Political Networks

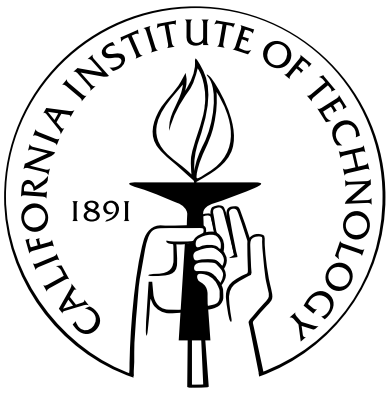
Thesis by

Betsy Sinclair

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2007

(Submitted May 25, 2007)

This dissertation is dedicated to all the individuals who have helped me persevere through graduate school, have inspired me to keep working, and have encouraged me to write it — to my family, to Brian, to my classmates, and to my advisors. I am so appreciative of all of your help, humor, and guidance.

Acknowledgements

Chapter 3 is joint work with Donald Green and Maggie McConnell. We thank Erin Hartman for her assistance in monitoring the experiment and Anne Kamsvaag for her data assistance. We also thank SCOPE for their cooperation and the opportunity to conduct a randomized experiment, as well as the James Irvine Foundation for funding.

Chapter 4 is joint work with Donald Green. We thank SCOPE for their cooperation and the opportunity to conduct a randomized experiment as well as the James Irvine Foundation for funding.

Chapter 5 is joint work with Charles Plott. The authors would like to thank Sarah Hill for assistance in conducting several of these experimental sessions and for helpful comments.

In Chapter 7 the author wishes to thank R. Michael Alvarez, Delia Bailey, Jonathan Katz, Rod Kiewiet, Jonathan Nagler, and Brian Rogers for their helpful comments and suggestions. In particular, the author wishes to thank Gary Jacobson for providing congressional district election returns in 2000.

In Chapter 8 the author wishes to thank R. Michael Alvarez and Jonathan Katz for the opportunity to procure this data, the staff of Polimetrix, and the participants of the CCES March 2007 Meeting for their helpful comments.

Abstract

This dissertation examines the degree to which relationships between individuals determine their political behavior. The dissertation finds, through theoretical and experimental examination, that the political context within which an individual exists affects their preferences. Additionally, using a series of canonical data-sets the correlation between context and preferences is revealed to be a consequence of the information voters obtain via horizontal voter-to-voter communication. The dissertation is divided into five sections, each of which uses a distinct methodological tool in developing my argument that voters are indeed affected by their surroundings, and demonstrates the components of the process by which this happens.

The normative concern regarding the implication of the contextual effect on voters is that if voters are increasingly selecting homogeneous political networks then candidates will use the lack of political discourse to their advantage and choose more ideologically polarized platform positions. Contextual effects, with some exceptions, have not been well studied in political science, due to the lack of detailed data on voters' networks, the difficulty in making causal inferences in the existing group membership data, and the lack of existing theory to guide empirical tests. This dissertation provides a theoretical framework in the first section, and then uses a broad array of tools to evaluate this theory, including laboratory experiments, field experiments, and empirical evaluation of canonical data-sets.

The first chapter discusses the overall findings of this dissertation and locates these results within the literature. The results of the dissertation support different aspects of the argument that political networks are influential in determining individual voter behavior. This first chapter ties together each of the findings into an overall

picture which provides compelling support that individuals' social connections are key components of the way in which they determine their political preferences, and that changes within these relationships have the potential to sway people's political preferences and behaviors.

The second chapter presents a formal mathematical characterization of the relationship between the amount of political information voters glean from their neighbors and the candidates' strategic platform position. This chapter connects the debate about elite polarization (whether or not elites are polarized because voters are polarized) to the idea that voters are heavily influenced by their political network. If voters have communication structures that reduce the amount of heterogeneity in signals about candidates, they are more likely to have correlated vote choices within their political networks, and this correlation can be exploited by ideologically motivated candidates.

The third and fourth chapters attempt to measure to what extent neighbors can influence voter behavior. In a field experiment in South Los Angeles, precinct walkers contacted voters in an attempt to get out the vote in the June 2006 primary and the November 2006 general in Los Angeles County. Three treatments are considered; one in which voters are not contacted, one in which voters are contacted by someone who does not live in their precinct, and one in which voters are contacted by someone who does live in their precinct. The success of voter mobilization is then evaluated.

The fifth chapter tests the theory that voters will aggregate their information based upon their communication structure in an experimental laboratory setting. This chapter evaluates an instance in which all voters observe a common signal. Uninformed voters can procure information from pre-election polls where the poll results include the choices of some informed voters in a simple-majority contest with two candidates in one-dimension. Each voter appears to update their priors about candidate location, considering the decisions and preferences of those voters around her.

The sixth through ninth chapters use survey data to analyze the effects of voters' social networks on vote choice and party identification. The sixth chapter summarizes

the survey data findings across these essays. The seventh chapter analyzes data from the 2000 American National Election Study (ANES) to look for contextual effects in the voters' candidate choice. The respondents are asked a social network battery and these responses, along with vote returns in their congressional district, are used as measurements of the voters' political network. Interestingly, voters whose candidate choice is for one party but whose partisan network distribution is another can easily be distinguished in the results. Individual candidate choices are seen to be determined by their social and contextual context, controlling for their demographic and socioeconomic characteristics and relaxing functional form assumptions. The next two chapters examine party identification within the social context using two distinct data-sets — one from the ANES pilot study in November 2006 and another from an internet panel from Polimetrix, which was also conducted in November 2006. Party identification is also observed to have a social component.

Tying these chapters together is the consistent set of results which demonstrate that the structure by which voters interact with particular groups is key to predicting voter behavior. The previous literature has observed correlations between voter behavior and membership within particular groups, but the research within this dissertation details that different political networks will produce starkly different outcomes. Note here that the contextual effects in the ANES and the elite polarization in the Senate are consistent with the results in the formal model: voters are indeed influenced by their peers (even after controlling for similarities in ideology and socioeconomic status) and elites are polarized. Whether or not these results are indeed tied together by voter communication structures is yet to be determined. However, as the laboratory experiment and field experiment demonstrate, voters are influenced by the ideology and behavior of their peers within their social group. Future research will determine the direction of causality; it is always possible that in fact ideology determines an individual's choice of social network. However, political networks appear to be key factors in understanding political behavior, both for voters and for senators, and the influence of horizontal communication may be so strong as to convince a Democrat to vote Republican.

List of Figures

2.1 Voter x_i's Ideology and Location 16
2.2 Candidate's Strategies in Equilibrium 23
2.3 Comparative Statics of Candidate Strategies as a Function of λ 25
2.4 Expected Candidate Distance from .5 as λ Increases. 27
2.5 Expected Candidate Benefit of Competency as a Function of Number of Signals. 30

4.1 Contact Rate by Percent Home-Turf Canvassing. Note that each dot represents one precinct. 59
4.2 Intent-to-Treat Effect by Percent Home-Turf Canvassing. Note that each dot represents one precinct. 60
4.3 Average Treatment Effect by Percent Home-Turf Canvassing. Note that each dot represents one precinct. 61
4.4 Doorhanger Used in SCOPE Campaign 68

6.1 Sample Payoff 105

8.1 Self-Identified Partisanship Combined with Network or Contextual Effects 143
8.2 ANES: Percent Republican Vote by Discussion Network 144
8.3 Pure Democrat and Republican Respondents' Two-Dimensional Factor Scores 145
8.4 All Respondents' Two-Dimensional Factor Scores by Vote and Majority Party Congressional District 146
8.5 All Respondents' Two-Dimensional Factor Scores by Vote and Majority Party Discussion Network 147

8.6 Two-Dimensional Mean Respondent Factor Score and 95% CI, Mean Calculated For Vote and Majority Party Congressional District Pair (from left-to-right, RR, DR, RD, DD, respectively) 148
8.7 Two-Dimensional Mean Respondent Factor Score and 95% CI, Mean Calculated For Vote and Majority Party Discussion Network Pair (from left-to-right, RR, DR, RD, DD, respectively) 149
8.8 Unbalanced vs Balanced Data 152
8.9 Unbalanced vs Balanced Data 153
8.10 Unbalanced vs Balanced Data 154

List of Tables

2.1 Order of Game 13

4.1 Treatment and Control Assignments and Outcomes 53
4.2 Effect of Any Contact (Using First Wave Assignment): Intent-to-Treat Effect and Treatment Effect 54
4.3 Effect of First Contact Only (Excluding Second Wave Treatment Assignment): Intent-to-Treat Effect and Treatment Effect 55
4.4 Effect of First Contact Excluding Second-Wave Precincts: Intent-to-Treat Effect and Treatment Effect 56
4.5 Marginal Effect of Second Contact: Intent-to-Treat Effect and Treatment Effect 57
4.6 Covariate Balance: Difference Between Contact and No Contact Means 58
4.7 Coefficients from 2SLS, Effect of Any Contact on Vote 62
4.8 Coefficients from 2SLS, Marginal Effect of a Second Contact on Vote . 63
4.9 Difference in Vote and Contact Rate Means: Home-Turf Canvassing vs Not 64
4.10 Logit Coefficients from Marginal Effect of Home-Turf Contact on Turnout Amongst Contacted 64
4.11 2SLS, Separated by Precinct: Home-Turf Canvassed vs Not 65
4.12 Effect of Home-Turf Contact on Voting for Contacted Individuals . . . 66
4.13 Meta-Analysis 66
4.14 Random Effects Meta-Analysis Regression 67

5.1 Treatment Assignments by Precinct 79

5.2 Treatment Assignments, Contacts, and Turnout 80
5.3 Covariate Balance: Difference in Means Across Treatment and Control Assignment 81
5.4 Covariate Balance: Difference in Means Across Realized Contact and No Contact 82
5.5 Effect of Contact: Intent-to-Treat Effect and Treatment Effect 83
5.6 Coefficients, Effect of Contact on Vote 84
5.7 Effects of Home-Turf Precinct Canvassing on Vote 85
5.8 Coefficients from Meta-Analysis, Effect of Contact on Vote 85

6.1 Experimental Sessions 98
6.2 Tabulated Error by Type 99
6.3 Logit Coefficients on Error 100
6.4 Logit Coefficients on Candidate Vote 101

8.1 Respondents' Partisan Majority Network and Vote 136
8.2 Respondents' Presidential Votes by Ideology and Context 137
8.3 Paired T-Tests On Factor Scores Grouped by Vote and Majority Party Congressional District Pair 138
8.4 Paired T-Tests On Factor Scores Grouped by Vote and Majority Party Discussion Network Pair 138
8.5 Robust Linear Regression Coefficients, Dependent Variable: Factor Scores 139
8.6 Logit Coefficients, Dependent Variable: Vote for Republican Presidential Candidate 140
8.7 Logit Coefficients, Dependent Variable: Vote for Republican Presidential Candidate 141
8.8 Logit Coefficients on Balanced Data, Dependent Variable: Vote for Republican Presidential Candidate 142
8.9 Original Logit Coefficients, Dependent Variable: Vote for Republican Presidential Candidate 150

8.10 Balanced Logit Coefficients, Dependent Variable: Vote for Republican Presidential Candidate 151

9.1 Respondent Party Identification and Discussant Party Identification . . 175
9.2 Summary of Variables 176
9.3 Respondent Relationship with Discussants 177
9.4 Treatment Effects 178
9.5 Respondent Party Identification and Discussant Party Identification Raw Numbers 179
9.6 Percent Improvement in Covariate Means From Matching 180

10.1 Respondent Party Change, November 04 to November 06 193
10.2 Summary of Variables 194
10.3 Discussion Network Disagreement: November 04 and November 06 . . 194
10.4 Logit Coefficients, Effect of Disagreement in 04 on Change 195
10.5 Ordered Logit Coefficients, Effect of Change on Disagreement in 06 . . 195

Contents

Acknowledgements **iv**

Abstract **v**

1 Summary **1**

2 Political Networks: The Relationship Between Candidate Platform Positions and Constituency Communication Structures **5**

2.1 Summary 5

2.2 Introduction 6

2.3 The Basic Model 13

2.3.1 Candidates and Voters 13

2.3.2 Signals and Communication 17

2.4 Preliminaries 17

2.5 Correlation Between Proximity and Votes 19

2.6 Candidate Polarization 20

2.6.1 Comparative Statics 22

2.6.2 Uniqueness 28

2.7 The Roles of Limited Information and Truth-Telling 28

2.7.1 Candidate Polarization with Limited Information 28

2.7.2 Truth Telling 31

2.7.3 Candidate Polarization with Most-Preferred Communication . 32

2.7.4 Information Distribution Changes 32

2.8 Conclusions and Future Research 33

3 Discussion: Theoretical Model and Empirical Chapters 35

4 Community-Based Mobilization: GOTV and SCOPE in South Los Angeles, November 2006 38

4.1 Summary . . . 38
4.2 Introduction . . . 39
4.3 SCOPE . . . 40
4.4 Experimental Setup and Data . . . 42
4.5 Methods and Results . . . 44
4.6 Conclusions . . . 51
4.7 Future Work . . . 52
4.8 Tables and Figures . . . 53

5 Community-Based Mobilization: GOTV and SCOPE in South Los Angeles, June 2006 69

5.1 Summary . . . 69
5.2 Introduction . . . 70
5.3 Data and Methods . . . 71
5.4 Results . . . 74
5.5 Conclusions . . . 77
5.6 Tables . . . 79

6 Are Straw Polls an Information Source for Voters? 86

6.1 Summary . . . 86
6.2 Introduction . . . 87
6.3 Experimental Setup . . . 90
6.4 Data . . . 91
6.5 Results . . . 93
6.6 Conclusion . . . 96
6.7 Tables . . . 98
6.8 Written Instructions . . . 102

7 Summary of Survey Data and Political Networks 106

8 What Makes a Democrat Vote Like a Republican? Political Networks and Contextual Effects 109

8.1 Summary . . . 109
8.2 Introduction . . . 110
8.3 A Respondent and Her Network . . . 113
8.4 Data and Variable Construction . . . 118
8.4.1 Voters . . . 118
8.4.2 Latent Utility and Thermometer Scores . . . 119
8.5 Methods . . . 123
8.5.1 Calculus of Voting with Contextual Effects . . . 123
8.5.2 Matching with Multi-valued Treatment . . . 124
8.6 Results . . . 129
8.7 Conclusions . . . 133
8.8 Tables and Figures . . . 136
8.9 Appendix . . . 150

9 Peer-Pressured Party Identification 155

9.1 Summary . . . 155
9.2 Introduction . . . 155
9.3 Data . . . 161
9.4 Method . . . 164
9.5 Results . . . 166
9.6 Conclusions . . . 169
9.7 Future Research . . . 170
9.8 Appendix . . . 170
9.9 Tables . . . 175
9.10 Additional Tables . . . 179

10 Dynamic Party Identification: A Social Network Explanation 181

10.1 Summary 181

10.2 Introduction 181

10.3 Data 184

10.4 Methods and Hypotheses 186

10.5 Results 188

10.6 Conclusions 189

10.7 Appendix 190

10.8 Tables 193

11 Conclusion 196

12 Future Work 200

Bibliography 201

Chapter 1

Summary

This dissertation validates the hypothesis that an individual's political behavior is influenced by her social connections and demonstrates the consequences for representation of the relationship between the individual and the discussants. A plethora of methodologies are used as components of different arguments to support these claims. Each chapter provides a unique contribution to the literature by itself, but tied together, they provide a coherent argument for the existence of network influence and the potential consequences of network influence for representation.

The initial chapter characterizes, in a formal mathematical framework, the relationship between network communication and elite polarization. Here the assumptions about voter behavior will be validated in the empirical chapters which follow — the principle contribution of this chapter for the dissertation is to document the consequences of the voter behaviors which are measured in the empirical sections. This chapter is also interesting in its own right, as it demonstrates that it is possible, in equilibrium, for voter communication within a limited network to produce elite polarization.

The next two chapters describe two get-out-the-vote (GOTV) experiments which are completed relying upon canvassing efforts from a local grassroots organization in South Los Angeles. This organization uses home-turf canvassing and a mobilization effect is produced in both June 2006 and November 2006 amongst a historically low turnout region. The effect in June 2006 is significantly smaller than that of November 2006 — one reason is the group of voters who are targeted. The June 2006 population

consists of all potential voters, whereas the November 2006 population consists only of new and occasional voters. One advantage of this difference is that it is possible to hypothesize that the increased effect in November could be due to the fact that new and occasional voters are most likely to be affected by the treatments — but of course, as these are two distinct elections no conclusions can be drawn from this fact. We do find variance in both experiments in terms of the effectiveness of different canvassers to contact voters — this may be both a function of precinct size, speed of canvassing, and safety of neighborhood. We do find in both experiments large effects in neighborhoods where there is traditionally low turnout — thus we conclude that it is possible to mobilize low propensity voters given the proper mobilization campaign. This campaign demonstrated that using home-turf canvassing has a particularly large marginal effect (4%) on top of an already 6% increase in turnout.

