won't be perfect in adjusting for what is important to every individual, we can see how as a whole that group's traits pull them left or right towards the opposing candidate.

After a sum of all the effects from demographic features are converted into ideological movement it should average out across the population and produce some increased predictive power of the model. The hope is that when all is said and done, I will have placed the individual in an ideological position that best reflects the real ideology we would expect from someone with their characteristics and their policy views to be, and thus by extension placed them ideologically closer to the candidate they will vote for. Once all individuals have been adjusted for, I will create a new generalized linear model that relies on only the adjusted ideology.

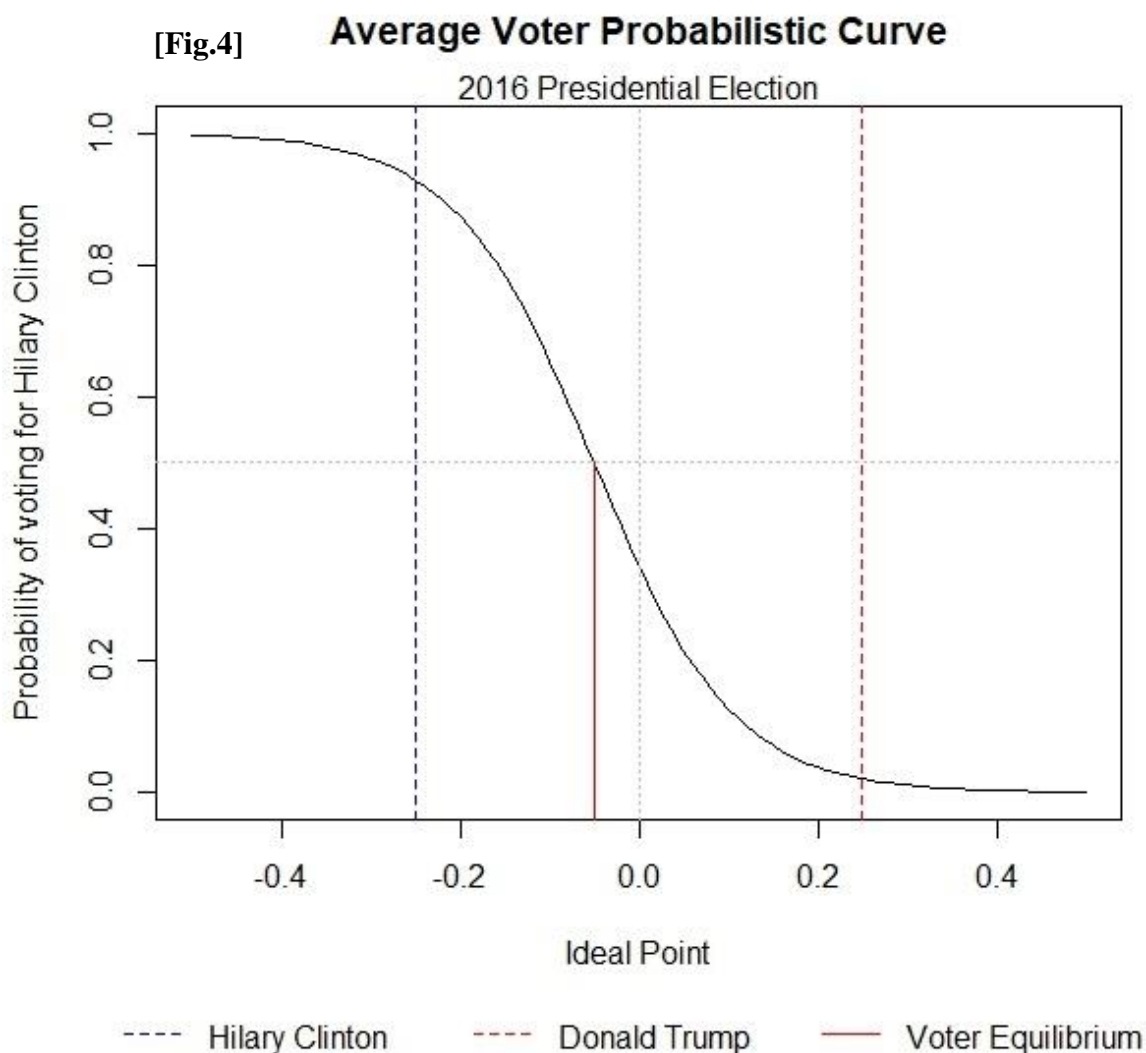
IV. Insights from the Initial Model

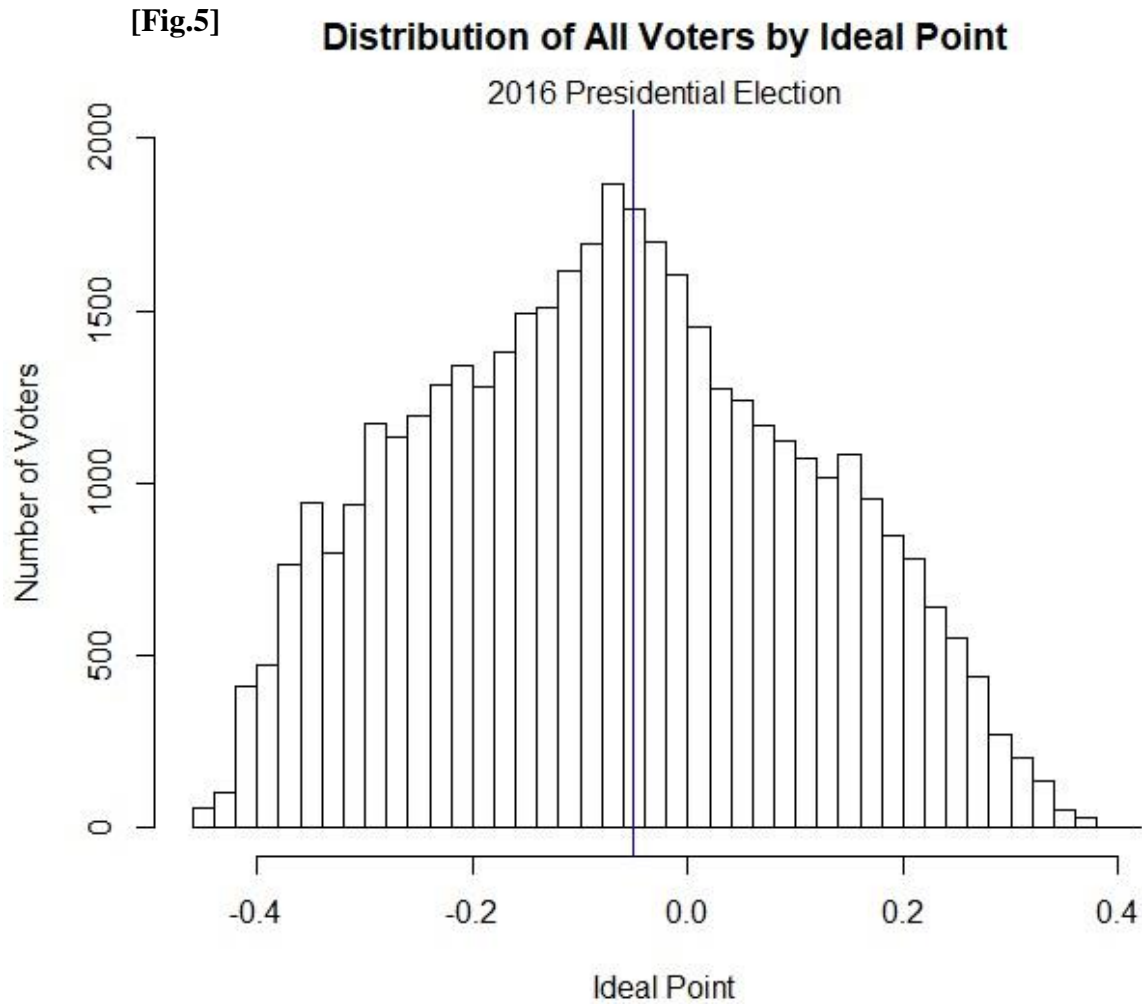
To get my initial spatial model’s probabilistic curve we set up a generalized linear model to run a probit regression to predict voter choice using ideology as our predictor variable. The result of this regression can be seen in figure 3. This regression predicts the likelihood of voting for Hillary Clinton at each ideal point, and after running it we end up with an intercept constant of -0.648 and the slope on ideology is -12.861. So, we can find the probability of someone voting for Clinton as the normal cumulative distribution function evaluated at $-0.648 + -12.861(X)$, where X is the ideal point as defined by our factor analysis. This is the function that is plotted for figure 4.