The GOTV field experiment chapters rely heavily upon the use of randomization of treatment assignment to ascertain canvasser effect. It would not be possible to determine the effect of turnout without the randomization assignment, as the only covariates available for each individual are those included within the Los Angeles County Registrar-Recorder's voter file. With only proxies such as age, gender, voter history, and party identification, there is simply not enough data to understand the process by which voters are able to receive the mobilization treatment. One disadvantage of the lack of covariates, however, is that it is very difficult to determine which characteristics make the voters most likely to turn out to vote after receiving the treatments.

The next chapter examines a laboratory experiment where voters update their priors about candidate location based upon a noisy, public signal. We anticipate that voters will reveal their information when it is a weakly dominated strategy, and that voters will assume that other voters are doing so as well. This experiment validates that component of the theoretical argument in the mathematical model, as well as validates the fact that voters will update their priors using Bayes Rule, as opposed to other behavioral models. This paper's contributions, however, are mostly to demonstrate that uninformed voters will be able to ascertain the locations of

candidates from pre-election straw polls, and to reinforce results previously described, but not verified, by McKelvey and Ordeshook (1987).

The final set of chapters all use survey data which ask each respondent to describe some characteristics of their political discussants and some of their own political characteristics. Using the politics of the discussants as an explanatory variable it is then possible to ascertain discussant effect upon respondent vote choice or partisanship. Each of these chapters is susceptible to the criticism that the respondent describes the discussant's politics as consistent with her own. However, existing survey work has documented over 80% accuracy of the survey responses from discussants in these type of surveys.

The first of these chapters uses the American National Election Study (ANES) from November 2000 to examine the consequence of discussant vote choice on respondent vote choice. Here a long list of respondent covariates are included as control variables, such as ideology, gender, age, marital status, etc. This chapter also develops a new methodological technique which permits the identification of discussant political preferences to be more nuanced and to take on multiple values. Additionally, each respondent is situated geographically so as to control for contextual effects as well. This chapter finds that contextual influence is minimal but that discussant influence is large.

One possibility for why contextual influence is minimal is explored in the next chapter, where each respondent is asked, in the course of an internet survey, to identify the geographic distance between themselves and each of their discussants. Over 40% of the responses indicate that the discussants are not even residents of the same city. This chapter, along with the one which follows it, examines the role of discussant partisan identification in the selection of partisanship. One advantage of the internet survey is that it documents a great deal about the relationship between the respondent and the discussant which provides additional control variables to help alleviate the selection problem.

The final chapter also explores the adoption of party identification but uses the panel of ANES 2004 and 2006 respondents in conjunction with the discussant party

identification so as to identify whether or not political disagreement is a good predictor of a change in party identification. Not only is discussant disagreement a good predictor of change, but the change is most often towards the direction of agreement with the discussant.

Each of the empirical chapters uses a range of methodologies, from nonparametric matching techniques to regression. The findings are robust across techniques and across data-sets, allowing me to conclude that so long as the respondents are able to correctly identify the discussants' political preferences, there are sizable effects from discussant political preferences.

The principle findings of each of these chapters provide an introduction to a new perspective on political behavior — that it is highly influenced by an individual's social context. The theoretical paper provides a cornerstone for the consequences of such a finding. If in fact individuals are influenced by their peers in their candidate or partisan choices, then candidates will be able to strategically take advantage of this communication, and, assuming that candidates prefer ideologies which are different than that of the median voter, would be able to be elected with more extreme policy positions. While this is effectively beneficial for candidates, it decreases overall voter utility.

Chapter 2

Political Networks: The Relationship Between Candidate Platform Positions and Constituency Communication Structures

2.1 Summary

This chapter examines the role of voter-to-voter communication in the candidates' strategic choice of platform positions. In the model, correlation in voter information in conjunction with a valence component in the voter's utility function leads to varying degrees of candidate polarization. This chapter presents a formal model of a political network — a network within a social framework wherein individuals have the opportunity to transmit political information to each other conditional upon being socially connected. The chapter then explores the role of information transmission on candidate platform locations in an election with two candidates competing in single ideological dimension. Two possible behavioral assumptions for voters are examined, one in which they communicate their competency signal truthfully and a second in which they communicate their most-preferred candidate. Correlation between preferences and voting choice is observed regardless of the distribution of preferences. Geographic voter polarization is seen to increase candidate polarization when voters

communicate their most-preferred candidate.

2.2 Introduction

Voters' information about candidates comes from many sources before an election, some trustworthy and some not. Voters know that campaigns would like to influence voter opinions by producing biased information but that (mostly) objective sources do exist. Not all of this information is free, however. While the campaigns sometimes spend millions of dollars to bring information into the voters' living rooms, voters must actively seek out objective sources. Yet, while the voters may often realize that they are viewing heavily biased information from the campaign, they may not have the interest or time to investigate unbiased sources and procure information. There is, however, a compromise between spending time and effort reading about candidates and simply viewing campaign advertising. It is possible to simply rely on a trusted source, specifically someone who has procured enough information to make an informed decision and will not relay heavily biased information (Downs 1957). If there is no single individual who is willing to pay the cost to procure all the information about the candidates, it is also possible for individuals to aggregate information (McKelvey and Ordeshook 1985, Converse 1962). This aggregation process occurs completely naturally through friendships (Huckfeldt and Sprague 2004). People are inherently social, and talking to friends is costless. If political information can be aggregated through a network of friends, then no single individual need spend too much time researching candidate positions. This chapter focuses on the process of aggregation that occurs when friends talk to each other — the transmission of political information via a social network.

Voters do talk to their families and friends about politics. In the 2004 National Election Study, 80% of the respondents reported having conversations with their families and friends about politics.[1] A large literature exists (Conover and Feldman 1989,

[1]Of the 1066 individuals who responded to the question "Do you ever discuss politics with your family and friends," 850 individuals responded "yes" in the 2004 National Election Study.

Hamilton 1981, Higgins and King 1981, Miller et al.. 1976, Kinder 1978, Page 1972) about how voters take cues from sources around them to minimize the costs of determining for whom to vote. Downs (1957) formalized this concept of taking cues from a trustworthy source; Lupia and McCubbins (1998) extend this idea to include endorsements of interest groups. Voters can be overwhelmed with the range of information available to them, and the process of filtering to attain unbiased information is difficult when faced with biased news sources such as Fox News or political advertisements. Instead, voters look to the people they can trust to help make their choice. These people are often neighbors, friends, and family members; people with whom the voter has a pre-existing relationship. It is not very costly to have conversations with friends, and it takes only a little time. In the process of sharing information in conversations, no single voter needs to spend hours researching the candidates to gather enough information to make an informed choice. Instead, each voter may aggregate the information she receives in the course of conversation.

Since the voter is aggregating information from her friends and family, the conclusions she can draw are constrained by the set of information available within her social framework. Then, regardless of the voter's preferences, she is likely to have the same beliefs about the candidates as her friends and family. Suppose also that the voter is likely to have the same preferences as her friends and family. In this case it is likely that since her beliefs are correlated with those around her, she will choose the same candidate as those in her social network, both because of similarity of preferences and because of shared information.

This chapter produces a mathematical characterization of the process by which voters communicate about politics with others who are socially connected to them. Individual voters possess preferences over both ideology and competency, a location in the social framework, and a private signal about the competency of candidates. Each individual communicates with a fixed number of other voters. After communication, individuals vote for one of two candidates. This communication has consequences for the amount of information available to certain types of voters to make decisions. The presence of correlation between social connection and ideology is also influential

on the amount of information available. The structure of this communication has consequences for candidate platforms.

Political scientists have long been interested in studying how people form opinions. Lazarfield et al. (1948) concluded that campaigns were not the catalyst for opinion shifts. Campbell et al. (1960) determined that party identification was the most crucial component in determining vote choice. Furthermore, a sizable literature exists to describe the formation of party identification as a product of parental ideology (Niemi and Jennings 1981, Franklin 1984, Achen 2002). Describing the inheritance of party identification is one example of how social relationships may influence voting behavior. One other possibility is that the changing nature of social relationships may explain how there are a group of individuals who do change their opinions (Abramowitz and Sanders 1998).

A possible explanation for the formation of political opinions is that they arise from the aggregation of information within a social network. Since social ties, especially those within families, tend to be fairly constant, it is unlikely that we will observe large changes in voting patterns due to social network changes. This explanation is consistent with the findings of Campbell et al. (1960) that people rely heavily on party identification, and of Niemi and Jennings (1981) that children tend to inherit their parents' party identification. Furthermore, aggregation of political information can then produce voting patterns where voters' choices for candidates resemble the voting choices of those around them.

The transmission of political information occurs over the structure of a social network, and insight into what that structure looks like can be drawn from sociology and microeconomics where a significant amount of work has been done to describe the formulation of social networks. In this literature, individuals are described as nodes in a graph where links (either directed or not) represent connections between people (friendships, coauthors, or simply people who know each other). Authors attempt to describe the shape of the social network, focusing on where links are formed, their distribution, and their density. Examples of social networks abound, from a favorite game for movie lovers "Six Degrees of Kevin Bacon", the online site "Friendster", and

the classic study of "small worlds" by Milgram (1967) where Milgram examined how many times a letter was handled before it was received by the intended recipient. An important distinction for different models within this literature is whether the links are formed as a result of the individuals searching for other individuals or whether the links are formed randomly, and models which assume these two different methods of link formation produce different theoretical predictions about the resulting shape of the network (Rogers and Jackson 2005).

Political scientists are also interested in the shape of the social network as it relates to the flow of political information. When individuals are described as isolated expected utility maximizers it is extremely difficult to produce formal predictions that result in positive levels of political participation, such as voting, much less produce formal predictions that result in positive levels of voters taking the time to research candidates' positions. However, political participation occurs at high levels, despite extremely low pivot probabilities for voters. The explanation that voters are non-rational is extremely unsatisfying, however, since voters do appear to exhibit other rational behavior, such as voting strategically (Alvarez and Nagler 2000). Furthermore, political participation appears to be a function of income, race, and the participation of individuals' neighbors, such as documented by Verba et al. (1995). Beginning in the 1940 Eric County study, Lazarsfeld et al. (1948) noticed that socioeconomic status, not exposure to campaign material, was the best prediction of people's vote choice. As Lazarsfeld writes, "Most people interacted during the campaign with others with shared social characteristics, shared attitudes and, thus, shared political predispositions" (137). This observation leads to the possibility that voters are choosing to participate because their behavior is observed by others within their social network and thus have the same proclivities to vote as the others within their social network. Just as a sports team performs better as a team than they would as individuals, perhaps voters who are part of a social network are more likely to participate in politics.

Participation based upon a social network is consistent with rationality of voters. In his paper, "Free Riders and Zealots: The Role of Social Networks", James

Coleman (1988) provides an example wherein if it is socially beneficial for everyone to participate and individuals are linked together, then people will participate when they would not have as individuals. Other authors (Sheingold 1973, Putnam 1966) have also suggested that voting should be couched in the framework of social interactions. Putnam describes the "social interaction theory" as an instance where "community influence is mediated primarily through the numerous personal contacts among members of a community. Such social interaction within the community would, on the whole, tend to support political attitudes commonly held by community members, and to undermine "deviant views" (641). Putnam finds that "community influence is to a considerable extent mediated through friendship groups" (649). Social interactions are valuable cues for predicting voting patterns. Furthermore, voters are often reported as poorly informed about the candidates' positions (Campbell et al. 1960, Palfrey and Poole 1987, Alvarez 1998).

Additional research has used community cues as predictions of voting behavior, studying questions from the National Election Study (Weatherford 1982). Weatherford cites the literature which collects "chain data" — data which attempt to trace a piece of information backward or forward from an individual to see with whom that information as been shared as evidence that political information is transmitted within a social framework. Weatherford (1982) uses NES data in the aggregate to examine variables that may affect local political networks, which include the number of years the individual has been a resident in their house, whether or not the individual is locally employed, whether or not the individual does local shopping and attends a local church, as well as neighborhood quality. Weatherford also includes a set of individual social and political characteristics (socioeconomic status, political interest, party affiliation and ideology) as well as a set of variables that affect the degree of social interaction with local residents (duration of friendships, frequency of contact). Weatherford then estimates linear regression coefficients for all of these variables using a series of dependent variables which are "network properties" such as politicization and party similarity. The paper finds that the variables that affect the degree of social interaction with local residents "do not contribute to network

politicization and overall have relatively small direct effects on the properties of local networks" (129).

Weatherford, however, is using aggregate network characteristics alongside NES data to try to understand the effects of an individual's social network. This is the best data available to him — individual data on an individual's social network were not available. As he says, "ideally, the study of social and political networks would include interviews not only with respondents but also with their contacts, and perhaps even with the contacts of their contacts" (124).

Huckfeldt and Sprague (1998) collect data in which they both ask respondents specifically about their political contacts and then interview the contacts to get a more complete understanding of an individual's network. Huckfeldt and Sprague attempt via survey analyses to test a theory similar to this paper's proposed theory of information transmission via friendships. They examine both the 1984 and 1996 presidential elections looking for the existence and the effects of political networks. In the 1984 presidential election they focus on the effects of an individual's preferences (and the distribution of preferences) upon the choice of "political discussion partners." They perform a two-stage survey where they first survey a set of individuals and then survey a subset of the people with whom these individuals said they had discussed politics. The authors are interested in the social effect on the content of "socially transmitted political information (468)" and specifically how majority and minority preferences are shaped by social context. Approximately 1500 respondents were interviewed in their initial survey, and these individuals were asked to provide names of "three people you talked with most about the events of the past election year." If no name was generated, they were then asked to provide names of people "with whom you were most likely to have informal conversations during the course of the past few months" (468). A followup survey of these names produced 950 interviews with discussion partners. Each respondent is asked for whom they cast their vote and how they believe their discussion partners cast their votes.

The Huckfeldt and Sprague results indicate that individuals are more likely to have conversations with people who agree with them. They find that two-thirds of

Reagan voters have discussants who reported voting for Reagan, and 57% of Mondale voters have discussants who reported voting for Mondale. Furthermore, people are much more likely to know how their discussants voted if their preferences are similar; 90% can identify how their discussant voted when the discussant shares the same preference, but a much smaller percentage can identify how their discussant voted when the discussant disagrees or does not vote. Conversations between people who disagree about politics are unlikely to happen and fail to produce a clear understanding of the vote choice of either person. Huckfeldt and Sprague (1988) argue that "when a citizen considers whether to discuss politics regularly with another individual, the choice is predicated upon her agreement or disagreement with the potential discussant. If agreement is present, the person is accepted as a discussant, but if disagreement is present then the citizen must either look for a new discussant or accept a politically disagreeable discussant and, hence, a politically dissonant relationship" (470). Disagreeable political conversations, however, are likely to produce a large amount of information about the candidates. Opportunities to have dissonant relationships are limited within the social network framework. Huckfeldt and Sprague (1988) remind us that "people living among Republicans, ceteris paribus, have fewer opportunities to choose political discussants who support Mondale" (470).

Huckfelt and Sprague predict communication between similar types. This model produces a similar result. In this model, a voter's "type", such as her geographic location, her age, or her probability of forming friendships, affects her ability to communicate political information. Consequently, communication among similar types produces biased information which voters use to make their voting decision. More formally, there exists a voting equilibrium function which maps a voter's preferences, types, and information into a candidate choice. There is positive correlation between type and voting choice when information and preferences are distributed randomly. This result exists regardless of whether or not preferences are correlated. If preferences are permitted to be correlated with type ex-ante, the correlation across voting choices increases.

This correlation affects candidate platform positions, with candidates choosing

more extreme points based upon the amount of information available to voters. A large literature exists in political science which describes candidate polarization (Fiorina 2004, Poole and Rosenthal 1984, Collie and Mason 1999), and a sizable component is devoted to including a valence component, that is, an additional candidate characteristic other than ideology which increases all the voters' utilities in the voter's calculus (for examples, see Kartik and McAfee 2006, Schofield 2003, Schofield 2004).

2.3 The Basic Model

The model is presented as a game, where candidates choose platform positions based upon the expected vote share they will receive, which depends on the aggregation of information via voter communication. The game proceeds as follows. First, with equal probabilities nature privately selects a candidate as relatively more competent. Next, candidates state platform positions. Then, each voter draws an informative signal about candidate competency. Voters then have the opportunity to share signals with each other and thereby aggregate information about the candidates' competency rankings. Finally, voters choose candidates in an election. Table 1 summarizes the game.

Table 2.1: Order of Game

1. Nature determines competency.
2. Candidates choose positions.
3. Voters observe competency signals.
4. Voters communicate.
5. Voters vote in the election.

2.3.1 Candidates and Voters

The players in this game are the candidates; voters communicate in an exogenous manner and then vote for the candidate who will maximize their expected utility.