Figure 4 is the probabilistic curve for the original spatial model. This graph shows the probability of voting for Hillary Clinton at each ideal point using ideology alone as the predictor. The voter equilibrium point, or the ideal point at which the vote splits 50 to

All Voters Regression Results	
[Fig.3]	<i>Dependent variable:</i>
	Vote_Choice
Ideology	-12.861*** (0.124)
Constant	-0.648*** (0.012)
Observations	40,891
Log Likelihood	-10,758.050
Akaike Inf. Crit.	21,520.090
Note:	* p<0.1; ** p<0.05; *** p<0.01

50 either way, is slightly left of the ideological midpoint at an ideal point of $-.0504$. This is expected, as this represents a slight conservative leaning bias, suggesting that Americans are generally more conservative in their ideology, which is consistent with surrounding research (Gallup, 2017). However, when looking at the raw distribution of voters at each ideological point in figure 5, the majority of the voters reside left of the voter equilibrium point, which is denoted by the blue line. This too is expected, as Clinton did win the popular vote, so one would expect more voters to be ideologically closer to her.





4.1 Demographic Differences

Now that the original curve has been created I can begin to separate out individual demographic groups and compare them against the base model. This will allow me to observe and adjust accordingly for what kind of effect that particular feature has on voting behavior at all ideological points. To estimate the effect I multiplied the voter behavior impact of all the demographic groups by 1 if they possessed that feature or by 0 if they didn't. The 1 represents features they possess whereas multiplying by the 0 cancels out the impact of unpossessed features. Finally this impact is divided by the impact ideology has on voter behavior, and that number is the equivalent ideological movement that will be produced by their adjustments.

Table 1. Ideological Effects of Race

	Estimated Effect
Black	-0.140
White	0.019
Hispanic	-0.034
Asian	-0.024
Native American	0.008
Middle Eastern	-0.032
Mixed	-0.025
Other	-0.006

Cells show the estimated amount of ideological movement that is equivalent to the effects of being different races.

Table 2. Ideological Effects of Gender

	Estimated Effect
Male	0.004
Female	-0.003

Cells show the estimated amount of ideological movement that is equivalent to the effects of being either gender.

Table 3. Ideological Effects of Education Level

	Estimated Effect
No High School	-0.013
High School Grad	0.012
Some College	-0.001
2-year	-0.003
4-year	-0.007
Post-grad	-0.005

Cells show the estimated amount of ideological movement that is equivalent to the effects of differing levels of education.

The estimated ideological distance for each racial feature is displayed in Table 1. Upon completing this it became clear that the only racial predictor that had a significant impact on voting behaviors from all other voters was the black demographic. The estimated ideological movement that would be equivalent to being black is over 4 times larger than any other race. This means that the racial feature that most affects voter behavior is the distinction between black and non-black voters. I expected that the black demographic would be overwhelmingly liberal, since no Democratic presidential nominee has received less than 82 percent of the black vote since Kennedy (Johnson, 2016). What was not expected was the lack of significant impact by every other race. This came as a surprise to me, as I expected race to be one of the more impactful demographic features.

Next I move onto gender, where I expect to find a significant gender gap. However, I am again surprised by my findings as gender appears to play no significant impact on vote choice at

each ideal point as seen in Table 2. It appears that gender differences do not cause changes in voter choice that are not already reflected by ideology alone.

Again surprisingly, when I do education level it appears to have no significant impact as well. Differences in education level must either be insignificant to vote choice or ideology already sufficiently encapsulates its effect. Table 3 shows much of the same and further shows how insignificant the impact was.

Income level will be the exact same way as well. As seen in Table 4, there is little movement away from the norm for all income levels as well. And not only is there little impact from the income level predictor, but the slight direction each income group leans is also seemingly random.

Party identification however, shows significant impact on vote choice and their impact is significantly different from the other like features, as shown in Table 5. Their relative impact is as expected, republicans vote for Clinton at lower rates at lower ideal points and democrats vote

for Clinton at higher rates for high ideal points.

Additionally, these effects are stronger for the strong partisans and weaker for the weak

Table 4. Ideological Effects of Family Income Level

	Estimated Effect
>\$50,000	-0.006
\$50,000 - \$99,999	0.003
\$100,000 - \$149,999	0.007
\$150,000 - \$199,999	0.011
\$200,000 - \$249,999	0.009
\$250,000 - \$349,999	0.017
\$350,000 - \$499,999	-0.014
\$500,000 or more	-0.013

Cells show the estimated amount of ideological movement that is equivalent to the effects of differing family income levels.

Table 5. Ideological Effects of Party Identification

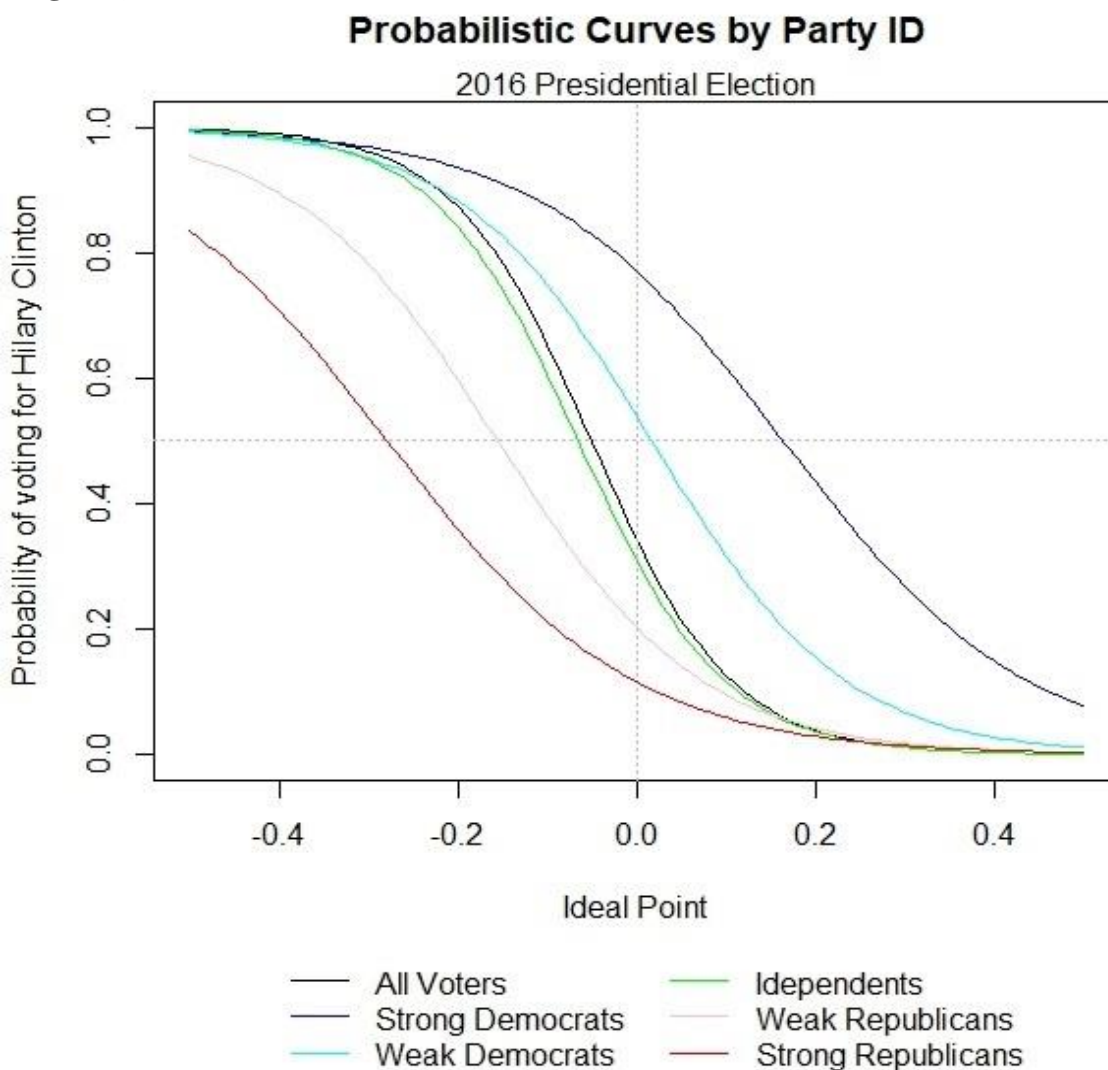
	Estimated Effect
Strong Democrat	-0.215
Weak Democrat	-0.068
Independent	0.014
Weak Republican	0.105
Strong Republican	0.229

Cells show the estimated amount of ideological movement that is equivalent to the effects of party identification.

partisans. This shows that party identification has a very significant impact on voter choice. Furthermore, this suggests there is something about party identification that affects vote choice that ideology does not encapsulate.

The curves for these features, shown in figure 6, are much more separated, and they do a much better job of showing just how significant an impact that party identification has on voter behavior at all ideologies. It should be noted however, that a large reason that these curves present themselves this way is because the distribution of people who identify as a partisan are

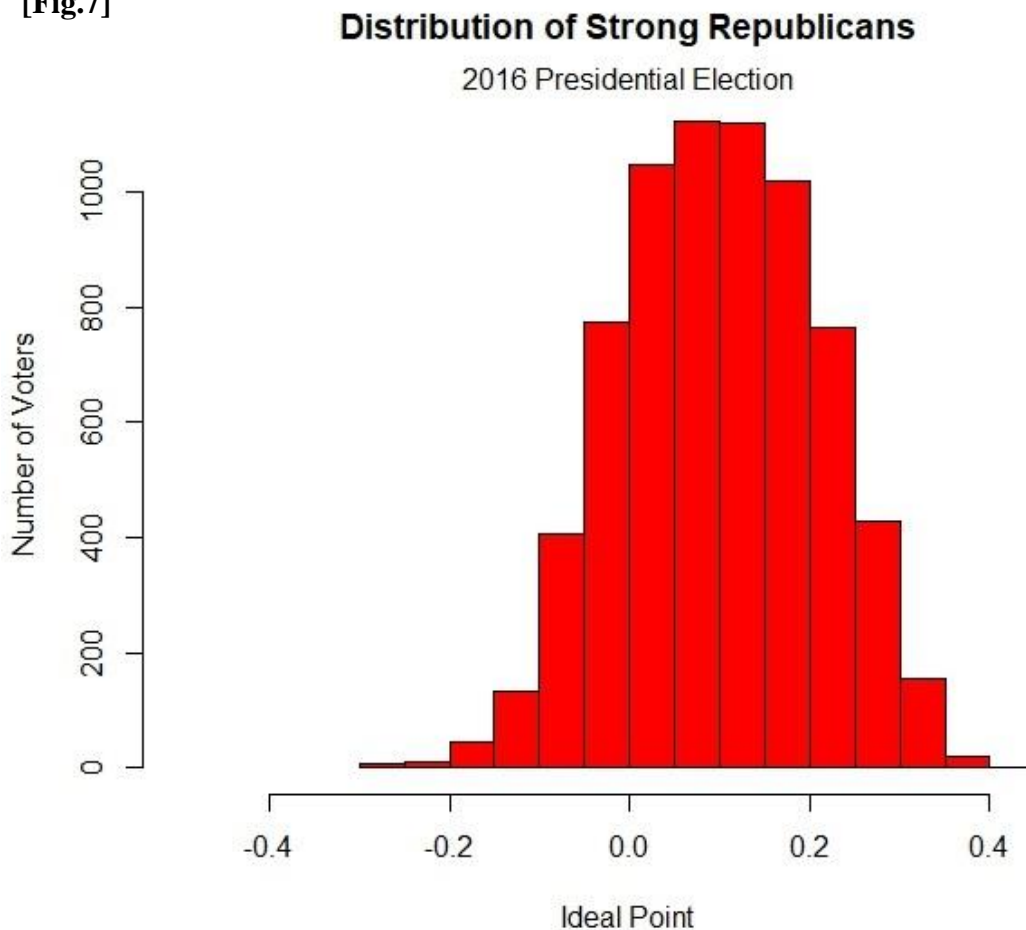
[Fig.6]



almost entirely on their ideological extreme anyway. Meaning of those who identify as a strong republican, the majority of them are at high conservative ideal points already. Figure 7 shows how skewed this distribution is as 75% of those who identify as strong Republicans are already right of the ideological midpoint.

However, it is clear that most of these demographic features were not as impactful party identification, as most features did not result in significantly different probabilistic curves. When features don't show significantly different voting behavior at each ideal point from the average voter, it reflects that these classes of voters do not deviate from the typical voting behavior based on ideological proximity. In other words, similar probabilistic curves between two groups will result in voters behaving a similar way as voters with like ideal points. The implication of this for

[Fig.7]



my result is that, despite differences in voting outcomes between demographical groups caused by differences in the groups' distributions, most voters behave consistently at any given ideal point regardless of their demographic features. Thus differences in election outcomes over time most likely derive from changing voter distributions.

V. Creating an Adjusted Model

With all my selected features observed and compared, I can adjust my final model accordingly. My assumption is that a voter's true ideology isn't just represented by policy issue opinions alone, but these demographic features provide clues as to how different demographics interact with policy issues and which they weigh more heavily. I hypothesize that a set of responses to an issue scale means different things for different demographic groups. An example could something like if two policy issue response sets are the same but the respondents differ in gender. The male respondent's policy issue response set may be relatively liberal for his demographic, but that same response set may be a more relatively conservative one amongst female respondents. I predict that differences like this result in spatial errors, as it seems logical that voters who are relatively more partisan in one direction for their demographic group would vote towards that party more frequently than other voters who are similar ideologically but are more relatively an opposing partisan among their demographic. Thus their true ideology would be a function of their initial policy issue ideology combined with what the equivalent ideological movements and effects for each of the demographic features are. The hope is that I strip away the differences

caused by these features by turning them into “featureless” voters whose demographic biases have already been accounted for.

Each respondent will receive 5 adjustments to their base policy issues based ideology, one for each demographic. Once all the respondents have been given new ideal points that have been adjusted, the model is run again using the adjusted ideal points. To get this new probabilistic curve I set up another generalized linear model to run a probit regression to predict voter choice using the newly adjusted ideology as our predictor variable instead.

The results of the regression can be seen in figure 8.

The result is a slightly more predictive adjusted model. The adjusted curve’s β coefficient is a larger magnitude than the original curve’s, meaning that in the newly adjusted model, the adjusted ideology