There are two candidates, denoted $C = \{A, B\}$. Candidates are assigned ideal points x_c where $x_A = 0$ and $x_B = 1$ and choose policy platforms p_c in the ideology space $[0, 1]$.

Candidates are assigned a relative competency rating by nature. The more competent candidate will be denoted θ where $p(\theta = A) = p(\theta = B) = 1/2$ (and that probability is common knowledge). Candidates then simultaneously choose their policy platforms without knowledge of their relative competency. In this sense candidates are forced to announce their ideology at the beginning of an electoral contest, when they do not know whether or not they would be more competent at handling the difficulties associated with the job. The model in this sense most closely resembles a contest with no incumbent.

The candidate who receives a majority of votes is declared the winner of the election. A coin flip decides in the case of a tie. Candidate payoffs are zero unless she receives a majority of votes. The winner receives a payoff of one minus the difference between her ideal point and her policy position. Formally, payoffs are $1 - |x_c - p_c|$ if the election is won, and zero otherwise. The intuition is simple — candidates are office-seeking. However, they find it costly, once in office, to implement a policy which they dislike. If they do not win office then their policy is not implemented and there is no discomfort associated with stating a policy which deviates from their own preferences.

There are a finite number N of voters located on a circle, and each voter is connected their nearest neighbors. Their position on the circle describes their geographic location. This concept can be considered as a simple model of the way in which voters form friendships or could be seen as a proxy for the people the voter would in fact reside next to in a neighborhood. Although a great deal of work has been done in the literature on social networks to estimate the likely structure of these networks (in particular, a small world, see Rogers and Jackson 2005 for a review of the literature), in this case we consider only this circle which connects nearest neighbors. The results would be consistent with other network structures, but would complicate the analysis.

Each voter has preferences over ideology and competency of the candidates. Vot-

ers can be described by their geography and their ideal point. Voters draw their (ideological) ideal point from a uniform distribution on the ideology space [0,1]. Voters' ideological preferences are independent of their geographical location. Figure 2.1 illustrates the ideology with red and blue of three voters j, k, and i. Each of these voters have drawn their ideology from a uniform distribution on the unit interval and are assigned a particular geography independently. All voters prefer more competency. In this figure, voter i is receiving a signal about competency which she will communicate with j and k.

Candidate policy positions and ideological preferences are common knowledge to all voters. All voters are given additional information to help them with their decision. As stated, each voter receives a signal about competency. Voters can obtain additional information by communicating with other voters. After communicating they vote for either candidate A or B without abstention.

Voters receive a payoff of $\gamma * \mathbf{1}_{\{\theta=C\}} - |p_C - x_i|$ for voting for candidate C, where $0 \leq \gamma \leq 1$. Thus the expected utility of voter i for voting for candidate C is $u_i(C) = p_i(\theta = C) * \gamma - |p_C - x_i|$, where p_i denotes i's posterior beliefs. Note that utility is separable into an ideology component and a competency component, so can be written $u_i(C) = u_i(I(p_C, x_i)) + u_i(\theta))$. We consider only the case where $\gamma \in (.1, .25)$.[2] Note that thus for some voters, conditional upon the specific candidate locations, the additional expected utility from competency may outweigh any particular ideological preferences, and vice versa. If a voter is indifferent between two candidates, he flips a coin.

[2]The reason for these bounds is that while we anticipate competency plays a role in the voter's decision-making process (so that the value of competency is nonzero), we do not anticipate that the value of competency will be so large as to outweigh all ideological value. While it is possible that voters do choose based upon perceived competency (Todorov et al. 2005), it is also the case that partisanship remains an extremely good predictor of vote choice (Campbell et al. 1960). Here we anticipate that the competency term is also akin to a general valence component which is orthogonal to ideology.

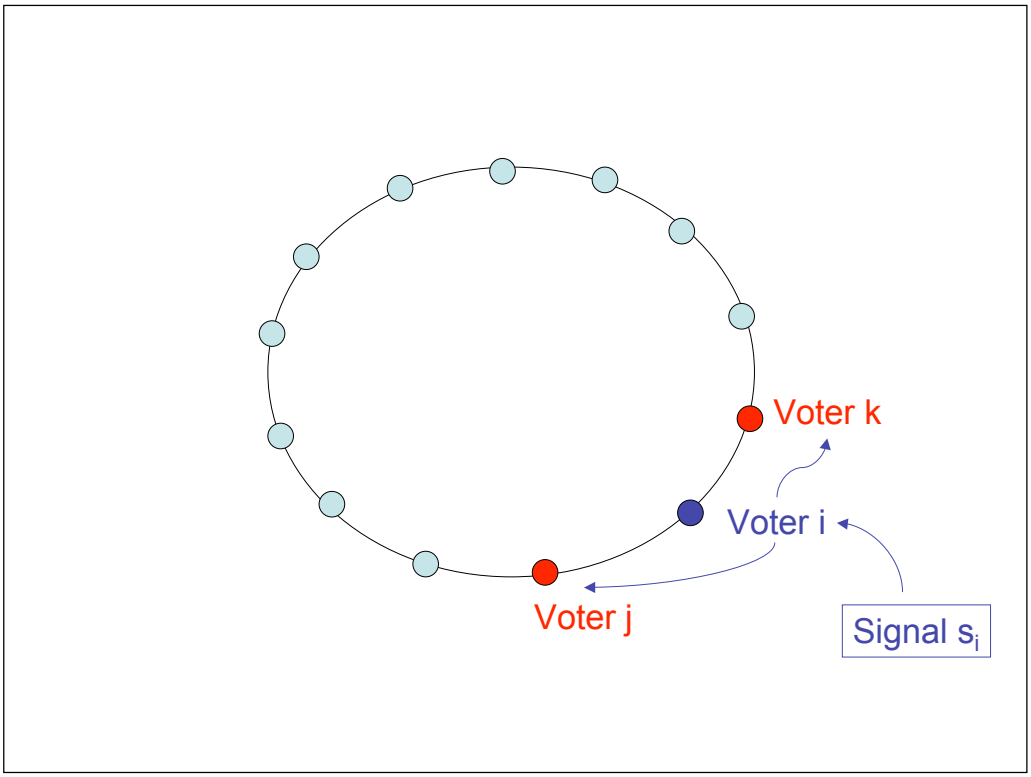


Figure 2.1: Voter x_i's Ideology and Location

2.3.2 Signals and Communication

Each voter i receives a competency signal s_i, where $p = P(s_i = \theta) = 2/3$. That is, she will receive a signal about the more competent candidate with probability p, and she will receive a signal about the less competent candidate with probability $1 - p$. These signals will simply consist of the candidate name. To model the communication stage among voters, suppose that each voter finds it optimal to communicate with S others. Assuming that communication costs increase with geographic distance, this means that voter i will announce her private competency signal with $S/2$ of her neighbors to her left and $S/2$ of her neighbors to her right.[3] Thus, voters receive, from communication, S signals (for ease of calculations, let S be even). Assume that $N >> S$.

At first pass, assume that the voters tell the truth when communicating their private signal. As the payoffs change to be a function of the candidate elected by majority rule, later sections will analyze whether or not the voters in fact have incentives to tell the truth.

2.4 Preliminaries

There are a few observations and assumptions to note before proceeding to the results. First, voters are Bayesian. Let n_i^C denote the number of signals voter i receives indicating that C is the more competent candidate. Then i's belief that A is more competent is a function only of $n_i = n_i^A - n_i^B$, the net number of signals in favor of A.[4] In particular, i's posterior belief that A is more competent is given by $\frac{2^{n_i}}{2^{n_i}+1}$.

Lemma 1: Assume without loss of generality that $\theta = A$. Then, ex ante, $p_i(\theta = A)$ is increasing in p and S.

Proof: Recall that n_i denotes the net surplus of i's signals that indicate A is the more

[3]Voters will all communicate their signals simultaneously — in reality neighbors would likely have some sequential communication as well — but as sequential communication significantly complicates the analysis, the voters currently only simultaneously communicate. Future research will attempt to address sequential communication.

[4]This is due to the voter updating her prior based upon Bayes' Rule.

competent candidate. Let X denote the random variable describing i's signals, so that it is distributed binomial with parameters $S+1$ and p. Then $N_i = 2X - n_i$. Since X obeys first order stochastic dominance with respect to p and S, it is clear that N_i does as well. The contribution to i's utility from competency is $\gamma\frac{2^{n_i}}{2^{n_i}+1}$ where the probability that candidate c is competent is $P(\theta = A) = \frac{2^{n_i}}{2^{n_i}+1}$. Since this is an increasing function of n_i, it's expectation increases with p and S by the definition of first-order stochastic dominance.

Now suppose voter i has ideology x_i and receives a signal about competency. WLOG, suppose she now believes candidate A is competent with probability 2/3. In expectation $p*(S+1)$ of these signals will be about the more competent candidate. Thus, the voter will have, in expectation, $p*(S+1) - (1-p)*(S+1) = (2p-1)(S+1) = n_i$ net signals about the more competent candidate. She will weight her expected benefit from competency by the amount of uncertainty she has about the more competent candidate. Thus, her expected benefit from competency may be described as $\lambda = P(\theta = A)*\gamma - [1 - P(\theta = A)]*\gamma = [2*P(\theta = A) - 1]*\gamma$. She will vote for candidate A if $p(\theta_A)*\lambda - |x_i - p_A| \geq p(\theta_B)*\lambda - |x_i - p_B|$. These are simply her expected utility calculations. Note that as the voter becomes more and more confident about the competency of any particular candidate there is more value attached to competency.

Note that candidates will consider the expected weight placed on competency in strategically determining their platform positions.

Lemma 2: For any two voters i and j where $j = i + (v+1)$, if both communicate with S voters (where $S + v < N$) then they will have X common signals and Y independent signals, where X and Y are described by: (a) If $S \leq v, X = 0, Y = S+1$ and (b) If $S > v, X = S - v, Y = v + 1$.

Proof: Assume i communicates with $i+1, i-1, i+2, i-2, \ldots$ and j communicates with $j+1, j-1, j+2, j-2, \ldots$, in those orders. Thus for each communication step the voters i and j are alternately communicating with one of the v voters which separate them, $i+1, j-1, i+2, j-2$, etc. When $s = v$, each voter separating i and j has

shared a signal with exactly one of i or j. At the next stage, and each subsequent stage, either i or j (but never both) communicate with a voter that the other has already communicated with, generating the claimed expressions.

2.5 Correlation Between Proximity and Votes

This section describes a relationship between the geographical locations of voters and the ways in which they vote. Recall that voters' positions are independent of both the information technology and ideological preferences. However, physical proximity induces correlation in voters' beliefs through the communication process, as voters who are closer together share more common signals, and thus have more similar beliefs. Given that all voters have a common preference for better competency, this effect implies that, on average, the votes of individuals closer together tend to be more similar than those of voters who are far apart. In other words, the network effects of voter communication induce correlations in voting behavior. The next theorem states this result formally. Note here that v describes the minimal number of voters which separate i and j.

Theorem 1: Let $j = i + (v + 1)$, with $v + 1 < N/2$. The ex ante probability that i and j vote for the same candidate is decreasing in v.

Outline of Proof:

- Lemma 2 proves that as v decreases the number of common signals X increases.
- Claim (a): More shared signals produce more similar beliefs about candidate competency.
- Claim (b): Similar beliefs about candidate competency produce correlated votes.

Proof(a): Without loss of generality, suppose candidate A is the more competent candidate. Let each voter have X common signals and Y independent signals, where $X + Y = S + 1$. I prove that $P(|N_i - N_j| \geq \alpha)$ is increasing in Y for all α. Note

that the only differences that can exist in the signals i and j receive is due to the Y independent signals. Thus, it is sufficient to show that given two independent and identically distributed binomial random variables B_1 and B_2 with parameters Y and p, $P(|B_1-B_2| \geq \alpha)$ is increasing in Y for all α. Under the normal approximation B_1-B_2 follows a $N(0, 2np(1-p))$. Thus, $P(|B_1-B_2| \geq \alpha)$ approximates $2P(Z \geq \frac{\alpha}{\sqrt{2np(1-p)}})$, which is clearly increasing in Y.

Proof(b): i votes for A if $p_i(\theta = A)\lambda - |x_i - p_A| > p_i(\theta = B)\lambda - |x_i - p_B|$, or if $p_i(\theta + A) > \frac{\delta_i^{AB}/\lambda+1}{2}$, where $\delta_i^{AB} = |x_i - p_A| - |x_i - b_B|$. Conditional on p_A and p_B, which are fixed at this stage of the game, δ_i^{AB} and δ_j^{AB} are independent and identically distributed random variables. Thus given Claim (a), which shows that $p_i(\theta = A)$ and $p_j(\theta = A)$ are less likely to be different the more signals are common to i and j, it is evident that the probability they vote for the same candidate is increasing in the number of shared signals.

2.6 Candidate Polarization

Given voter behavior as outlined in the previous sections, we now turn to analyzing candidate policy choices. The candidates strategically choose policy platforms in order to maximize their expected utility. Recall that candidates prefer winning office under any policy commitment to losing the election but, conditional on winning, prefer to implement a platform close to their own ideology.

The strategy space is $P = [0, 1]$. A mixed strategy for candidate C is $\sigma_C \in \triangle P$. A mixed strategy profile is $\sigma = (\sigma_A, \sigma_B) \in \triangle P \times \triangle P$. The expected payoff functions are $\pi_C : P \times P \to R$. Let $\pi'_C : \triangle P \times \triangle P \to R$ be the mixed strategy extension of π_C. Then $\pi'_C(\sigma) = \int_{p_A} \int_{p_B} \sigma_A(p_A)\sigma_B(p_B)\pi_C(p_A, p_B)$.

I will restrict my attention to symmetric equilibria and to cases where $\lambda \in [0, .25]$.

Theorem 2: σ^* is an equilibrium, where σ_A^* places weights $(p, q, 1-p-q)$ on the locations $(.50 - 2\lambda, .5 - \lambda, .5)$ where $p = \lambda/.50$ and $q = \frac{(.25-\lambda^2/.50)}{.25+\lambda}$ and σ_B^* places weights $(p, q, 1-p-q)$ on the locations $(.50 + 2\lambda, .5 + \lambda, .5)$.

Proof: I start by finding candidate A's best response to σ_B^*.

Suppose A plays a mixed strategy σ_A which puts positive probability on some $z \in (.50 - \lambda, .50)$. I want to show that σ'_A is a better response than σ_A, where σ'_A transfers any probability weight that σ_A has attached to z to the strategy .50, and is otherwise identical. That is, I want to show that that $\pi'_A(z, \sigma^*_B) < \pi'_A(50, \sigma^*_B)$. This is clearly seen as $\pi'_A(z, \sigma^*_B) = 1/2(\pi'_A(.50 - \lambda + \epsilon, \sigma^*_B)) = 1/2(.50 + \lambda - \epsilon) = .25 + \lambda/2 - \epsilon/2 < 1/2(.50) + 1/2(p * .50) = 1/2(.50) + 1/2(\lambda/.50)(.50) = .25 + \lambda/2 = \pi'_A(50, \sigma^*_B)$, where $\epsilon = z - (.50 - \lambda) > 0$. A similar argument holds for any continuous density putting mass on $(.50 - \lambda, .50)$.

Now suppose A plays a mixed strategy σ_A which puts positive probability on some $y \in (.50 - 2\lambda, .50 - \lambda)$. I want to show that σ'_A is a better response than σ_A, where σ'_A transfers any probability weight that σ_A has attached to y to the strategy $.50 - \lambda$, and is otherwise identical. That is, I want to show that $\pi'_A(y, \sigma^*_B) < \pi'_A(50 - \lambda, \sigma^*_B)$. This is clearly seen as $\pi'_A(y, \sigma^*_B) = 1/2(.50 - 2\lambda + \epsilon)(p + q) = (.50 + 2\lambda - \epsilon)\frac{(\lambda/2 + .25)}{(.50 + 2\lambda)} = \frac{(.50 + 2\lambda - \epsilon)}{(.50 + 2\lambda - \epsilon)}(.25 + \lambda/2) < .25 + \lambda/2 = \pi'_A(.50 - \lambda)$, where $\epsilon = y - .50 - 2\lambda > 0$. A similar argument holds for any continuous density putting mass on $(.50 - 2\lambda, .50 - \lambda)$.

Now suppose A plays a mixed strategy σ_A which puts positive probability on some $x \in [.50 - 4\lambda, .50 - 2\lambda)$. I want to show that σ'_A is a better response than σ_A, where σ'_A transfers any probability weight that σ_A has attached to x to the strategy $.50 - 2\lambda$, and is otherwise identical. That is, I want to show that $\pi'_A(x, \sigma^*_B) < \pi'_A(50 - 2\lambda, \sigma^*_B)$. This is clearly seen as $\pi'_A(x, \sigma^*_B) = 1/2(\pi'_A(.50 - 4\lambda + \epsilon, \sigma^*_B)) = 1/2(p(.50 + 4\lambda) = \lambda/2 + (\lambda^2)/.25) < .25 + \lambda/2 = \pi'_A(.50 - 2 * \lambda)$ where $\epsilon = x - (.50 - 4\lambda) > 0$. A similar argument holds for any continuous density putting mass on $(.50 - 4\lambda, .50 - 2\lambda)$.

Suppose A plays a mixed strategy σ_A which puts positive probability on $x' \in [0, .50 - 4\lambda)$. There are no points in x' which would produce a win for A and thus playing any mixed strategy which put positive probability on x' would imply including points which had zero expected utility. Thus, candidate A will never select a strategy profile that would place any positive probability on any such x'.

Therefore, there exists support for A's best response only on the points $.50 - 2\lambda, .50 - \lambda$, and .50. Moreover, since the expected utility from all three points are identical, any mixture on these points is a best response. Suppose A plays these

points with probabilities $(p, q, 1-p-q)$. Then, by symmetry, B's best response is σ_B^*. Thus σ^* is an equilibrium.

The important conclusion from this equilibrium, and that will be clarified in the following figures, is that candidate polarization increases with λ. This is because as voters reduce their uncertainty about which candidate is more competent — and thus anticipate receiving greater utility from candidate competency — they will place more and more value on candidate competency when making their vote choice. This is comparable to the consequence of living in a highly homogeneous neighborhood where all neighbors agree on which candidate is more competent. In this case, the voter will have greater certainty in her beliefs about competency and thus attach more weight to competency in expectation — with the "real world" implication that the entire neighborhood might vote for a candidate different than the one they would support if they had full information.

2.6.1 Comparative Statics

In Figure 2.2 the points are plotted that receive positive probability from each candidate in equilibrium. Two of the values which the candidate places positive probability upon are moving increasingly away from zero (the location of the median voter) as λ, the expected value of competency, increases. That is, note that as λ increases the furthest point away from the median voter $(.50 \pm 2 * \lambda)$ is moving further from zero as is the middle point in each candidate's strategies $(.50 \pm \lambda)$. Note, however, that this particular fact does not necessarily imply that as λ increases we will observe increase in candidate polarization, in the sense of expected distance between candidates' platforms and the median voter's ideal point, because the third value which the candidates place positive probability upon is in fact zero. Thus, the figures which follow examine how an increase in λ will affect the relative probabilities of the platform locations in the support of the candidates' equilibrium strategies and thus will help to determine how an increase in λ affects candidate polarization in expectation.

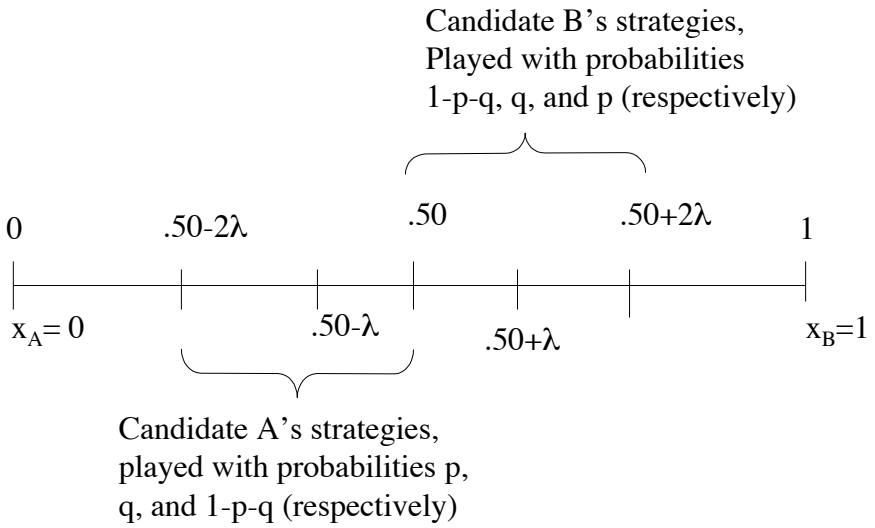


Figure 2.2: Candidate's Strategies in Equilibrium

In Figure 2.3 the change in the probabilities of the three different strategies is plotted as a function of λ. Note that as λ increases the most extremist strategy $(.50 \pm 2 * \lambda)$ is played increasingly often. However, the middle strategy $(.50 \pm \lambda)$ is played with decreasing probability, and the strategy which represents the preferences of the median voter is played with increasing probability (.50). Therefore, the effect on expected voter welfare is not obvious. Note, however, that at all possible values of λ both the middle or extreme strategies are played with higher probability than the most moderate strategy, the strategy which accords with the median voter theorem.

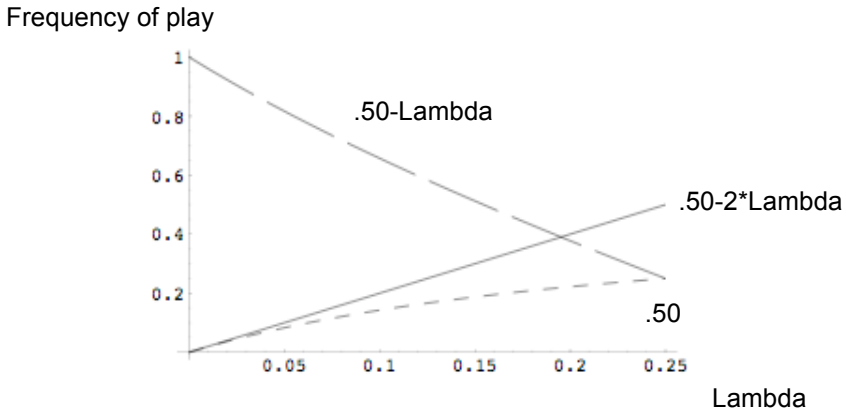


Figure 2.3: Comparative Statics of Candidate Strategies as a Function of λ

Note: The solid line represents the extreme strategy, the long dashed line represents the middle strategy, and the short dashed line represents the moderate strategy.

Even with the addition of the valence dimension the majority of voters would prefer candidates who did not locate at increasingly extreme policy positions. Thus, as λ increases, voter welfare decreases. The expected polarization is calculated by determining the expected candidate position, which is a function of both the probabilities p and q as well as λ. Consider Figure 2.4, which shows that the expected distance from the median voter to the candidates' platforms increases with λ. Thus the candidates do in fact exhibit increased polarization, and the consequence of this increased polarization is the collapse of the equilibrium predicted by the median voter theorem. Again, higher values of λ are associated with greater certainty on the part of the voter as to which candidate is more competent — more certainty results in higher expected utility for the voter from choosing the more competent candidate based upon her signals. For candidates, this means that voters will attach greater importance to competency, which then allows the candidates to adopt more polarized ideological positions.

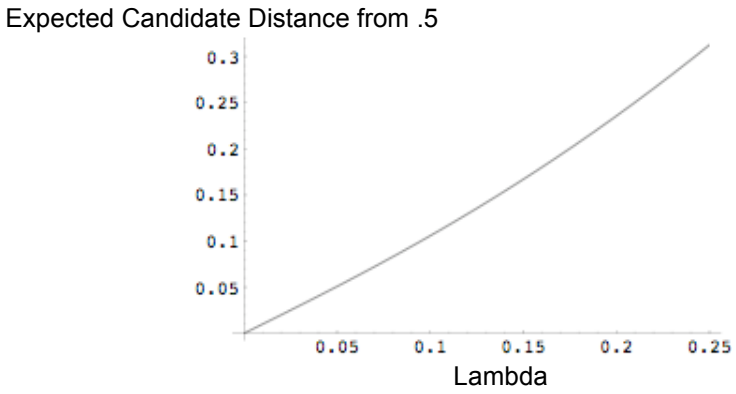


Figure 2.4: Expected Candidate Distance from .5 as λ Increases.

Note: As λ increases, the expected candidate location moves further away from .50.

2.6.2 Uniqueness

No evidence beyond intuition is presented here that the equilibrium under discussion is unique. Future work will attempt to prove the conjecture that any point that has positive support in the best-response correspondences will have distance at least λ from any other such point. The intuition for this conjecture is that there is some relationship between competency and ideology that is optimal — at some level it is possible that playing even more extreme ideological strategies yield low-enough payoffs that they are not worth the increased risk in the election for candidates. Note that since there are no pure strategy equilibria, it is also possible to conjecture that any other equilibria would place positive probability on points even further away from .50.

2.7 The Roles of Limited Information and Truth-Telling

This section clarifies the effects of some of the modeling assumptions. In particular, changes in the accuracy of voter's beliefs about competency are seen to affect their relative utilities for competency versus ideology. This is an important question since this analysis shows how changes in voter's information affect the basis on which voting choices are made. Next the assumption that voters communicate their information honestly to their neighbors is analyzed. Here the voters are also assumed to communicate their sincere preference (as opposed to simply their signal). Additionally, voter communication is discussed under majority rule, in which case voters may have incentives to intentionally misrepresent their information, and what effects this has on voting behavior and the policy choices of candidates.

2.7.1 Candidate Polarization with Limited Information

Note that while voters know exactly where the candidates are located, they do not know with certainty which candidate is more competent. Voters factor this informa-

tion into their choice, and this changes the magnitude of their expected payoff from the competency dimension (see Lemma 1). In Figure 2.5 the competency component of utility changes as a function of $S + 1$, the number of signals that each voter has access to. In Figure 2.5 S ranges from 0 to 25 and the value of $\gamma = .10$ is fixed. The expected benefit of competency to each voter is calculated (and thus to the more competent candidate). That is, as the number is signals increases, the voter becomes increasingly more certain about which candidate is relatively more competent. This certainty adds additional weight to the component of their utility that is based upon competency, which thus allows the candidates to move increasingly further away from the median voter's policy platform preferences. Note that this effect is similar to the comparative statics in p as well. As the number of signals increases the voter's expected benefit from competency converges to γ, as expected, since as the number of signals grows the beliefs about competency converge to the truth. Thus any social structure which produces fewer signals reduces the voter's expected benefit of competency, and any structure which produces more signals will increase the voter's expected benefit of competency. When voters truthfully reveal their signals in the communication phase of the game then increasing the number of available signals adds additional weight to competency, but when voters have the option of distorting the message, however, more signals do not necessarily increase the amount of information — and as a consequence, the voters' expected utility.

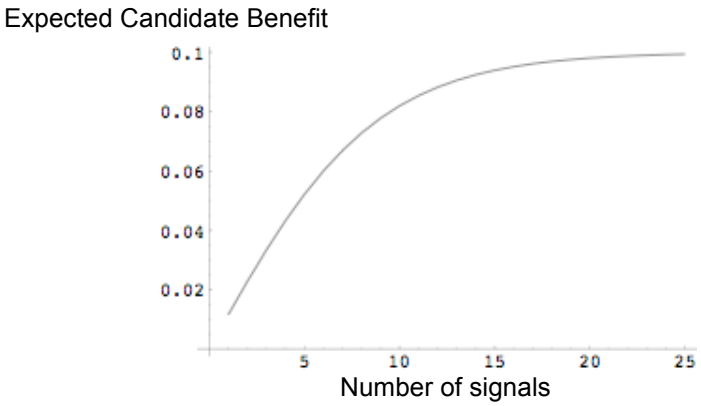


Figure 2.5: Expected Candidate Benefit of Competency as a Function of Number of Signals.

Note: Fixed $\gamma = .1$.

2.7.2 Truth Telling

The following sections analyze candidate location outcomes under a different assumption of voter behavior. Here voters will be assumed to communicate their sincere preference in the communication phase, as opposed to communicating their competency signal. There is a large literature which suggests that voters will indeed communicate based upon partisan bias (Bartels 2002, Gerber and Green 1998, Achen 1992, Achen 1975, Zaller and Feldman 1992) and this behavioral assumption — that voters will communicate their candidate preference and not their competency signal — aligns much more closely with observed voter behavior. Here again voters will have payoffs which are determined based upon for whom they vote. This ensures that voters will not condition their vote upon their pivotality in the election.[5]

Note first that all voters observe each other's ideological preferences, so that a voter's ideology is common knowledge. Note that then voter i observes p_A, p_B, s_i (her private competency signal), and x_j, the ideology of voter j, for all $j \in N$. Voter i can state a message $m_{ij} \in \{A, B\}$ to voter j. That is, she can claim that she observed either $s_i = A$ or $s_i = B$ when communicating with j.

Recall λ is defined as the marginal expected benefit of voting for the more competent candidate. The voters are sorted into three groups. A voter is defined as an *ideological voter* if $||x_i - p_A| - |x_i - p_B|| \geq \lambda$. A voter is defined as a *competency voter* if $||x_i - p_A| - |x_i - p_B|| < \lambda$. Finally, a voter is defined as an *indifferent voter* if $||x_i - p_A| - |x_i - p_B|| = \lambda$.

Lemma 3: Voter j will update her beliefs about competency if voter i is a competency voter.

Proof: i sends j a message m_i which reveals her most-preferred candidate. j has observed the ideology of i, x_i, the candidate locations p_A and p_B and knows the value

[5]Suppose that voter payoffs were based upon the majority winner. This clearly affects their desire to tell other voters the truth when they reveal their private competency signal. With a large electorate, however, it is unlikely that voters will communicate strategically to others within their social context. Additionally, if voters are assumed to communicate sincerely, then even if a voter conditions upon her pivotality (consistent with the setup of Austen-Smith and Banks 1996) when casting her ballot under majority rule she will continue to vote sincerely. Note here that all voters observe each other's ideological preferences and the common prior on the competency of either candidate is 1/2.

of λ. Thus, it is possible for j to calculate which group i will belong to by examining the value of $||x_i - p_A| - |x_i - p_B||$ and λ. If i is an ideological voter, then the message m_i will be based upon i's ideological preference. If i is an indifferent voter, then the messagem_i is equally likely to be based upon i's ideological preference as it is to be based upon her information. Thus, the only informative message which i can send to j is if i is a competency voter, in which case her most preferred candidate is based upon her own signal.

2.7.3 Candidate Polarization with Most-Preferred Communication

Recall that since voters randomly draw their ideological preferences from the uniform distribution, there is no correlation, ex ante, between their location on the circle and their preferences. Given the cases outlined above, communication which consists of the voters revealing their most-preferred candidate has the consequence of dramatically reducing the amount of information available for the competency voters, as now the only voters from whom they can receive useful signals are other competency voters, and being a competency voter is dependent upon ideology. This implies that each competency voter will have less information (so the expected utility of competency is smaller) implying that λ will in fact be smaller than in the situation in which all voters are required to tell the truth. Referring to earlier work with respect to the comparative statics of λ and candidate locations in equilibrium, note that as λ is reduced less candidate polarization is observed.

2.7.4 Information Distribution Changes

Next the distribution of voter ideology changes. Here there is now strong correlation between voter ideology and their geographic location. Assume that after each voter draws their ideology, they sort themselves by ideology, and within ideology, by groups — thus competency voters have competency neighbors, indifferent voters have indifferent neighbors, and ideology voters have ideology neighbors. This produces a circle

with the extremists together and the moderates together.

By considering this special distribution, the aim is to capture the notion that voters may separate themselves into groups in a way that depends on their ideology. In other words, people who live in close proximity to each other, or who have close relationships with each other, are more likely to hold similar political views. Under this extreme form of correlation, for example, every voter on the left half of the circle is a Democrat while every voter on the right half of the circle is a Republican, and within parties, the voters are again sorted by group. This situation is referred to as "voter polarization".

The voters know the polarization setup (by allowing them, as before, to observe each other's types). Now the competency voters, because they are neighbors, will be able to receive a great deal more information via competency signals. They will again have more certainty about which candidate is more competent. Here it is again possible refer to earlier work with respect to the comparative statics of λ and candidate locations. With an increased number of signals λ will increase relative to the uncorrelated case, which will produce an increase in candidate polarization.

2.8 Conclusions and Future Research

This model relates horizontal communication to candidate platform location by describing candidate location as a function of voters' expected information. To the degree to which candidates are able to predict the amount of information that will be available to voters conditional upon their social connections and the amount of polarization, candidates will then be able to strategically choose locations which they prefer. The model formalizes the consequences of particular communication structures on elite polarization and demonstrates that civic discourse across party identification and ideology amongst voters produces more moderate candidate policy platforms. This conclusion is drawn from the interaction between voters' political networks and the behavior of the political elite.

One nice feature of the model is that voters who are moderates are unable to

update their information about candidate competency from communication (when communication is based upon a most-preferred candidate) with those voters from whom they are ideologically polarized. This model captures nicely the correlation that can occur when neighbors agree about which candidate is more competent. It also produces a candidate polarization result that is smaller when the voters themselves are not polarized.