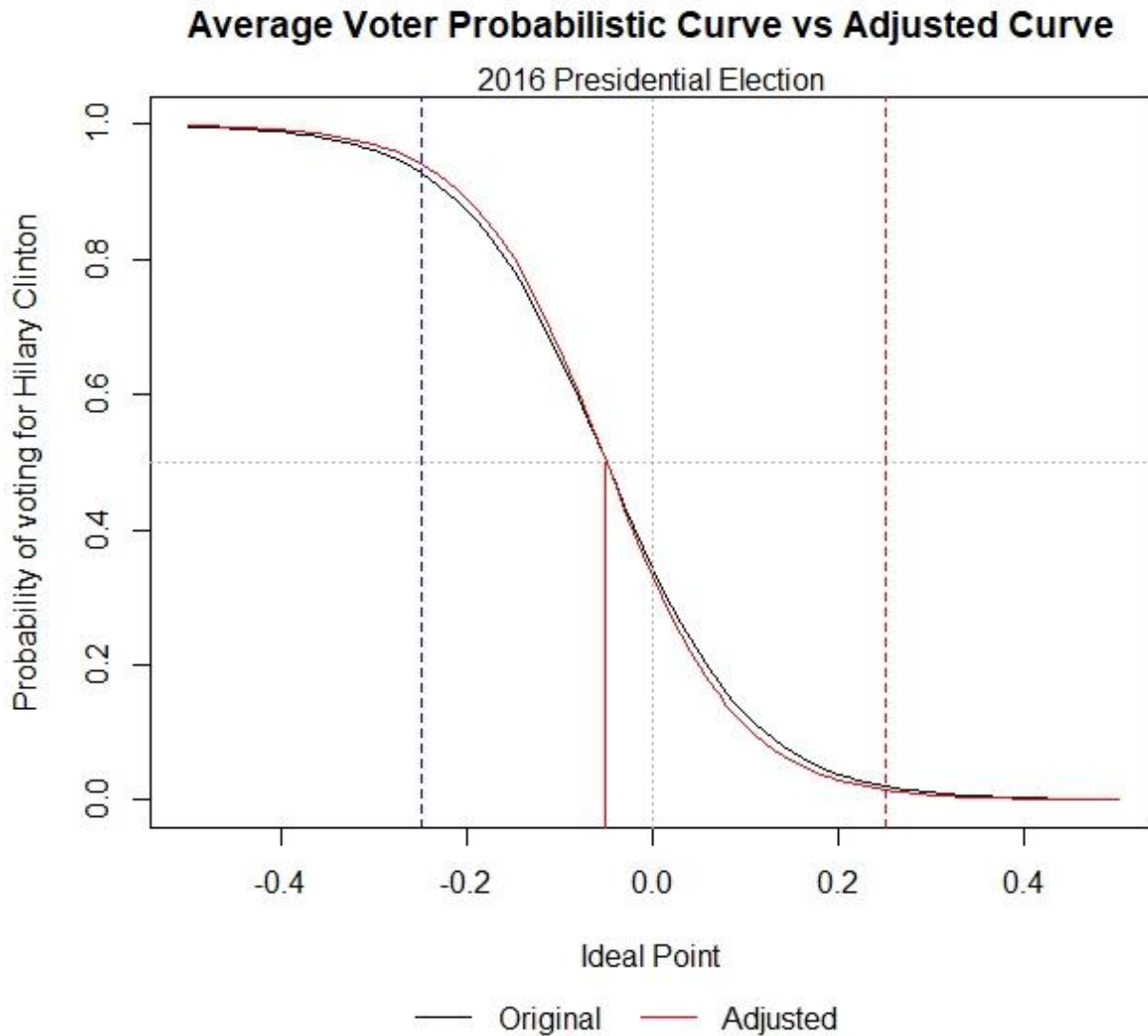
has a larger effect on voter choice than just policy issue ideology had. The increase in predictive power is small, but that may be because it was already very predictive and there wasn’t much room for improvement. The effect especially appears small when shown on a graph, as in figure 9. The adjusted curve is slightly steeper, which as discussed earlier, means is more predictive. This was my expectation but the magnitude of the effect is much lower than I would have expected. As it currently stands, my model is only slightly more predictive as using policy issue ideology alone. While this is not the magnitude I initially expected it is fairly consistent with my findings. Most of the demographic groups, when separated from and compared to the rest,

[Fig.8] Adjusted Curve Regression

<i>Dependent variable:</i>	
Vote_Choice	
Ideology	-13.985*** (1.595)
Constant	-0.705*** (0.137)
Observations	40,891
Log Likelihood	-10,756.050
Akaike Inf. Crit.	21,570.090
<i>Note:</i>	* p<0.1; ** p<0.05; *** p<0.01

showed little change in their probabilistic curve, so it follows that the culmination of all that is miniscule as well. Even the features that did show significant differences were either a small particular minority or were already very predictive on their own.

[Fig. 9]



VI. Initial Results

My analysis of voter behavior has resulted in findings that were different from what I had hypothesized. My initial hypothesis assumed that each demographic feature I adjusted for would have a significant impact on voter behavior. I will define a demographic feature having significant impact on voting behavior as a feature in which differentiation produces an effect on voting behavior that is equivalent to an ideological movement of .05 in either direction from the average voter's probabilistic curve. I chose this amount to reflect the overall slight ideologically conservative leaning bias. My reasoning for this is that despite this bias, American elections outcomes are typically very even and election results differ over time. So if an advantage of that size isn't enough to tip voting behavior consistently in Republican's favor for election outcomes, then it is unlikely that any difference smaller would produce a more significant or consistent effect. Thus these small differences do not have a significant impact on voter behavior. Furthermore, if the demographic features did not have a significant enough effect to at least been able to overcome a small initial bias than it is unlikely to have a significant impact on voter behavior.

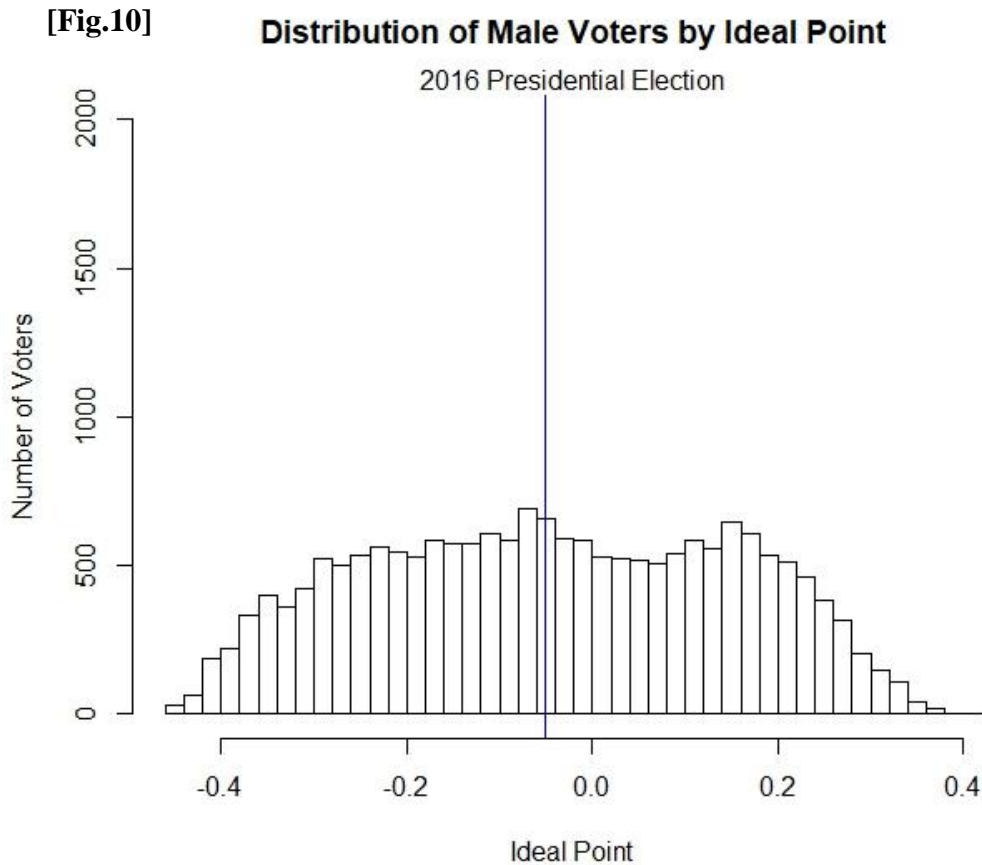
However, most demographic features I adjusted for did not produce a significant effect. Differences in demographic features did not produce significant changes in voting behavior at each ideological position. Only changes in two of the five demographic features produced any meaningful departure from normal voting behavior. These two features are race and party self-identification. Yet even then, the only distinction that was relevant within race was the difference between the voting behaviors of black voters vs. non-black voters, instead of differences between each race.