Future research will investigate results allowing the voters to be strategic and consider their behavior when their utility is based upon the majority rule outcomes. Future research may include work in which the voters did not know each other's types. This would have as a consequence voter j under-weighting even more dramatically the competency signals she receives via communication. Another possibility would be to additionally correlation candidate signals with ideology. A final possibility is to let each voter draw a vector with $N-1$ entries which describe the probabilities of friendship with any other voter (or intensity of friendship). Information could be exchanged with these probabilities. Empirical testing of this model will examine the degree to which individually reported communications from other voters (using the 2000 ANES, the 1987 GSS and the 1988 GSS) relates to vote choice, the degree to which geographic proxies for network effects determine vote choice (using election returns from specific geographies), and the amount of information available to both respondents and to those within their discussion network (using the 2000 ANES and data from the Huckfeldt and Sprague 1996 St Louis -Indianapolis Study).

Chapter 3

Discussion: Theoretical Model and Empirical Chapters

Each of the following sections will analyze data which are related to different components of this model. No single data-set is provided in order to empirically test the model directly. However, different parts of the model or assumptions are verified in the sections which follow. In all cases, data is presented about the voter behavior — the empirical analysis discusses only the actions which would occur as a consequence of voter-to-voter communication. These are the assumptions which underly the model — if in fact voters do behave consistently with these assumptions, then it is possible to assume that candidates will behave consistently with the model.

Two get-out-the-vote (GOTV) mobilization campaigns were completed in South Los Angeles, where the door-to-door canvassers were identified as individuals who were residents of South Los Angeles. In both instances individuals are identified as being precinct-residents or not, and the turnout from precincts contacted by residents is higher than the turnout from precincts contacted by non-residents. The script, read by both non-residents and residents, is the same for all canvassers. Thus, the increased turnout is consistent with voters more readily believing information from others within their social framework. These canvassers, however, are not necessarily known personally to each voter, although they do often report seeing their friends while walking in their neighborhoods. Again, it is also possible that this is a "peer-pressured" effect, where voters are more likely to turnout to vote if they believe their

friends will observe their behavior. The voter mobilization experiments validate one of the findings derived from the theoretical model which states that individuals are likely to be socially connected to others of similar types – and that communication with similar types allows individuals to update about a valence dimension. That we observe higher turnout rates with canvassing conducted by home-turf canvassers indicates that indeed individuals who are more likely to be socially connected increases the persuasiveness of communication.

The next section presents results from a laboratory experiment which investigates the degree to which individuals can update their beliefs about candidates when presented with almost no information, but instead a noisy, public signal which closely resembles the "cheap talk" communication phase in the theoretical model. Voters are readily able to do so, and are able to converge upon the candidate midpoint through Bayesian updating. The experimental setup is not identical in that all voters observe the same signal (as opposed to observing a signal only within their own social framework) but the behavioral assumption in the model, that the voters will update their priors from their signals, is validated here.

The final section presents a series of results from three survey data-sets, each of which asks the respondent to identify discussant characteristics and then uses those characteristics as a "treatment" variable to understand the voter's candidate and partisan choices. Assuming that the respondent's answers to the discussant characteristics are not endogenous to the respondent's characteristics, then it is possible to observe large effects of discussant characteristics on candidate choice and party identification. This is consistent with the results of the model which claim that there will be correlation between social connections and vote choice as a consequence of communication.

Each of these sections have findings which are consistent with the assumptions about voter behavior in the theoretical model — that individuals are Bayesians (laboratory experiment chapter), that they are likely to be connected to individuals of similar types (survey data chapters), and that they communicate about politics (survey data chapters). The preliminary results in the theoretical model are also validated

in these chapters – that voters of similar types are more persuasive in their communication (GOTV experiment chapters) and that voters who are socially connected are likely to have correlated choices, controlling for their preferences (survey data chapters). If candidates are aware of these behaviors, then it is possible to assume that they would consider the voter choice correlation in their own choice of policy platform positions — resulting in elite polarization.

While the empirical results are interesting by themselves, they provide collectively a particularly strong argument for the fact that individuals will condition their political behavior based upon the actions of others within their political network. The traditional calculus of voting, which considers each individual as an independent agent maximizing her preferences, is simply an insufficient explanation. Here individuals behave consistent with others within their social framework. Whether those behaviors are due to a change in the reliability of information when information is procured from a friend or whether the individuals are simply trying to behave like others within their social group is not resolved here. However, what is clear is that political preferences and behaviors condition upon group preferences.

Chapter 4

Community-Based Mobilization: GOTV and SCOPE in South Los Angeles, November 2006

4.1 Summary

This chapter analyzes the efficacy of the November 2006 GOTV campaign of Strategic Concepts in Organizing and Policy Education organization (SCOPE). SCOPE's campaign focused on using community-based precinct walking as a way to mobilize low-propensity voters. Voters were randomly assigned to have a SCOPE volunteer visit their home and speak with them about the upcoming election — this randomization process happens both early in the campaign and then again during the final weekend of the campaign — and the advantage of randomization and control is that we are able to evaluate the efficacy of the campaign. Additionally we are able to evaluate the efficacy of canvassing if the canvasser is a resident of the neighborhood where she is canvassing or not. Here we describe the campaign, the data, and analyze the efficacy of both the early and late campaigns. We find large effects of canvassing during the early campaign but find no statistically significant effect of a recontact during the late campaign. The campaign as a whole has a 6% increase in turnout, and home-turf canvassing increases turnout an additional 4%. This finding corresponds with the theoretical framework which demonstrates that individuals who are socially connected will be more persuasive.

4.2 Introduction

Politics is often studied as an event by which individuals, in isolation from others, think about their own preferences, the location of the candidates' policy platforms, the probability that their vote is pivotal, and then decide whether or not to vote. However, this is a calculus which is unlikely to produce the high levels of turnout observed in general elections. Furthermore, it seems unlikely that these are the conditions under which individuals are determining whether or not to vote. Individuals are unlikely to make these choices in isolation — at a minimum their choices are likely to be observed by their immediate household, and potentially by others in their community. Additionally, individuals are likely to determine whether or not it is important for them to cast a ballot — either in terms of pivotality, in terms of sending a symbolic message, or by learning information about particular ballot propositions, for example. Particularly individuals are most likely to listen to election information from individuals whom they believe have their best interests at heart or have similar interests themselves (Downs 1957) and this paper analyzes the effects of communication about an election while attempting to leverage those close contacts.

Individuals in this experiment were randomly assigned to either be contacted by a canvasser from South Los Angeles who was likely to match the subject on racial and language characteristics or to receive no contact. The canvasser went to the individual's home and spoke for a few minutes about the upcoming election and, in detail, discussed one of the ballot propositions. The randomization process was repeated during the last week of the campaign, with some individuals then randomly assigned to receive a second treatment. Voter turnout is then observed in the election to determine the effect of one or two canvasser contacts.

The individuals in this study are new or occasional voters (defined as having participated in fewer then four of the last ten general elections). This is a group who we expect to be particularly vulnerable to campaign contacts (Niven 2004). Furthermore, we expect these voters to be additionally receptive to treatment as the demographic profile of this group of voters is such that these are voters who are likely to pay a

higher cost for procuring election information and for participation — for example, many of these voters are highly mobile, making registration, locating a polling place, and getting to the polls more costly — and thus the canvasser contact dramatically reduces the cost of voting. A significant body of research has documented that face-to-face mobilization is likely to be most effective (Green and Gerber 2001; Green and Gerber 2000; Green, Gerber and Nickerson 2003) — here we extend that research to examine the effects of mobilization using canvassers who are most likely to fit into the Downsian "trusted source" mold and apply this treatment to a group of voters whose vote history is such that they are unlikely to participate without canvasser contact.[1] We expect to observe significant effects from contacts, however, as these occasional voters are also most likely to be persuaded to participate (Niven 2004), and we believe they are most likely to be persuaded by someone from their neighborhood. It is impossible, however, in this study to distinguish the effect of neighborhood contact from simply obtaining additional information — and it is possible that simply procuring the additional information about the election, as opposed to receiving it from a potentially trusted source, is the causal mechanism (Lassen 2005). With that caveat in mind, however, this experiment provides a unique opportunity to evaluate mobilization.

Because each canvasser is identified geographically, it is possible to measure the effect of home-turf canvassing. We hypothesize that, consistent with the theoretical framework set out in the first chapter, that individuals are more likely to be persuaded by an individual with whom they are socially connected. Consequently we anticipate increases in turnout from home-turf canvassing.

4.3 SCOPE

The voter mobilization effort in this study was conducted by the grassroots political organization Strategic Concepts in Organizing and Policy Education (SCOPE) whose

[1] There is some evidence that in terms of cost effectiveness, phone calls are also likely to produce similar turnout results (Nickerson 2006).

offices are in South Los Angeles.[2] One of their stated mission strategies is to "build models of increasing civic participation where poor and disadvantaged communities and communities of color can become active participants in public policy-making and initiatives that impact their lives." The intent of their mobilization campaign in the fall of 2006 was twofold — to ensure that low propensity voters were contacted and turned out to vote, and to encourage the registered voters to support SCOPE's "Green Jobs" campaign, a component of which was to cast a vote in favor of Proposition 87.[3]

One unique component of this campaign is the particular voting population which SCOPE intended to contact — SCOPE's campaign was dedicated to mobilizing only new and occasional voters — defined as those voters who had participated in fewer than four of the last ten general elections. They intended to send representatives from SCOPE to the doorways of these voters, to both convince these voters to vote in the election and to convince them to support Proposition 87. SCOPE has spent years recruiting volunteers from the South Los Angeles area who are willing to walk in the precincts prior to the election. One of the other unique aspects of this campaign is the population of canvassers — the walkers are often residents in the areas where SCOPE is campaigning.

SCOPE began their campaign on Saturday, October 28th, 2006, and proceeded to organize their volunteers to walk each Saturday and the Sunday, Monday and Tuesday immediately prior to the election, November 7, 2006. A paid team (many of whom were recruited from the SCOPE volunteer population) walked through precincts 6 days each week excluding Sundays. This process continued until the final week prior to the election, when the final "GOTV Weekend" campaign began and SCOPE re-contacted earlier contacts.

One significant concern with this process was the political nature of the canvassing,

[2]Their website describes their mission and ongoing campaigns; http://www.scopela.org/agenda/index.html

[3]Proposition 87 was a constitutional amendment which proposed establishing a program to reduce petroleum consumption by establishing a tax on California oil producers. It read "Should California establish a $4 billion Clean Alternative Energy Program to reduce California's oil and gasoline consumption by 25 percent through incentives for alternative energy, education, and training?" It failed, 45.3% Yes to 54.7% No.

which consisted of administering a political survey about Proposition 87. The survey process was confusing both to the volunteers and, according to many of the volunteers, to the voters as well.

Approximately 20 volunteers arrived each Saturday, participated in a brief training session, and spent approximately 2 to 3 hours canvassing in the local neighborhoods. They then returned for lunch and were debriefed.

During the campaign each canvasser was asked to complete a survey upon return to campaign headquarters. The content of this survey was geared to evaluate whether or not the canvasser had lived in the precinct or was particularly familiar with the precinct. Each volunteer also initialed each contact, making it possible to evaluate whether or not there was a "home-turf" canvassing effect. However, during the course of the campaign, these surveys were misplaced. As a consequence, each canvasser's address is geocoded and located in a particular precinct and zipcode.

4.4 Experimental Setup and Data

Among the new and occasional registered voters in 50 precincts in South Los Angeles, we drew a control group which consisted of approximately 20% of the eligible voters in our sample. South Los Angeles is located west of the 110 Freeway and south of the 10 Freeway in an area that has historically been economically depressed. As a consequence, however, South Los Angeles is particularly well-suited to evaluate home-turf canvassing as many of the residents do not own cars (many of the volunteer canvassers themselves did not own cars). This is rare in Los Angeles — and it increases the likelihood that individuals will know others within their neighborhood. The randomization is conducted by household, and households with more than 3 individuals are removed from the experiment. In our initial randomization, we assigned 3578 individuals to the control group and those remaining were assigned to the treatment group.

In the final week before the election SCOPE opted to re-contact some their earlier contacts. We again randomized the assignment treatment and control among the

voters that SCOPE had determined would be advantageous to re-contact, removing 360 individuals for the control group. This second contact then provides another randomized treatment and thus we are able to evaluate the effectiveness of a second contact in the campaign. These assignments are described in Table 4.1.[4]

Table 4.1 Goes Here

Based upon records kept by the canvassers, we are able to discern which of the individuals in the sample received either one or two contacts. The contacts are tabulated by the random assignments. Of the individuals assigned to receive the first contact, 3907 were successfully contacted. A total of 2 individuals were inadvertently contacted from the control group. Of the individuals assigned to receive the second contact, 481 were successfully contacted, and 11 individuals were inadvertently contacted. Additionally, SCOPE placed doorhangers on the doors of many of the individuals in this sample. The rate of doorhangers (where one doorhanger is placed for each household) is documented as well. Fortunately no individuals in the control group for the early campaign received a contact, so they will be useful as a comparison. Finally, it is possible to identify which of these individuals voted in the November 2006 election using turnout data provided by the Los Angeles Registrar Recorder.

Each contact was initialled by the canvasser. Consequently, it is possible to identify which canvasser contacted each individual in our sample. Each canvasser provided their street number and address on a sign-in sheet at the beginning of each GOTV walk session. Using these addresses, we geocode each canvasser into a precinct and a zipcode. We then classify each individual within a zipcode and map these into precincts. Thus it is possible for us to discern whether or not an individual was contacted by someone who lives in their neighborhood — hereafter referred to as "home-turf canvassing".

[4]The data provided to us by SCOPE indicated contacts which they had scanned into their contact dataset using a bar scanner. However, our duplicate efforts discovered a total of 942 discrepancies. We include data for all individuals for whom SCOPE did not provide a walksheet or a walker code (for a total of 399) but when otherwise our scanned data disagrees with the SCOPE data, we use ours instead.

The principle advantage of the randomization is to ensure that in fact the mobilization campaign is the cause of turnout. If the mobilization efforts were not randomized, then several problems could emerge with the data — either the ability to successfully administer treatment could be correlated with the probability that an individual turns out to vote (being home during the day, for example, might correlate with age — and older voters are more likely to turn out to vote (Wolfinger and Rosenstone 1980) or worse, the organization could select voters to contact whom they believe are likely voters. We avoid both problems by assigning treatment and control groups.

4.5 Methods and Results

We estimate the intent-to-treat effect[5] and the treatment effect[6] on the group which received any contact from SCOPE whatsoever, the group which received one contact, the group which received one contact excluding those precincts which were included in the second round of randomization, and on the group which received the second contact. One disadvantage of incorporating all contacts together is that if there is an effect of a second contact this will bolster the appearance of an effect from any contact whatsoever. However, an examination of those individuals who only received one contact is also flawed. SCOPE selected the individuals which they wanted to re-contact — and many of these were selected because they had already successfully received the first contact. Thus, looking at only those individuals who received the first contact is likely to bias the results downwards. Consequently, here we present all analyses. Fortunately, the marginal benefit of a second contact is not statistically significant — and thus the effect of any contact can be deemed appropriate.

Table 4.2 presents the intent-to-treat and treatment effects from having received any contact from the SCOPE campaign. Here we observe that of the 11789 individuals

[5]The intent-to-treat effect is defined as the observed difference in turnout between those assigned to the treatment and control groups. If the contact rate is 100%, the intent-to-treat effect is identical to the treatment effect. Generally this is not the case, however, and to calculate the treatment effect we must adjust for the contact rate.

[6]The treatment effect is defined as the intent-to-treat effect divided by the contact rate.

who are eligible for treatment, only 5341 are actually treated — a contact rate of 45%, which is particularly high. The SCOPE canvassers did not attempt to contact every household in the treatment list, but many of the households which they did attempt to contact were not home or were unwilling to open their doors. Canvassers reported being unable to contact individuals for a variety of reasons, ranging from that household having a mean dog in the front yard to being unable to convince the individual to open their front door. One of the reasons it is essential to use the randomization assignments in the results is that there may be systematic differences between those individuals who are able to receive the mobilization "treatment" and those who are not. Given that we have so few covariates in the voter file, we are thus quite reliant upon the use of randomized experimentation. Looking at only the raw differences in the percent voting among treatment and control, it is clear that there is in fact an observable difference between the voting rates, 36.68% voting amongst voters within the treatment group and 33.68% voting amongst voters within the control group. Thus, controlling for the contact rate, this provides an estimated treatment effect of 6.63%.

Table 4.2 Goes Here

Table 4.3 presents identical calculations, except on the subset of the sample which was treated only once and comparing this group to the control group. This essentially has the disadvantage of potentially biasing the results downwards — it is possible, for example, that SCOPE has chosen the most likely voters to re-contact by re-contacting those who were successfully contacted originally. We find that the estimated treatment effect is smaller (close to 5.4%). We also examine this calculation excluding those precincts which had observations drawn for the second-wave of contact in Table 4.4 and here find an enormous effect of first-wave contact, almost 9.5%.

Table 4.3 and Table 4.4 Go Here

We next calculate the effect of a second contact during the final GOTV weekend (the last weekend prior to the election). The GOTV estimate should be considered

the additional marginal benefit of re-contacting an earlier contact. Our results are displayed in Table 4.5. Here it appears that the fraction of individuals who voted from the treatment group is in fact lower then that of the control group — however, incorporating the very low contact rate demonstrates that there is simply not enough statistical power to make any type of prediction regarding the effect of a second contact. The estimated treatment effect, as well as the intent-to-treat effect, is not statistically significant.

Table 4.5 Goes Here

Before proceeding to any additional analyses, we evaluate to what extent our treatment and control assignments are balanced in terms of observable covariates. In the dataset we observe partisan registration, age, vote history, a surname classification for ethnicity, and gender. This data-set does not incorporate many control variables, which makes randomization essential, but it is important to ensure that the covariates that are observed are balanced across treatment and control. In Table 4.6 the difference in means for each observed covariate is presented. Two variables are not perfectly balanced — for each contact group, the vote history mean is different and statistically significant between treatment and control; and for the second contact group, the decline-to-state registration is different and statistically significant between treatment and control. As such, we will calculate effects incorporating the covariates as well as without them so as to ensure that the covariates are not themselves driving the results. If we had observed more covariates about each individual voter, we would be able to better construct a model which would allow us to determine which covariates, for example, increased the probability that an individual was able to successfully receive the mobilization treatment. We would also be able to consider using other methodological techniques in order to analyze the effects of treatment, such as using a matching technique combined with linear regression, and would not be as reliant upon randomized experimentation. However, as we are in a case where there are few observable covariates, we rely upon our treatment assignment as an instrument to determine treatment effects.

Table 4.6 Goes Here

We evaluate to what extent the door-to-door contact treatments were able to successfully mobilize a particular individual. We do so using two-stage-least-squares estimates (2SLS), where we want to estimate the effect of contact on whether or not an individual votes in the election but the application of the contact treatment is nonrandom. In order to implement this analysis we construct instrumental variables for the contact with their respective randomization assignments — this process eliminates the correlation between the implementation of the campaign and the error terms. Both of our instruments here are fairly good — the correlation between the treatment assignment and contact for the first contact is .32, and the correlation between the treatment assignment and the second contact is .45. Note that we are assured that there is no correlation between unobserved covariates and assignment, as the assignment was completed randomly prior to conducting the experiment. We also incorporate a fixed effect for each precinct. Our results are displayed in Table 4.14. There is a statistically significant and positive effect of receiving any contact for all methods of estimation — whether we include the control variables[7] or incorporate precinct fixed effects.

Table 4.7 Goes Here

We next examine the effect receiving the second wave contact. We again present three models — the two-stage-least-squares estimate, the two-stage-least-squares estimate with covariates, and then finally the two-stage-least-squares estimate with precinct fixed effects. Again none of the effects are statistically significant, although they are all negative. Note again here that we do calculate robust standard errors (we expect clustering to occur at the household level).

Table 4.8 Goes Here

[7]We include indicators for whether or not the age or female control variables are missing and then fill in values for age and female to match those missing spaces. This allows us to directly compare the estimates of any contact across all models.

We next present some observational tables on the difference between home-turf canvassing effects vs not-home-turf. First, in Table 4.9, we simply present the difference in mean contact and turnout rates between precincts which experienced home-turf canvassing vs those which did not. We observe a slightly higher (and statistically different) contact rate for the home-turf precincts as well as a slightly higher (but not statistically significant) turnout rate for the home-turf precincts. This suggests that home-turf canvassers are more likely to have individuals open their doors, but are not more likely to persuade their contacts to vote.

Table 4.9 Goes Here

However, our next analysis demonstrates that there is a statistically significant improvement when the contact is conducted by a home-turf canvasser. Table 4.10 presents coefficients from a logistic regression where the dependent variable is whether or not the individual voted. These estimates are calculated only within the individuals who were successfully contacted — as the home-turf canvassing was not randomly assigned, this is the best initial way to observe whether or not there was a home-turf effect. The first model aggregates the home-turf canvassers to the zipcode level, the second model aggregates the home-turf canvassers to the precinct. For the zipcode level model, we observe a positive and statistically significant effect of home-turf canvassing. For the precinct level model, we observe a positive effect but not one that is statistically significant. These estimates may actually underestimate the effect of home-turf canvassing if indeed home-turf canvassers are also more successful at convincing individuals to open their doors (and thus increasing their contact rates).

Table 4.10 Goes Here

To demonstrate that in fact home-turf canvassers will not have higher contact rates we plot the contact rate, by precinct, in Figure 4.1. Here each green dot represents a precinct, with the yellow bars representing the 95% confidence interval for the precinct contact rate. The x-axis demonstrates the intensity of the home-turf effort

— the fraction of individuals within that precinct who were contacted by a home-turf canvasser. The blue line is fitted through the green estimates, with the grey shading demonstrating the 95% confidence interval for the fitted line. The blue line is essentially flat; it slopes downward very slightly but is effectively pulled downwards by the one lone estimate with over 60% home-turf canvassing. Thus we may conclude that home-turf canvassers will not have higher contact rates.

Figure 4.1 Goes Here

Given that we observe home-turf canvassers will not have higher contact rates, we next plot the ITT and ATE estimates by precinct against the percent of home-turf canvassing. Again the green dots demonstrate the precinct-level estimates, with the yellow bars representing the 95% confidence interval for each precinct estimate. The x-axis demonstrates the intensity of the home-turf effort — the fraction of individuals in that precinct who were contacted by a home-turf canvasser. The blue line is fit through the green dot estimates, with the grey shading demonstrating the 95% confidence interval for the fitted line.

Figure 4.2 and Figure 4.3 Go Here

These figures demonstrate that as the intensity of the home-turf effort increases, the effectiveness of the campaign increases. Both blue lines are remarkably upward-sloping and in fact have intercepts extremely close to zero. This provides powerful support for the hypothesis that home-turf canvassing is more effective at increasing turnout.

Our final sets of tables are to provide additional support for Figure 4.2 and Figure 4.3. Table 4.11 breaks the data into two sets — the precincts which had some home-turf canvassing and those which did not — and repeats the 2SLS estimates across the two models (with contact and incorporating control variables) and finds that for all estimates of the effect of contact, the home-turf coefficient is positive and statistically significant while the not-home-turf coefficient is not statistically different from zero in the model which incorporates the control variables and is smaller in the

simplier model. This provides additional evidence that in fact home-turf canvassing is better at increasing turnout, and suggests that canvassing without home-turf canvassers has the potential to have zero impact on turnout.

Table 4.11 Goes Here

We next explore the possibility that the home-turf effect is due simply to a set of particularly effective walkers who were also walking in their own precincts. Thus we estimate another set of models, two of which incorporate fixed-effects for walkers within the population of individuals who were all successfully contacted in Table 4.12. Across all models there is a positive and statistically significant effect of home-turf contact, regardless of the inclusion of walker fixed-effects or control variables. In particular, controlling for the walker fixed-effects, the estimate of the impact of the home-turf contact increases dramatically.

Table 4.12 Goes Here

Finally we calculate a meta analysis by precinct of the 2SLS estimates to evaluate the effectiveness of campaign and the effect of home-turf canvassing in Table 4.13 and Table 4.14. This method allows us to incorporate the fixed effects of each precinct without the assumption of homoskedasticity which we are concerned with, as the campaign was implemented with different contact rates across precincts (and in particular the second wave was implemented much less thoroughly across precincts). We perform this analysis both simply on the 2SLS estimates of any contact and then again including a variable which describes the fraction of contacts within the precinct which were canvassed by home-turf canvassers.

Table 4.13 and Table 4.14 Go Here

Table 4.13 presents results that are effectively identical to those presented earlier. Here the estimate for any contact is again statistically significant and is approximately 5.6%. The results in Table 4.14 provides us with a measurement of the effectiveness of home-turf canvassing, however. The coefficient here is positive and statistically

significant — and states that if a campaign were to increase the intensity of home-turf canvassing by 10%, they would increase the fraction of individuals who voted by 30%.

4.6 Conclusions

The analysis above has demonstrated that SCOPE was able to mobilize a group of low-propensity voters to turnout to vote with very high rates of success. A 6% increase from a single contact demonstrates that while this is a demographic which typically has not participated in politics, it is indeed possible to mobilize this group to cast a ballot. SCOPE contacted new and occasional voters by using individuals from those neighborhoods, and one potential explanation of their success is their choice of canvassers. Across all methods of estimation, there is a positive and statistically significant effect of canvasser contact.

This campaign has demonstrated that, consistent with the rest of the voter mobilization literature, door-to-door canvassing is likely to have an effect on turnout. Unfortunately, due to the small number of voters who were contacted during the second GOTV phase, the results are unable to provide definitive estimates on the marginal effect of the second contact. However, initial estimates indicate that this is also likely to be positive, and future experiments should be completed to help analyze the timing component of contacts during campaigns. Campaigns typically have additional volunteers during the final weekend, and if a second contact is effective, it would be useful for campaigns to know whether to allocate their resources to making new contacts or to re-contacting earlier ones. From these results alone, it is clear that a single contact does have the ability to mobilize voters.

Mobilization campaigns are particularly rewarding when they enable a group of voters to express their political preferences who otherwise might not have participated in the democratic process. In particular, a 6% increase in turnout is a large enough increase that it has the potential to change election outcomes, which in turn has the potential to make elected representatives more responsive to the registered voters'

preferences.

Here we have also observed a significant increase in turnout as a result of home-turf canvassing. This increase (an additional 4% above the base increase) is dramatic — when the graph is presented which demonstrates that with additional home-turf contacts turnout continues to increase, this presents compelling evidence for community-based mobilization campaigns. This additionally suggests that home-turf canvassers are more persuasive (and a quick look at the walker-fixed-effects table indicates that it is not merely that there were more persuasive walkers who *also* walked in their home-turf, but rather that walkers in their home-turf are most effective) which is consistent with the theoretical model that individuals are best able to communicate with individuals with whom they are socially connected (and that within those social connections, there is likely to be correlation of preferences).

These results also argue for campaigns or organizations to have a regular presence in a community — it is not sufficient to simply produce a set of walkers to contact a group of low-propensity voters, but instead it is necessary to recruit individuals from those neighborhoods to contact their neighbors. This argues for the reemergence of the "machine" politics as a way to increase voter turnout and suggests that the accumulation of social capital (community-based canvassing) is key to establishing political participation.

4.7 Future Work

Future research will examine election returns for Prop 87 to attempt to determine if there is an increase in Prop 87 "yes"-voting in precincts where there was high levels of campaigning by SCOPE. This is particularly interesting as the SCOPE mobilization campaign consisted not only of a GOTV campaign but also an issue campaign. If Prop 87 returns are higher, this suggests that additional work should be done to explore the efficacy of partisan or issue-based mobilization campaigns.

4.8 Tables and Figures

Table 4.1: Treatment and Control Assignments and Outcomes

Assignment	Total Obs.	First Contact	Second Contact	Doorhanger	Voted
First Wave Treatment Only	9356	2949	32	1342	3288
First Wave Control Only	3578	2	0	0	1205
First and Second Wave Treatment	2073	2040	548	890	875
Second Wave Control and First Wave Treatment	360	352	5	37	161
Experiment Totals	15367	5343	585	2269	5529

Each cell entry describes the number of individuals who fall into the row assignment category.

Table 4.2: Effect of Any Contact (Using First Wave Assignment): Intent-to-Treat Effect and Treatment Effect

N in any first wave treatment	11789
N in control group	3578
N in the treatment group who are actually treated	5341
N in the control group who are inadvertently treated	2
N who voted in the treatment group	4324
N who voted in the control group	1205
Percent Voting — Treatment	36.68%
Percent Voting — Control	33.68%
Percent Contacted — Treatment	45.30%
Percent Contacted — Control	.06%
Contact Rate	.4524
Estimated ITT Effect	3.0002
SE of ITT Effect	.9064
Estimated Treatment Effect	6.6304
SE of Treatment Effect	2.00

Note that here errors are neither robust nor clustered by household.

Table 4.3: Effect of First Contact Only (Excluding Second Wave Treatment Assignment): Intent-to-Treat Effect and Treatment Effect

N in treatment group	9716
N in control group	3578
N in the treatment group who are actually treated	3301
N in the control group who are inadvertently treated	2
N who voted in the treatment group	3449
N who voted in the control group	1205
Percent Voting — Treatment	35.5%
Percent Voting — Control	33.68%
Percent Contacted — Treatment	33.97%
Percent Contacted — Control	.06%
Contact Rate	.3391
Estimated ITT Effect	1.8201
SE of ITT Effect	.9275
Estimated Treatment Effect	5.366
SE of Treatment Effect	2.7344

Note: These values include individuals who were later assigned to be the control group for the second wave of treatment. Individuals who were inadvertently given the second treatment are included (37 total).

Table 4.4: Effect of First Contact Excluding Second-Wave Precincts: Intent-to-Treat Effect and Treatment Effect

N in treatment group	4907
N in control group	1490
N in the treatment group who are actually treated	1636
N in the control group who are inadvertently treated	2
N who voted in the treatment group	1752
N who voted in the control group	485
Percent Voting — Treatment	35.7%
Percent Voting — Control	32.55%
Percent Contacted — Treatment	33.34%
Percent Contacted — Control	.13%
Contact Rate	.3320
Estimated ITT Effect	3.1537
SE of ITT Effect	1.3939
Estimated Treatment Effect	9.4975
SE of Treatment Effect	4.1977

Note: These values were calculated using the population of individuals from the precincts where there was no second-contact experiment.

Table 4.5: Marginal Effect of Second Contact: Intent-to-Treat Effect and Treatment Effect

N in treatment group	2040
N in control group	352
N in the treatment group who are actually treated	548
N in the control group who are inadvertently treated	5
N who voted in the treatment group	861
N who voted in the control group	156
Percent Voting — Treatment	42.21%
Percent Voting — Control	44.32%
Percent Contacted — Treatment	26.86%
Percent Contacted — Control	1.42%
Contact Rate	.2544
Estimated ITT Effect	-2.1122
SE of ITT Effect	2.8690
Estimated Treatment Effect	-8.3023
SE of Treatment Effect	11.2765

Note: These values were calculated conditioning upon only those individuals who were successfully contacted in the first wave of contacts. This excludes the 41 individuals who were assigned to the second experiment but were not successfully contacted during the first-wave of the campaign.

Table 4.6: Covariate Balance: Difference Between Contact and No Contact Means

Variable	Any Contact	Marginal Second Contact
Democratic Registration	.052* (.007)	.002 (.021)
Republican Registration	-.014* (.004)	-.011 (.012)
Decline-to-State Registration	-.036* (.006)	.011 (.017)
Other Registration	-.002 (.003)	-.002 (.010)
Female	.026* (.011)	-.068* (.032)
Age	4.87* (.29)	.031 (.872)
Vote History	.196* (.017)	.065 (.047)
Latino	-.01** (.005)	.03* (.013)
Number of Observations, Received Treatment	5343	585
Number of Observations, No Treatment	10024	1839
Total	15367	2424
$* = \alpha = .05$ and $** = \alpha = .10$		

Note: Age and female have some missing observations.

Figure 4.1: Contact Rate by Percent Home-Turf Canvassing. Note that each dot represents one precinct.

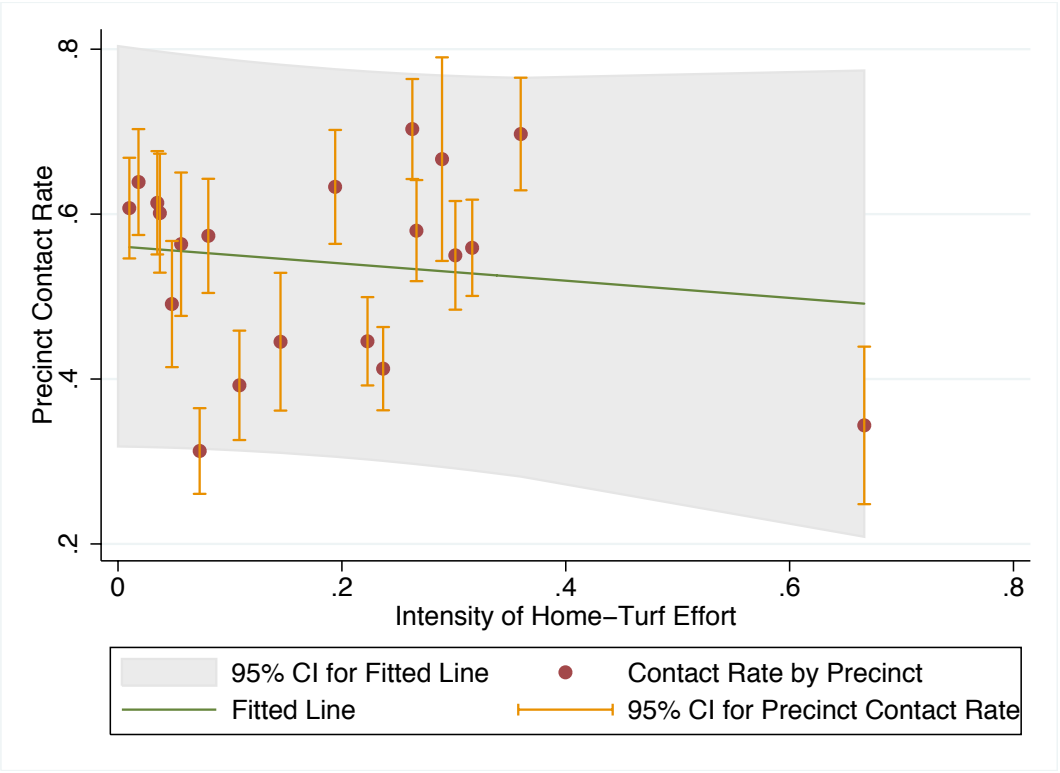


Figure 4.2: Intent-to-Treat Effect by Percent Home-Turf Canvassing. Note that each dot represents one precinct.