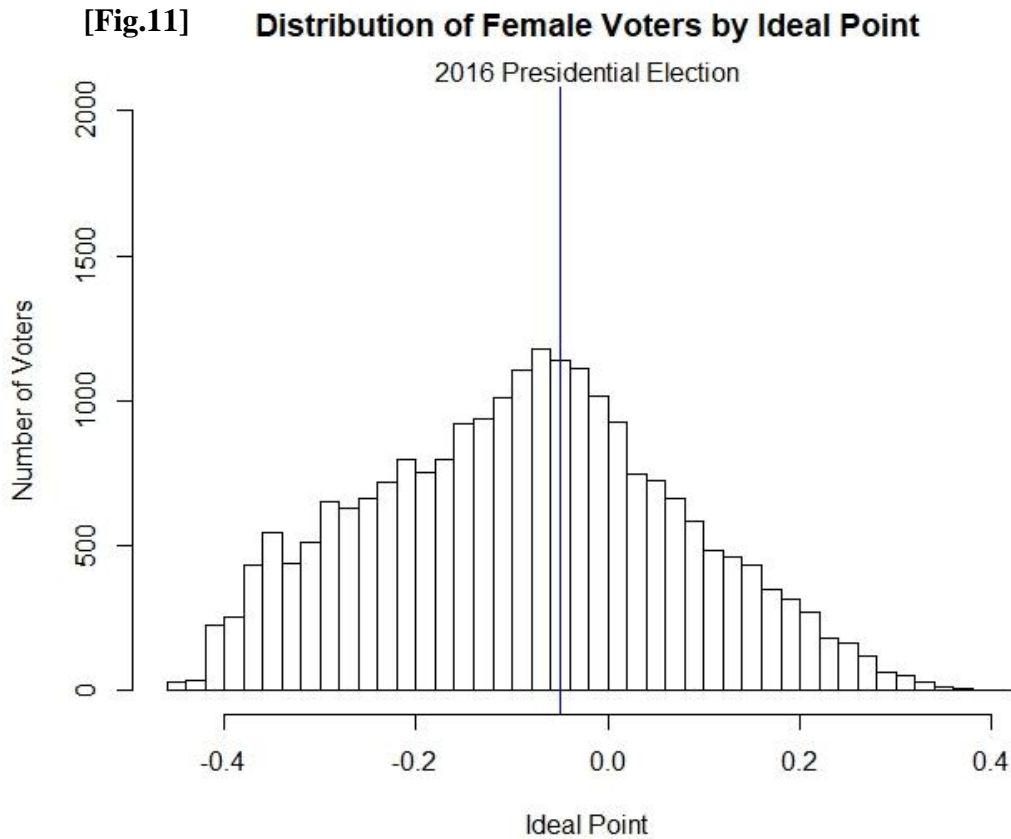
As a result, my adjusted model with all demographic features adjusted for doesn't change much from the original unadjusted curve. This suggests voters will vote similarly to those who are ideologically similar to them regardless of almost all demographic differences. It appears that the data does not support my hypotheses. The data does support that adjusting for demographic features slightly improves the predictive power of spatial theory, but not nearly to the extent I hypothesized it would. Additionally the data does not support that all or most demographic differences are also accompanied with differing interpretations and interactions with policy issue sets.

However, the very low coefficients from demographic in my first model are very telling. Low coefficients suggest that these demographics were not effective at predicting voter behavior compared against the baseline, which was ideology. But these demographics do have different voting behaviors, so knowing that is the case but seeing small predictive coefficients means that ideology has captured and can replicate any their effect, because it was already doing it so well.

6.1 Source of Voter Variation in Election Outcomes

This result isn't as counterintuitive as it may initially seem. This result does not suggest that there are no differences in voting behavior between different demographic groups. It's important to note again that predictive curves for spatial models do not denote how an electorate will vote. Predictive curves in spatial models simply tell you what the likelihood of a voter voting a particular way is at any given ideal point. Voting differences are generated by how different groups are distributed among different ideal points. For example, in this model there was shown to be very little difference between the male and female probabilistic curves. However, we know there is a gender gap and women tend to vote more liberal than men. This gender gap occurs because women tend to have more liberal ideologies so they are distributed





more frequently on ideal points that have high percentages of liberal voters. For example, men are more evenly spread out. This is shown in figures 10 and 11, where the blue line again represents the average voter equilibrium point. It is important to recognize that just because differing features did not produce significantly different probabilistic curves does not mean that these groups do not have different voting behaviors; it only means that their voting behavior is similar among voters with similar ideologies.

This is true for most of the features I use. The curves merely show the proportion at that ideal point that will vote one particular way. But this strengthens the case for using large issue scales to assemble spatial models to predict voting behavior. The lack of deviation is good for an unadjusted spatial model, because that means that the issues scales are capturing and properly representing issue dimensions and ideology regardless of demographic context. It is ideal if a

survey set can accurately represent all respondents. This means that this model does not need to know your race or gender; it can provide a probabilistic curve using only policy issue ideology, which strengthens spatial theories claim that ideology is the driving force of voter behavior.

6.2 Implications for Spatial Theory

These results strengthen the case for spatial theory as a framework for explaining and understanding voter behavior. While there was no evidence to suggest that adjusting ideology for demographic differences improves the predictive power of the model, there is evidence to suggest that the model was very predictive on its own. Even without any adjustments, voters with an ideal point below -0.05038, which is the left or liberal side past the voter equilibrium point, voted “correctly” for Hillary Clinton 88.6% of the time. And the conservative ideologues on the other side of the voter equilibrium point voted “correctly” for Donald Trump 87.1% of the time.

This already is incredibly predictive and already shows a better predictive correlation than party identification. Party identification correlates to vote choice around 80% of the time (Bonneau, 2013), which means using policy issue ideology with large issue scales provides nearly 10% more accuracy at predicting voter behavior. In the model adjusted for demographic differences, voters who were left of the voter equilibrium point were actually 92.8% likely to vote for Hillary Clinton, and those ideologically right of the voter equilibrium point voted 92.1% for Donald Trump. Thus, my model is nearly 5% more predictive than the original model. However, most of this change came from moving partisans who were ideologically far from their party’s expected location over to the more ideologically fitting side. This occurred as a small percent of partisans had surprisingly neutral or seemingly opposite ideologies from their party but still voted for their party at near 80%. Upon removing adjustments for party identification the

same correlations fell to 89.6% and 88.3% for Clinton and Trump respectively, which is basically the same as the unadjusted model with a very small increase in predictability.

Most importantly though, the fact that the regression predictor coefficients were insignificant for nearly all races, income levels, education levels, and genders suggest that using large issue scales to derive ideology is not biased towards any particular set of demographic features. Using large issue scales created a definition of ideology that is similar among nearly all respondents. There was little to no varying in how these sets of policy issue opinions reflected voting behavior. For all voters except for partisans and Blacks, the original policy issue only ideology was consistent in predicting their voter behavior.

This highlights the advantage of using large issue scales for policy issue ideology. The large amount of question essentially ensures an accurate measure of ideology due to the law of large numbers. While poll respondents may be poor at answering questions that reflect their ideology, by requiring many responses the likelihood that they misrepresent their ideology due to confusion or mistakes falls as they answer more questions. Conversely, their answers will likely converge to their actual ideology as they are presented more opportunities to properly represent it. Respondents are likely to converge to their ideology as policy issue opinions are typically interconnected by an underlying ideological framework is used to evaluate policy issues, which typically leads to connected policy opinions. For example, a respondent may oppose regulatory practices as they are against government intervention in the economy, it would then follow that this logic would extend to issues like taxes, that they likely to oppose as well because their underlying ideological framework logically applies to both issues. These large issue sets also increase the likelihood that issues that are key in defining some voter's ideology are included. In this way, these large issue sets are able to capture and represent the defining policy issues of

entire demographics. By capturing the defining issues of demographic groups in the measure for ideology, ideology is able to effectively capture and account for oddities in the way that some demographic interacts with that issue.

The consistency across all these demographic features is significant. It was able to loosely represent an incredibly diverse and divided population and provide a model that was very representative for nearly anyone. The ability for the model to accurately predict seemingly contrasting groups suggests that this model is a good fit for understanding voter behavior due to its wide reach across demographic lines and its consistency in its relative predictive accuracy. The inclusiveness of this approach strengthens the theory and model as a complete and comprehensive way to understanding and predicting voting behavior.

6.3 Racial Analysis

However, this model is not perfectly inclusive in its relatively accurate predicting power. Initially, race was a feature I expected to have a very large impact on voter behavior for every race in a way that wasn't already encapsulated by ideology alone. I expected race to have a much larger impact especially across all races due to the fact that Clinton won the Hispanic vote by 36 points, and the black vote by 80 points (Pew, 2016). And while this did manifest itself for Black voters it was surprisingly not true for Hispanic voters.

I expected racial differences to affect how one interacts with and responds to their ideology. My hypothesis was that different racial issues and cultural differences would lead to different issue priority and issue conceptualization. I had hypothesized these racially charged differences in experiences and treatment would result in a different level of issue emphasis and different political expectations that would affect voter behavior yet be missed by policy ideology.

But this turned out to not be the case as shown in Table 1. Not a single race outside of black voters had a significant impact.

This largest surprise was how little of an impact being white had on voting behavior and how small of a difference it had on the adjustments. White voters tend to vote more conservatively and even favored Trump by 21 points in the 2016 election (Pew, 2016), yet the estimated effect of being a white voter is one of the smallest differences. It should be noted though, that one explanation of this may be the size of the white demographic group. White voters account for about 70% of the respondents, so it's likely hard for this feature to move any large ideological distance away from the average without dragging the whole average with it. Additionally, the White demographic was fairly normally distributed with a peak very close to the ideological center. It is most likely that white voters were too diversely spread in their ideology and voting behavior to result in a significant enough advantage in either direction to adjust for.