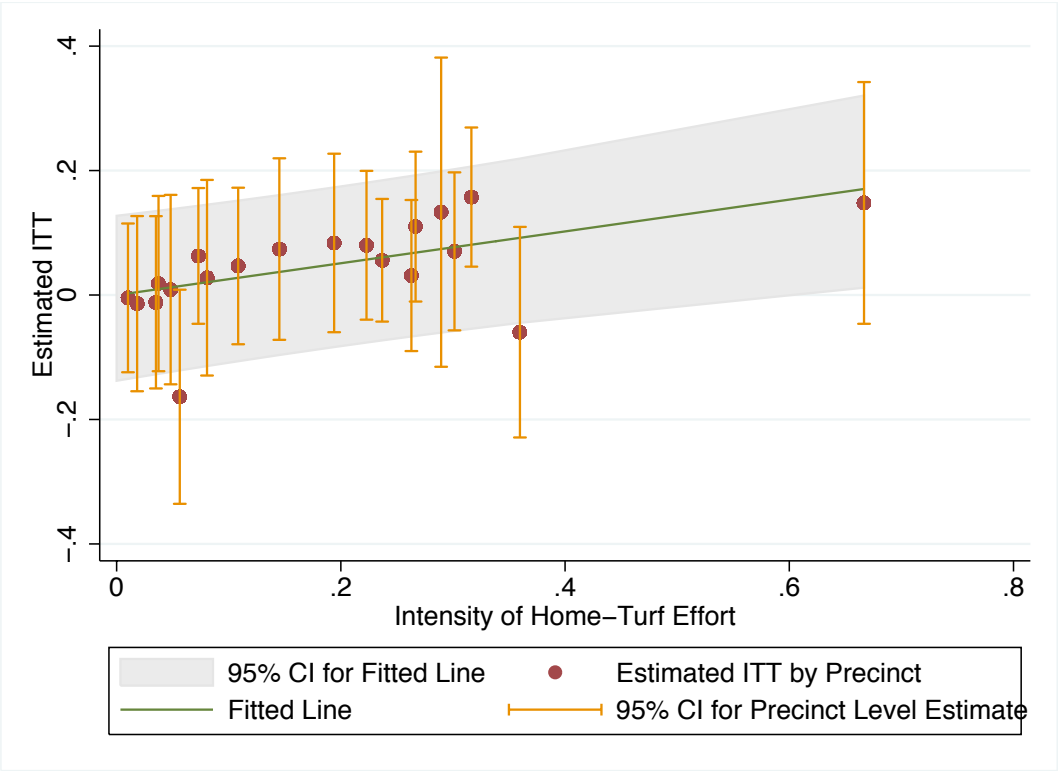


Figure 4.3: Average Treatment Effect by Percent Home-Turf Canvassing. Note that each dot represents one precinct.

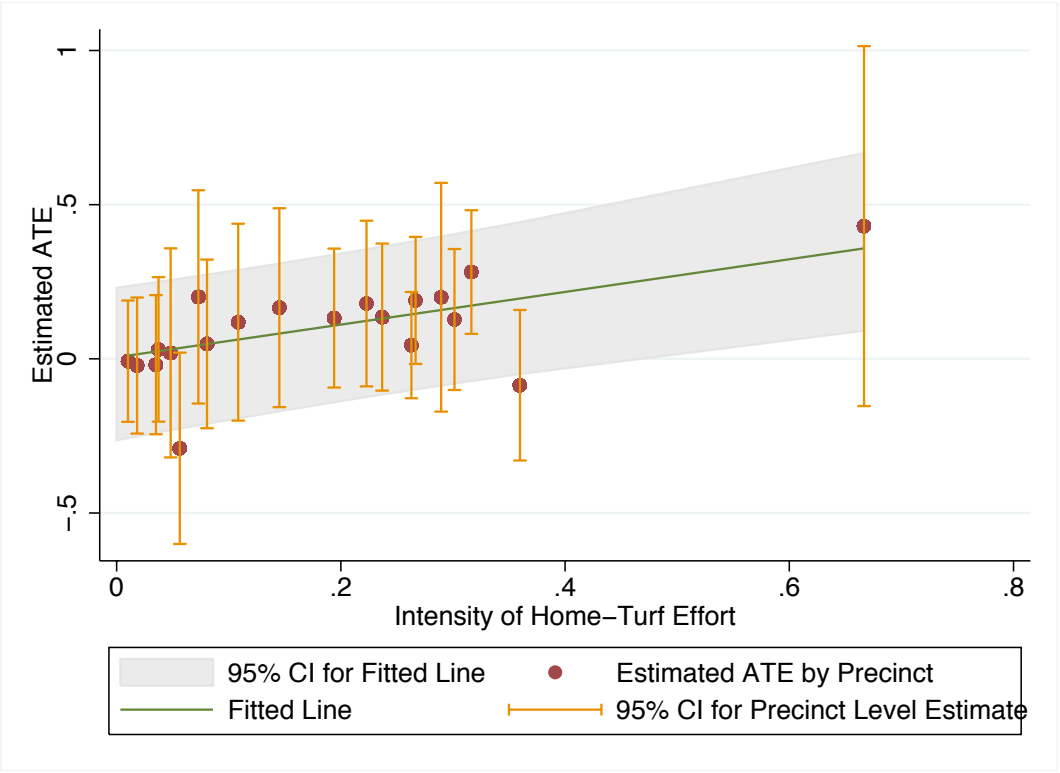


Table 4.7: Coefficients from 2SLS, Effect of Any Contact on Vote

Variable	Coefficient
2SLS	
Any Contact	**.066***
	(.021)
Constant	.34*
	(.008)
Clusters	13354
N	15367
F(1,13353)	10.05
2SLS with Control Variables	
Any Contact	**.051***
	(.020)
Democratic Registration	.016**
	(.008)
Age	.002*
	(.0003)
Vote History	.16*
	(.004)
Latino	.096*
	(.014)
Female	.002
	(.01)
Missing Age	-1.63*
	(.247)
Missing Female	-.045
	(.079)
Constant	.017
	(.012)
Clusters	13354
N	15367
F(8,13353)	302.74
2SLS with Precinct Fixed Effects	
Any Contact	**.048***
	(.02)
Constant	.58*
	(.033)
Precinct Fixed Effects	Included
Clusters	13354
N	15367
F(50,13353)	3.94
$* = \alpha = .05$	
$** = \alpha = .10$	

Note: The random assignment for treatment in the first wave is used as an instrument and has correlation .4016 with the treatment variable.

Table 4.8: Coefficients from 2SLS, Marginal Effect of a Second Contact on Vote

Variable	Coefficient
2SLS	
Second Contact	**-.083**
	(.112)
Constant	.44*
	(.028)
Clusters	2128
N	2392
F(1,2127)	.55
2SLS with Control Variables	
Second Contact	**-.088**
	(.101)
Democratic Registration	.0001
	(.023)
Age	.001
	(.001)
Vote History	.172*
	(.01)
Latino	.048
	(.048)
Female	-.012
	(.028)
Missing Age	-.88
	(.616)
Missing Female	.059
	(.23)
Constant	.146*
	(.042)
Clusters	2128
N	2392
F(8,2127)	44.77
2SLS with Precinct Fixed Effects	
Second Contact	**-.081**
	(.09)
Constant	.65*
	(.068)
Precinct Fixed Effects	Included
Clusters	2128
N	2392
F(26,2127)	1.76
$* = \alpha = .05$	
$** = \alpha = .10$	

Note: The random assignment for treatment in the second wave is used as an instrument and has correlation .2138 with the second wave contact indicator.

Table 4.9: Difference in Vote and Contact Rate Means: Home-Turf Canvassing vs Not

Variable	Not Home-Turf	Obs	Home-Turf	Obs	Difference
Any Contact	.316	9822	.402	5545	.086*
	(.005)		(.007)		(.008)
Voted	.356	9822	.367	5545	.012
	(.005)		(.006)		(.008)
$* = \alpha = .05$					

Table 4.10: Logit Coefficients from Marginal Effect of Home-Turf Contact on Turnout Amongst Contacted

Variable	Coefficient
Home-Turf Contact by Zip	.225*
	(.098)
Constant	-.266*
	(.03)
Home-Turf Contact by Precinct	.171
	(.279)
Constant	-.248*
	(.03)
N	5343
Number of Home-Turf contacts by zip	453
Number of Home-Turf contacts by precinct	52
$* = \alpha = .05$	

Table 4.11: 2SLS, Separated by Precinct: Home-Turf Canvassed vs Not

Variable	Home-Turf Coefficient	Not Home-Turf Coefficient
2SLS		
Contact	.085*	.52**
	(.03)	(.029)
Constant	.33*	.34*
	(.013)	(.01)
Clusters	4807	8547
N	5545	9822
F(1,N-1)	8.53	3.20
2SLS with Control Variables		
Contact	.07*	.037
	(.028)	(.027)
Democratic Registration	.001	.024*
	(.014)	(.010)
Age	.002*	.001*
	(.0004)	(.0003)
Vote history	.155*	.163*
	(.007)	(.005)
Latino	.117*	.090*
	(.03)	(.016)
Female	-.007	.007
	(.016)	(.011)
Missing Age	-2.14*	-1.37*
	(.395)	(.319)
Missing Female	.030	-.086
	(.136)	(.07)
Constant	.008	.021
	(.022)	(.014)
Clusters	4807	8547
N	5545	9822
F(8,Clusters-1)	110.22	194.22
$* = \alpha = .05$		

Note: The instrument used here is the treatment assignment from the first wave.

Table 4.12: Effect of Home-Turf Contact on Voting for Contacted Individuals

Variable			WFE	WFE
Home-Turf Contact by Zip	0.056**	0.043**	0.113**	0.091**
	(0.024)	(0.023)	(0.031)	(0.029)
Democratic Registration		0.007		0.011
		(0.015)		(0.015)
Age		0.001		0.001
		(0.000)		(0.000)
Missing Age		-0.017		-0.024
		(0.037)		(0.017)
Female		-0.012		-0.015
		(0.017)		(0.017)
Missing Female		0.060		0.085
		(0.147)		(0.146)
Vote History		0.178**		0.176**
		(0.007)		(0.007)
Latino		0.112**		0.113**
		(0.026)		(0.025)
Walker Fixed Effects			Included	Included
Constant	0.434**	0.108**	0.500**	0.212
	(0.007)	(0.023)	(0.217)	(0.215)
* = α = .05				
N	5343	5343	5343	5343

Note: There is a separate fixed effect coefficient for contacts made by unknown walkers (399 observations). Home-Turf Contact by Zip indicates the number of walkers that made a contact who were from the same zip code as the voter.

Table 4.13: Meta-Analysis

Method	Pooled Estimate: No Cov	p-val	Pooled Estimate: Cov	p-value
Fixed Effects	0.057	0.004	0.045	0.015
Random Effects	0.057	0.004	0.045	0.015
Test for Heterogeneity	46.293	0.584	47.480	0.535

Table 4.14: Random Effects Meta-Analysis Regression

Variable	No Covariates	With Covariates
Share of Contacts that are Home-Turf	0.382**	0.321**
	(.157)	(0.153)
Intercept	0.025	0.001
	(0.023)	(0.022)
* = α = .05		
N	50	50

Figure 4.4: Doorhanger Used in SCOPE Campaign

Chapter 5

Community-Based Mobilization: GOTV and SCOPE in South Los Angeles, June 2006

5.1 Summary

This chapter examines the effects of door-to-door canvassing by canvassers who are residents of the precincts in South Los Angeles in June 2006. Here we expect to find that canvassing efforts are more successful when canvassers are drawn from the neighborhoods themselves. This is consistent with the concept that within-network individuals are more likely to be similar types, and thus (as stated as a result in the theoretical chapter) will be able to communicate valence characteristics, such as the value of participating in an election. In this experiment we find no statistically significant effect of door-to-door canvassing. However, we do find a positive effect of home-turf canvassing. We discuss these conclusions in light of community-based GOTV efforts.

5.2 Introduction

The existing literature on get-out-the-vote (GOTV) field experiments has consistently used a range of individuals (paid, volunteer) who contact households which may, in the case of community-based organizations, be known to the canvasser. Paid, volunteer, and home-turf canvassing may account for differences in contact rates and treatment effects across experiments. For example, the existing literature has found that canvassers who match the contact on racial or ethnic characteristics (Michelson 2003) are more effective mobilizers. Analyzing the different types of campaigning has important implications for understanding the effectiveness of voter mobilization campaigns and in particular, for understanding turnout behavior.

This experiment was conducted with the Strategic Concepts in Organizing and Policy Education organization (SCOPE), a grassroots community-based organization located in South Los Angeles which was founded shortly after the Los Angeles riots in 1992 in an effort to provide members of the South Los Angeles community with a political voice.[1] As is consistent with all of SCOPE's campaign work, SCOPE's June 2006 GOTV campaign focused on using community-based precinct walking as a way to mobilize voters within their community. Walkers were split into volunteer walkers and paid walkers. One of SCOPE's stated mission strategies is to "build models of increasing civic participation where poor and disadvantaged communities and communities of color can become active participants in public policy-making and initiatives that impact their lives." As a consequence, SCOPE actively ensures that the individuals contacted during the campaign are contacted by individuals who are similar along sociodemographics, in particular, language. SCOPE has spent many years recruiting volunteers from the neighborhoods in South Los Angeles, and the neighborhood volunteers form a crucial component of this particular experiment.

The goal of this experiment is twofold; the first is to evaluate the efficacy of a grassroots GOTV campaign and the second is to attempt to directly measure the effect of home-turf canvassing. In the 2004 presidential contest, almost 2/3 of the

[1]SCOPE describes their mission, history and current campaigns on their website, http://www.scopela.org/agenda/index.html.

adult population surveyed (64%) in the Pew Internet and American Life Project Survey reported being contacted directly by political actors, with the respondents approximately split in half as to whether they had been contacted by Republicans or Democrats (Cornfield 2004). However, the focus of the Republican GOTV campaign was to link to pre-existing social ties (church congregations, neighborhood parties) in contrast to the focus of the Democratic GOTV campaign was to hire recruits or bus college students to pivotal states (Cornfield 2004). According to Democratic National Committee (DNC) spokeswoman Karen Finnley, the Democratic Party simply was not prepared to mobilize grassroots canvassers in 2004. As she says, "the ability to have this kind of infrastructure was something we were lacking" [in 2004] (Cillizza 2006). According to Cillizza, "The idea of neighbor-to-neighbor contact was used to incredible effectiveness during the 2004 presidential race by the Bush campaign, which relied heavily on community leaders to deliver its message" (2006). If indeed community-based mobilization campaigns are more efficacious, then this experiment should result in higher-than-average mobilization effects. Additionally, as it is possible to know whether or not a precinct was canvassed by a resident of that precinct, we are also able to test directly whether or not there is higher turnout within those precincts.

SCOPE assigned each canvasser a precinct to contact beginning four weeks prior to the election. These canvassers are sometimes assigned to their home precinct (referred to hereafter as "home-turf" canvassing) and several of the canvassers are paid employees. The variance in type of canvasser enables a comparison to be made of the success of different types of canvassing. The variance in canvassing may affect the success of the contact rate or even whether or not the contacts find the canvasser's get-out-the-vote message compelling.