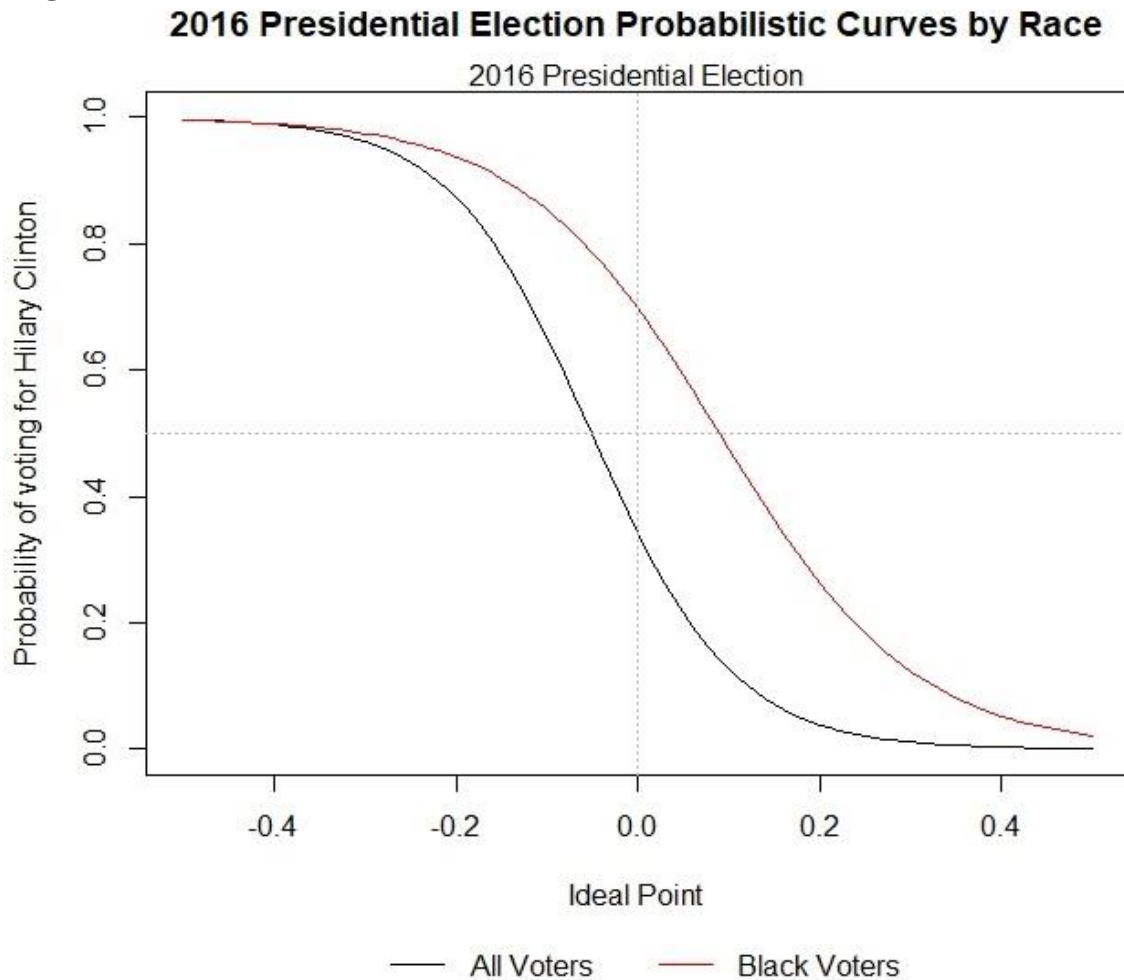
As a whole, race seems to offer little explanation on voting behaviors beyond that of what ideology alone tells us. Meaning either that large policy issue scales do a good enough job of identifying and accounting for differences in different races' conceptualization and interaction with ideology, or race has no effect on one's relationship between ideology and voter behavior at all. Instead it appears that there is only one key racial dynamic in American elections that can't be explained or accounted for by ideological differences alone, which is black vs. non-black voters.

6.4 Insights on the Black Voter

As shown in figure 12, there is a significant departure from the average probabilistic curve by black voters. Black voters are significantly more inclined to vote Democrat, shown by Blacks voting 90% for Hillary Clinton at the same ideological point that the average voter would be expected to vote for Clinton 50% of the time. Similarly, the average voter votes for Hillary Clinton 35% of the time at the same ideological point where Black voters vote. This creates an interesting question though: why are black voters the only racial demographic that has a significant difference in voter behavior at nearly all ideologies?

I argue that this differentiation must be resulting from some characteristic that the

[Fig.12]



Black demographic possesses, as they are the only race whose probabilistic curve is different. On top of that, Black voter's probabilistic curve deviates significantly from the original and all others, so it is likely that they are the ones who possess some quality that results in this. I find it unlikely to be any other race as no other race departs significantly from the original curve and all the other races end up in a similar location to one another.

One possible explanation for this difference may rest in the current American political climate. Whites and Blacks have been shown to have huge discrepancies on the views of race and inequality in America. According to the Pew Research Center, only 36% of whites say racial discrimination is a major reason blacks in the US have a hard time getting ahead; whereas 70% of blacks say it is (Pew Research Center, 2016). This may provide some evidence that the way Black and Non-Black voters identify with and interpret civil rights policy are very far off from one another, resulting in different interactions between the policy issue opinion and ideology for both groups. Furthermore, the entire history of racial conflict in America has in large part surrounded the mistreatment and inequality faced specifically by Blacks. As a result Black voters may identify more closely with this issue giving it much more value relative to other policy issue opinions. Thus black voters may be the only demographic with a significant departure because they are more single issue oriented than other races. Even very conservative black voters vote Democrat frequently, which I argue is because race policy is so important of an issue that they are willing to vote ideologically further away from their ideological location just to promote favorable race policy.

This makes sense given the results, as single issue voters would be poorly by reflected by a 37 question large issue set, because a large issue set weighs all the responses equally which is the complete opposite of how they should be represented. Furthermore, it makes sense that the

Black demographic would lean heavily liberal as the Democratic Party has been far more focused on addressing and rectifying civil rights issues in recent history. Black voters thus heavily identify with the Democrats as they have been more sympathetic to racial minority issues. This is shown by the Pew Research center, as they find 81% of people who identify as Democrat believe the country needs to continue making changes to give blacks equal rights with whites, whereas only 36% of Republicans believe the same (Pew, 2017).

This does raise the question of why this is not true for other races though. Hispanics, for example, have divisive issues like immigration that large proportion of Hispanics are personally affected by or care about. However, immigration isn't a purely racial issue like black racial politics are. Immigration also involves a wide array of topics from foreign policy to economics. Issues like immigration are a far cry from away from solidarity among Hispanics, unlike the Black demographic which has shown to be overwhelmingly consistent in their support of civil rights. For example, only 30% of Hispanics say that a candidate must share their views on immigration for them to vote for them, while 47% say it is one of many factors they consider when selecting who to vote for (Gallup, 2018). Plus, one also must acknowledge that Hispanics are not a monolithic race, as the opinions and culture of differing Hispanic ethnicities can vary widely from ethnic origin to ethnic origin. There just simply hasn't been another historic and universal issue that has been in the center of the American political stage for other races like black civil rights for other races to rally around or care exclusively about.

6.5 Party Self-Identification Analysis

Partisans are unique from other demographic groups in that their membership in that the demographic implies ideological position. Partisanship is unlike other demographic features where there is no ideology that is inherent to those features, but there is for partisanship. It's

logical for strong republicans to be ideologically conservative since the party they are a member is ideologically conservative and pursues a conservative agenda. Thus, this is the demographic feature I expected to see the widest departures from the norm, because of how predictive party self-identification can be at predicting voter choice and because of the inherent ideology it brings. This was the only set of demographics features in which the results were more in line with what I had expected. Each party identity shows a significant departure, except for independents which were expected to be close to the average voter. It comes as no surprise that Republicans vote for Clinton at lower rates and at lower ideal points and Democrats vote for Clinton at higher rates and at higher ideal points. As expected, the effects are stronger for the strong partisans and weaker for the weak partisans.

But significant departures from the average probabilistic curve mean something. It means that ideology alone is poor at predicting and modeling how partisans are going to vote, yet it is a good predictor for independents. Spatial models that use ideology to predict voter behavior have a hard time predicting how ideologically neutral partisans will behave. This is because the policy issue ideology of a partisan doesn't always match up with their self-identified ideology. Problems in predicting voting behavior arise when a voter is coded and placed based off of their policy issue ideology but they behave based on their self-identified ideology. Based on their ideology alone, those partisans near the center should be voting around 50/50 for either candidate, however because they're partisans they actually vote for their party's candidate much more frequently than the model would predict. This strengthens the Michigan model of voter behavior because if the voter's partisanship is known then how they will vote is essentially known. For example, strong republicans only start voting for Hillary Clinton at a -0.27 ideal point which would make that voter more ideologically liberal than Clinton, which simply almost

never occurs. However, the vast majority of these conservative partisans exist far to the right at ideal points where they would be voting 0% for Hillary Clinton anyway, so spatial models only struggle at identifying a small proportion of partisans that end up being ideologically neutral.

Party identification is too predictive to provide much utility to my model. Partisans predominantly occupy the ideological extremes so they vote towards the ideologically closest candidate in very high percentages already. For this reason, adjusting for them doesn't add much predictive power to my model. Essentially I'm pushing the far right even further right and the far left even further left, but that doesn't make my model more predictive because I'm adjusting voters who don't need to be adjusted. I would be adjusting for a group that already votes 80% correctly towards the ideologically closest candidate. The only subgroup of partisans that actually add predictive power to my model is the ideologically neutral partisans, since the adjustments won't move them too far out of my model but will remove them from the location they are hard for spatial theory to predict. So even though they showed significant departures from the norm, they were not effective at improving model predictability.

VII. Conclusion

In conclusion, my research yielded mixed results. On the negative side, I was unable to create the near perfect spatial model that I had set out to at the beginning of my thesis. The most surprising negative I ran into was just how many of my demographic features had little to no impact in their predictive power. I was shocked to see that gender, family income, and education level all produced miniscule correlatory effects. On top of that, nearly all racial features produced no impact as well, with the exception of the Black demographic.

As for the positives, my model was ultimately more predictive than using policy issue ideology only. My final adjusted model actually produced a predictive correlation of 92% between ideology and vote choice. My model has a 4% better predictive correlation than the policy issue ideology only model, and my model is much more predictive than the traditional Michigan Model that only has a predictive correlation of 80%.

And while the lack impact from the predictors among demographic features was not useful in making my model more predictive, it did provide a good support for the use of policy issue ideology in spatial modeling to predict voter behavior. The lack of deviation between the predictive power and probabilistic curves of assorted demographic features and the predictive

power if the ideology only original model shows the model's ability to capture and predict nearly all voters regardless of their demographic features.

This model and its predictive strength suggest that elections and voting are issue based. This means that policy issue opinions are the key to understanding the viability and utility of both voters and candidates. The model suggests elections are issue based due to the models ability to distribute all voters based on these policy issue opinions, regardless of their demographics, and the ability it grants candidates to follow and situate themselves around high distributions of voters with like issue dimensions. Policy issue ideologies serve as a tool to connect voters to ideologically similar candidates and serve as a tool for candidates to communicate their platform to voters.

7.1 Looking Forward

Despite my findings, the issue I set out to address still remains: what are the factors that missing to build the perfect spatial model? The model I created was slightly more predicative but it remains far from perfect. At its best there was a near 92% correlation between ideology and voting behavior. While it is likely impossible to create a completely perfect spatial model, I would like to see a more significant improvement. While 93% is a strong predictive correlation, ideology alone produced an 88% predictive correlation. Only improving by 5% does not seem like the best it can get and adding variables to account for almost always improves the predictive power of predictive model. I would especially like to find room for improvement considering nearly all the demographic features I adjusted for had little to no impact.

I believe that there must be some data fields or features out there that play an impact on a voter's ideology or their voting behavior. Finding these features will be especially difficult considering the features I have already tried adjusting for. Demographic features like race,

gender, education, and income are the most obvious features that one would intuitively think of when considering features that would affect voting behavior. After the obvious features like the ones I have already used, it is unclear what features would be a good next step to try and improve upon the predictability of the model. Some possible options to explore would be features like age, location, urbanization, or occupation. It is also possible that I would see diminishing returns on the predictive power that these new features add as the model gets closer to being as predictive as possible. Additionally if the obvious features had little effect on my model, these secondary features may have an even smaller effect. The question remains if there is still room for improvement.

7.2 Reflections

The first thing I considered when looking back was whether or not the results were significantly different due to the election year I chose. The previous election in 2016 was a very unique and incredibly divisive. Everything in the previous election was out of the ordinary, the candidates, the news coverage, and even the public response.

My biggest concern in reflection would have to be the candidates and how that may have affected voting behavior. Both the main candidates, Trump and Clinton, had huge negatives and carried a lot of baggage. Donald Trump ran an unprecedented campaign focusing almost exclusively on the negatives of Clinton while throwing decorum out the window. Meanwhile, Clinton's campaign was marred with political scandals all while she was vehemently attacked by Trump through social media. The extreme negatives associated with both these candidates concerned me as it is possible that voting behaviors were abnormal in this election as a result. Additionally, this election saw some of the highest third party voting in years, this was a huge

concern for me as it undoubtedly added more abnormal voting behavior. I also removed third party votes from my data set, so this increase in 3rd party turnout resulted in me removing a larger proportion of voters than I would in a more normal election.

Furthermore, this election was incredibly divided along gender and racial lines. In Trump's campaign, he used a lot of divisive rhetoric when referring to Mexican immigrants, Syrian refugees, and immigration policy. This created a lot of racial tensions, which I would expect to lead those racial demographic groups to change their voting behavior or lead them to vote significantly more against Trump, yet both Hispanics and Middle Easterners did not show a significant departure from the policy issue ideology only probabilistic curve. However it is hard to imagine that Trump's rhetoric did not get more Hispanics or Middle Easterners to vote against him, so it would be interesting to compare it to other election years. Similarly, Trump also made several comments that were derogatory towards women. These comments were not received well and I would have expected to see the women's probabilistic curve move away from the average but it essentially went unchanged. Upon reflecting on this I was surprised to see neither of these demographic groups had significant movement from the average curve, especially with how inflammatory Trump's actions were towards these groups. I would have expected these curves to move significantly towards the Democrat side.

In addition to the racial and gender divide, the 2016 election was one of, if not the most, partisan as in 2016 where "58% of Republicans have a very unfavorable impression of the Democratic Party, up from 46% in 2014 and just 32% during the 2008 election year" and Democrats highly negative opinions on Republicans moved "from 37% in 2008 to 43% in 2014 and 55% currently." (Pew Research Center, 2018). As tensions rise across both sides of the isle it is likely that this has resulted in more party voting by partisans. This could have a significant

impact on my results as the curves I separated out by party identification were significantly separated from the average curve. Not only was the effect significant for the differing party identifications, but it was also one of the few features I adjusted for. It's very likely that this divisive election, between two extremely negative candidates, led to more partisan voting which may have affected my results. Upon reflection, I believe I could improve upon this thesis by possibly expanding my model with a few more demographic features and by using a less divisive and abnormal election or by even using multiple elections.

Something I found interesting that I would like to revisit would be to measure the correlations between Ideology and the other demographic features at predicting voting behavior. I'd like to see if it is indeed the case that ideology already captures and represents those issues or if it is that people are just more naturally ideological. This could also shed some light onto whether these demographic factors affect the way in which voters use ideology or if they are independent and unrelated in their impact of voter behavior.

[Figure 2]

Question Code	Question Text
CC16_330a	On the issue of gun regulation, do you support or oppose each of the following proposals? Background checks for all sales, including at gun shows and over the internet
CC16_330b	On the issue of gun regulation, do you support or oppose each of the following proposals? Prohibit state and local governments from publishing the names and addresses of all gun owners
CC16_330d	On the issue of gun regulation, do you support or oppose each of the following proposals? Ban assault rifles
CC16_330e	On the issue of gun regulation, do you support or oppose each of the following proposals? Make it easier for people to obtain concealed carry permit
CC16_331_1	What do you think the U.S. government should do about immigration? Grant legal status to all illegal immigrants who have held jobs and paid taxes for at least 3 years, and not been convicted of any felony crimes
CC16_331_2	What do you think the U.S. government should do about immigration? Increase the number of border patrols on the US-Mexican border
CC16_331_3	What do you think the U.S. government should do about immigration? Grant legal status to people who were brought to the U.S. illegally as children, but who have graduated from a US high school
CC16_331_4	What do you think the U.S. government should do about immigration? Fine US businesses that hire illegal immigrants
CC16_331_5	What do you think the U.S. government should do about immigration? Admit no refugees from Syria
CC16_331_6	What do you think the U.S. government should do about immigration? Increase the number of visas for overseas workers to work in the US
CC16_331_7	What do you think the U.S. government should do about immigration? Identify and deport illegal immigrants
CC16_331_8	What do you think the U.S. government should do about immigration? Ban Muslims from immigrating to the US
CC16_332a	Do you support or oppose each of the following proposals? Always allow a woman to obtain an abortion as a matter of choice
CC16_332b	Do you support or oppose each of the following proposals? Permit abortion only in the case of rape, incest, or when the woman's life is in danger
CC16_332c	Do you support or oppose each of the following proposals? Prohibit all abortions after the 20th week of pregnancy
CC16_332d	Do you support or oppose each of the following proposals? Allow employers to decline coverage of abortions in insurance plans
CC16_332e	Do you support or oppose each of the following proposals? Prohibit the expenditure of funds authorized or appropriated by federal law for any abortion
CC16_332f	Do you support or oppose each of the following proposals? Make abortions illegal in all circumstances
CC16_333a	Do you support or oppose each of the following proposals? Give EPA power to regulate carbon dioxide emissions
CC16_333b	Do you support or oppose each of the following proposals? Raise required fuel efficiency for the average automobile from 25 mpg to 35 mpg

Analysis of Dimension Expansion in Spatial Modeling of American Voter Behavior

CC16_333c	Do you support or oppose each of the following proposals? Require a minimum amount of renewable fuels (wind, solar, and hydroelectric) in the generation of electricity even if electricity prices increase somewhat
CC16_333d	Do you support or oppose each of the following proposals? Strengthen enforcement of the Clean Air Act and Clean Water Act even if it costs US jobs
CC16_334a	Do you support or oppose each of the following proposals? Eliminate mandatory minimum sentences for non-violent drug offenders
CC16_334b	Do you support or oppose each of the following proposals? Require police officers to wear body cameras that record all of their activities while on duty
CC16_334c	Do you support or oppose each of the following proposals? Increase the number of police on the street by 10 percent, even if it means fewer funds for other public services
CC16_334d	Do you support or oppose each of the following proposals? Increase prison sentences for felons who have already committed two or more serious or violent crimes
CC16_335	Do you favor or oppose allowing gays and lesbians to marry legally?
CC16_351a	Congress considers many issues. If you were in Congress would you vote for or against each of the following? Supreme Court Nomination. Approve the nomination of Merrick Garland to the Supreme Court of the United States
CC16_351b	Congress considers many issues. If you were in Congress would you vote for or against each of the following? Trans-Pacific Partnership Act Free trade agreement among 12 Pacific nations
CC16_351c	Congress considers many issues. If you were in Congress would you vote for or against each of the following? USA Freedom Act: Ends the US government's phone surveillance database program. Allows individual phone companies to keep such databases, and allows the government to access those records if there is reasonable suspicion an individual is connected to a terrorist organization
CC16_351d	Congress considers many issues. If you were in Congress would you vote for or against each of the following? Trade Adjustment Assistance Act. Provides education assistance and retraining to workers who have lost their jobs as a result of foreign trade
CC16_351e	Congress considers many issues. If you were in Congress would you vote for or against each of the following? Education Reform. Repeals the No Child Left Behind Act, which required testing of all students and penalized schools that fell below federal standards. Allows states to identify and improve poor performing schools
CC16_351f	Congress considers many issues. If you were in Congress would you vote for or against each of the following? Highway and Transportation Funding Act. Authorizes \$305 billion to repair and expand highways, bridges, and transit over the next 5 years
CC16_351g	Congress considers many issues. If you were in Congress would you vote for or against each of the following? Iran Sanctions Act. Imposes new sanctions on Iran, if Iran does not agree to reduce its nuclear program by June 30
CC16_351h	Congress considers many issues. If you were in Congress would you vote for or against each of the following? Medicare Accountability and Cost Reform Act. Shifts Medicare from fee for service to pay for performance. Ties Medicare payments to doctors to quality of care measures. Requires higher premiums for seniors who make more than \$134,000. Renews the Children Health Insurance Program
CC16_351I	Congress considers many issues. If you were in Congress would you vote for or against each of the following? Repeal Affordable Care Act
CC16_351K	Congress considers many issues. If you were in Congress would you vote for or against each of the following? Minimum wage. Raises the federal minimum wage to \$12 an hour by 2020

References

Christopher H. Achen. Mass political attitudes and the survey response. *The American Political Science Review*, 69(4):1218-31, 1975.

James F. Adams. A theory of spatial competition with biased voters: party policies viewed temporally and comparatively. *British Journal of Political Science*, 31:121-58, 2001.

James F. Adams and Samuel Merrill III. Voter turnout and candidate strategies in American elections. *The Journal of Politics*, 65(1):161-89, 2003.

James F. Adams, Samuel Merrill III, and Bernard Grofman. *A Unified Theory of Party Competition*. New York: Cambridge University Press, 2005.

Stephen Ansolabehere, Johnathan Rodden, and James M Snyder. The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting. *American Political Science Review*, 102:215-32, 2008.

Ansolabehere, Stephen and Brian F. Schaner, COOPERATIVE CONGRESSIONAL ELECTION STUDY, 2016: COMMON CONTENT. [Computer File] Release 2: August 4, 2017. Cambridge, MA: Harvard University [producer] <http://cces.gov.harvard.edu>

Angus Campbell, Warren E. Converse, Philip E. Miller and Donald E. Stokes. *The American Voter*. Chicago: University of Chicago Press, 1960.

Bonneau, C. W., & Cann, D. M. (2013). Party Identification and Vote Choice in Partisan and Nonpartisan Elections. *Political Behavior*, 37(1).

Boorstein, Michelle. The stunning difference between white and black evangelical voters in Alabama. December 13, 2017

Philip E. Converse The nature of belief systems in mass publics. In Richard Nieme and Herbert Weisberg, editors, *Ideology and Discontent*, 322-57. New York: Free Press, 1964.

Anthony Downs. An Economic Theory of Political Action in a Democracy. *Journal of Political Economy*, vol. 65, no. 2, 1957.

Enelow, James M., and Melvin J. Hinich. *Advances in the spatial theory of voting*. Cambridge University Press, 2008.

ENELOW, JAMES M., and MELVIN J. HINICH. "Estimating the Parameters of a Spatial Model of Elections: An Empirical Test Based on the 1980 National Election Study." *Political Methodology*, vol. 11, no. 3/4, 1985

Enelow, James M., et al. Equilibrium in multicandidate probabilistic spatial voting. *Public Choice*, vol. 98, 1999.

Enelow, James M., and Melvin J. Hinich. *The Spatial theory of voting: an introduction*. Cambridge University Press, 1984.

George Rabinowitz, and Stuart Elaine Macdonald. "A Directional Theory of Issue Voting." *The American Political Science Review*, vol. 83, no. 1, 1989

Immigration. (2018, January). Retrieved from
<http://news.gallup.com/poll/1660/immigration.aspx>

Stephen A. Jessee. *Ideology and Spatial Voting in American Elections*. New York: Cambridge University Press, 2012.

^{fig 1} Stephen A. Jessee. Spatial Voting in the 2004 Presidential Election. *American Political Science Review*, vol. 103, no. 1, Feb. 2009.

Maniam, Shiva and Tyson, Alec. "Behind Trump's victory: Divisions by race, gender, education." June 27. 2016.

Pew Research Center. *On Views of Race and Inequality, Blacks and Whites Are Worlds Apart*. June 27. 2016.

Pew Research Center. *Partisanship and Political Animosity in 2016*. June 22. 2016.

Analysis of Dimension Expansion in Spatial Modeling of American Voter Behavior

Pew Research Center. The Partisan Divide on Political Values Grows Even Wider. October 5, 2017.

Stokes, Donald E. "Spatial Models of Party Competition." *The American Political Science Review*, vol. 57, no. 2, 1963, pp. 368–377.