5.3 Data and Methods

The experimental population universe consists of registered voters within South Los Angeles. This population was chosen simply because it is the population from which SCOPE draws its volunteers — additionally we knew, from contacts with other grass-

roots organizations and campaigns, that there were no other campaign efforts attempting to contact this voter population. Individuals were aggregated to the household level, and, by precinct, each household was randomly assigned to treatment or control. For those households in the treatment group, a canvasser was assigned to go to that address and speak for a few minutes with an individual in that household about the upcoming June 2006 general election. Canvassers began to contact households four weekends prior to the election and continued to do so up until election day.[2] Canvassers varied in type — some were volunteers, some were paid employees, and some were residents of the precinct within which they were speaking to the registered voters. Note here that while the contact was randomly assigned, the type of contact is not — and consequently all comparisons about the effectiveness of type of contact will be observational results and not the result of random assignment. Additionally, a canvasser could be both, for example, paid and walk in their home precinct.

This chapter has several significant differences with the previous chapter. In particular, the experimental population consists of all registered voters within twenty-six precincts in South Los Angeles, as opposed to the November population of occasional voters within a larger set of precincts. Additionally, the information about canvassers is at the precinct level. Thus we have recorded, for example, whether or not a precinct was contacted by a resident or by a paid canvasser, but no measurement of the intensity or frequency of that contact nor do we know the particular households contacted by the different types of canvassers. As a consequence our analyses of home-turf canvassing is less precise – here we observe similar results (home-turf canvassing appears to increase turnout) but with larger standard errors.

Table 5.1 describes the breakdown by treatment and control assignments for all 7212 individuals in the experiment. These individuals are randomly assigned at the household level, by precinct, from 26 precincts — a total of 1782 individuals were assigned to the control group while 5430 individuals were assigned to the treatment group.

[2]During the final weekend prior to the election, some individuals were contacted a second time by the campaign. Unfortunately, no data was recorded during this weekend to verify re-contacts.

Table 5.1 Goes Here

Ideally each individual in the treatment group would be contacted — however, this is simply not possible. The canvassers often make only a single attempt to contact a household, so that, for example, it would not be possible to contact someone in the treatment group if they were running an errand when the attempt was made. Canvassers report all types of problems with making contact with individuals in the treatment group — some individuals had frightening dogs in their yards, other individuals refused to open the door, and most often the canvassers simply did not have time to attempt contact with all of the individuals in the treatment group. Table 5.2 describes the number of individuals who were assigned to control and treatment and then tabulates which individuals were successfully contacted. Approximately 35% of the individuals assigned to the treatment group were contacted. Additionally, approximately 1.5% of the individuals assigned to the control group were contacted accidently.

Table 5.2 also describes the number of individuals who live in a precinct that was canvassed by a paid canvasser as well as the number of individuals who who live in a precinct that was canvassed by someone who lives in that precinct. Almost all of the precincts had some paid canvassing, whereas only six precincts were canvassed by a resident. Finally, we tabulate the number of individuals by control and treatment assignment group who voted in the June 2006 election. SCOPE recorded the individuals who were contacted and the classification of canvasser by precinct. The Los Angeles Registrar-Recorder's office provided data on which individuals participated in the June 2006 election.

Table 5.2 Goes Here

Prior to analysis, we first determine to what extent the data is balanced across the observable covariates. One of the reasons that it is necessary to conduct a randomized experiment is that there are remarkably few observable characteristics for each individual voter. This prevents us from using more traditional methods of analysis, such

as regression techniques, and requires the use of randomized experimentation. We check to ensure, however, that for the variables which we can observe that there is not a statistically significant difference between treatment and control groups. Table 5.3 describes the differences. No variable has a statistically significant difference across treatment and control. Note that there is some missing data for the age variable — not every individual in the data-set provided by the Los Angeles County Registrar Recorder had a birth-date, and as such, 462 individuals have missing age variables.

We also present an identical table for the differences in means, by covariate, describing the differences between the individuals who were contacted against those who were not. Table 5.4 is clear evidence of the need to incorporate randomization experimentation in analyzing mobilization efforts. Here we observe statistically significant differences in the vote history, age, and partisan registration (Democratic and Decline-to-State) between individuals who were successfully contacted vs not.

Table 5.3 and Table 5.4 Go Here

5.4 Results

We first present the intent-to-treat effect and the treatment effect calculated directly from the data summary for the treatment and control groups. Table 5.5 presents these calculations. There is an observable increase in the percent voting for the treatment group over the control group and very few of the control group individuals are inadvertently contacted during the experiment. Thus, with a contact rate amongst the treatment group of 33.47%, the estimated intent-to-treat effect is just over 1%. Again, the contact rate is affected by many factors. First and foremost, the canvassers did not attempt to contact every household in the treatment group, due to time constraints. For those they did attempt, some households posed some threat to their physical safety (dog in yard, unsafe street) and other households were either not home or had residents who were unable to open the door to the canvassers. It is possible that there are some unobserved characteristics which relate the ability of the canvasser to contact the voter with the voter's propensity to cast a ballot. As a consequence, it is

crucial to incorporate the treatment assignments as well as outcomes in the analysis. Controlling for the contact rate, this presents a treatment effect of just over 3%. Neither of these results are statistically significant at traditional levels, but they do demonstrate some support for a positive effect.

Table 5.5 Goes Here

We next analyze the effect of contact on vote using an instrumental variables approach where we use the randomization assignment as an instrument in a 2SLS process. We have a particularly high quality instrument because we are assured that the instrument (the treatment assignment) is uncorrelated with any unobserved characteristics since the randomization was completed independent of any covariates and prior to the start of the experiment. Note that the correlation between contact and treatment assignment is .32. We find no statistically significant effect of contact on vote with or without incorporating the control variables in Table 5.6, even when including a fixed effect by precinct. Robust standard errors are calculated with clusters by household — the randomization is done by household, and it seems likely that household members will communicate with each other about the choice to participate.

Table 5.6 Goes Here

Table 5.5 is consistent with Table 5.6; the contact coefficient remains statistically insignificant across all methods of estimation. We next examine whether or not we observe a positive effect of home-turf canvassing.[3] Table 5.7 indicates the coefficients from this analysis.

Table 5.7 Goes Here

We do find positive and statistically significant effects for the marginal effects across Table 5.7. The first coefficient looks at the marginal effect of home-turf canvassing amongst those individuals who were successfully contacted. Here the effect

[3]As almost all of the precincts were contacted by a paid canvasser, it is not possible to distinguish the effect of paid canvassing from that of canvassing.

is positive but not statistically significant. The next analysis incorporates an indicator variable those precincts where home-turf canvassing occured and those precincts where there was no home-turf canvassing. Here again there is no effect of contact, but the coefficient on home-turf canvassing is positive and statistically significant. This suggests that while ordinary canvassing may not have an effect, home-turf canvassing is likely to increase turnout.

We next proceed with a meta-analyses of the contact effects by precinct. This yields positive and statistically significant estimates. The estimate for contact effect controlling for precinct fixed effects is estimated to be slightly above 8%, and is statistically significant at traditional levels ($\alpha = .05$). Also, the estimate for contact effect controlling for precinct random effects is estimated to be slightly above 6.2%, and is also statistically significant at traditional levels ($\alpha = .10\%$) . These estimates are described in Table 5.8.

Table 5.8 Goes Here

The first two values in this table represent the meta-analysis when looking at precinct fixed- or random-effects. This is a sensible concept in that there will be variation across precincts, in particular in terms of contact rates, which is incorporated into the analysis. We find that both of these values do produce positive and statistically significant coefficients for the effect of contact. The second two values in this table represent the coefficients from the meta-regression. Here we are particularly interested in the coefficient on the home-turf indicator. We find it is positive (as hypothesized) but not statistically significant at traditional values. One problem with this measurement of home-turf canvassing is simply that it is at the precinct level — we do not know what fraction of individuals within that precinct were contacted by home-turf canvassers — and this additional noise makes it difficult to produce precise home-turf measurements. The only statistically significant home-turf estimate came in the 2SLS estimate, where the effect of home-turf incorporated the contact rate, which helps to reduce the noise in the estimate.

5.5 Conclusions

The only statistically significant effect of any particular type of canvassing found in this experiment is the positive effect of home-turf canvassing (a 2.7% increase in turnout) in the instrumental variable 2SLS regression. The contact effect appears to be positive across all analyses but at no point would be considered statistically significant at traditional levels except in the meta-analysis which incorporates fixed effects (which then finds an 8% increase in turnout). Thus the estimates do weakly correspond to our hypotheses, which are consistent with the theoretical framework — that there would be a positive effect from contact and that this effect would increase with home-turf contact.

One caveat in this analysis is that voters across all types of vote histories are included in both treatment in control. Thus some individuals are contacted who are regular or likely voters, and for this group, as they are likely to vote already, there is very little effect of canvasser contact. Additionally, there was very little on the ballot to motivate individuals to get to the polls in June 2006. One disadvantage of the small turnout is the loss of statistical power for the experiment — and thus an increase in the standard errors. However, we do find our principle two hypotheses are supported — that canvassing increases turnout, and that home-turf canvassing in particular increases turnout.

Home-turf canvassing is particularly important in understanding the voter mobilization process. Political "machines" have long relied upon neighborhood groups to mobilize voters, and experiments which involve home-turf canvassing efforts allow comparisons to be made between home-turf effects and other volunteers which enable a direct view into why political machines have been so historically effective. Perhaps it is because individuals are more likely to feel peer-pressured into participating or perhaps it is simply that they are more willing to accept the idea that it is in their best interest to participate from a known individual, but it does appear that home-turf canvassing does increase turnout. Interestingly as well in this experiment is that the preliminary evidence suggests that in fact paid canvassing does not have a sim-

ilar effect. Future research will be dedicated to evaluating the comparisons between volunteer and paid canvassers.

5.6 Tables

Table 5.1: Treatment Assignments by Precinct

Precinct	Control	Treatment	Total
9000288A	59	228	287
9001026B	92	206	298
9001189B	55	184	239
9001314A	68	123	191
9001747A	72	264	336
9001755A	65	256	321
9001779A	52	195	247
9002024A	48	162	210
9002025A	32	135	167
9002087A	93	260	353
9002090A	80	263	343
9002091A	72	165	237
9002101A	73	232	305
9002137A	79	254	333
9002162A	77	187	264
9002167A	69	238	307
9002179A	80	247	327
9002211A	106	242	348
9002219A	76	249	325
9002461A	57	200	257
9002477A	42	158	200
9002481A	57	190	247
9002482A	79	185	264
9002484A	60	229	289
9002497A	45	109	154
9005414A	94	269	363
Total	1782	5430	7212

This table indicates the number of individuals, by precinct, who were randomly assigned to treatment and control groups. Individuals were aggregated to the household level and randomization was done by household, within each precinct.

Table 5.2: Treatment Assignments, Contacts, and Turnout

	Random assignment	Any Contact	Paid Precinct Contact	Home-Turf Precinct Contact	Vote
Control	1782	26	25	5	308
Treatment	5430	1897	1844	458	994
Total	7212	1923	1869	463	1302

This table indicates the number of individuals who were assigned to treatment and control groups and then indicates within those groups which fraction were successfully contacted. Individuals are then tabulated by paid precinct contact and Home-Turf precinct contact amongst those successfully contacted.

Table 5.3: Covariate Balance: Difference in Means Across Treatment and Control Assignment

Variable	Difference in Means (Treat - Control)	N
Democratic Registration	.003 (.011)	7212
Republican Registration	.002 (.006)	7212
Decline-to-State Registration	.003 (.009)	7212
Other Registration	-.008 (.006)	7212
Age	.135 (.463)	6850
Vote History	.058 (.076)	7212
Female	-.002 (.013)	7212
$* = \alpha = .05$ $** = \alpha = .10$		

This table presents the difference in means, by covariate, for the treatment and control random assignments. No difference is statistically significant.

Table 5.4: Covariate Balance: Difference in Means Across Realized Contact and No Contact

Variable	Difference in Means (Treat - Control)	N
Democratic Registration	.029* (.011)	7212
Republican Registration	.004 (.006)	7212
Decline-to-State Registration	-.028* (.009)	7212
Other Registration	-.004 (.005)	7212
Age	3.13* (.463)	6850
Vote History	.256* (.077)	7212
Female	.016 (.013)	7212
$* = \alpha = .05$ $** = \alpha = .10$		

This table presents the difference in means, by covariate, between those individuals who were contacted against those who were not, regardless of random assignment.

Table 5.5: Effect of Contact: Intent-to-Treat Effect and Treatment Effect

N in treatment group	5430
N in control group	1782
N in the treatment group who are actually treated	1897
N in the control group who are inadvertently treated	26
N who voted in the treatment group	994
N who voted in the control group	308
Percent Voting — Treatment	18.31%
Percent Voting — Control	17.28%
Percent Contacted — Treatment	34.94%
Percent Contacted — Control	1.46%
Contact Rate	.3347
Estimated ITT Effect	1.0217
SE of ITT Effect	1.0385
Estimated Treatment Effect	3.0521
SE of Treatment Effect	3.1021

Note that these estimates do not calculate robust standard errors nor cluster the estimates by household. Neither the ITT nor the ATE are statistically significant at traditional levels.

Table 5.6: Coefficients, Effect of Contact on Vote

Variable	Coefficient
2SLS, Effect of Contact on Vote	
Contact	.03
	(.03)
Constant	.17*
	(.01)
Clusters	4883
F(1,4882)	.82
N	7212
2SLS, Effect of Contact on Vote with Covariates	
Contact	.03
	(.03)
Democratic	.03*
	(.01)
Female	.02**
	(.01)
Vote History	.02*
	(.002)
Age	.003*
	(.0003)
Missing Age	-2.39*
	(.310)
Constant	-.03**
	(.015)
Clusters	4883
F(6,4882)	45.51
N	7212
2SLS, Effect of Contact on Vote with Fixed Effects by Precinct	
Contact	.047
	(.03)
Constant	.34*
	(.03)
Fixed Effects	Included
Clusters	4883
F(26,4882)	2.95
N	7212
$* = \alpha = .05$	
$** = \alpha = .10$	

These coefficients from the 2SLS analysis present no evidence that contact has a statistically significant effect on turnout. However all twenty-six precincts are included in this analysis — including those in which there was no canvassing.

Table 5.7: Effects of Home-Turf Precinct Canvassing on Vote

Variable	Coefficient
Logistic Regression, Estimates Based Upon Contacted Individuals Only	
Home-Turf Walker	.15
	(.14)
Constant	-1.12*
	(.07)
Clusters	1448
Psuedo R2	.0007
N	1923
2SLS with Home-Turf Precinct Indicator	
Contact	.29
	(.034)
Home-Turf Precinct	.027*
	(.012)
Constant	.166*
	(.010)
Clusters	4883
F(2,4882)	2.95
N	7212

These coefficients all focus on the impact of home-turf canvassing. The first coefficient is calculated based upon contacted individuals only. The second estimate incorporates an indicator for home-turf contact at the precinct level.

Table 5.8: Coefficients from Meta-Analysis, Effect of Contact on Vote

Method	Pooled Estimate	p-value	N
Fixed Effects	.083	.006	26
Random Effects	.062	.105	26
Test for Heterogeneity	Q=36.52	.064	25
Meta-Regression	Estimate	z-value	N
Home-Turf	.08	.85	26
Constant	.044	1.01	26

Chapter 6

Are Straw Polls an Information Source for Voters?

6.1 Summary

Voters are often assumed to use cues, such as endorsements or opinion poll results, to guide their choice of candidate. In the theoretical model in chapter two, voters are presented with information about candidates from individuals within their social framework and the voters are assumed to use this information to Bayesian update their priors on the candidates' competency. This chapter examines whether or not voters will indeed behave consistently with this assumption — that is, update their beliefs via Bayes Rule. This experiment differs in that the voters will gain information about candidate locations using information contained in the straw polls, or opinion polls, that occur before an election, as opposed to information about competency. In order to test this one aspect of voter behavior, we conduct laboratory experiments. We undertake a series of laboratory experiments where we use a similar experimental setup to that of McKelvey and Ordeshook (1985). However, unlike McKelvey and Ordeshook, we control the candidate locations and the informed voters candidate choices. Each subject is uninformed about the candidate locations but knows (1) their location in the distribution of voter ideal points and (2) which candidate is further left in a 1D space. We ask each subject to vote in a series of straw polls where they are asked which candidate they would most prefer if the election were

held today. We incorporate the informed voters choices into the results and present the total poll results for all the uninformed voters. This process repeats until the election, at which point a candidate wins. We test a series of hypothesis about the amount of information the uninformed voters will procure from the straw poll results, including the convergence result presented in the McKelvey and Ordeshook paper.

6.2 Introduction

There is extensive research on American voters that indicates a large proportion of voters simply do not know a great deal about politics (Almond and Verba 1963; Berelson, Larzarsfeld and McPhee 1954; Campbell, Converse, Miller and Stokes 1960; Converse 1964; Kinder and Sears 1985). The range and depth of this ignorance is astounding to political scientists — that voters may not know the number of senators from their state, for example, and (less surprisingly) that they may have difficulty locating themselves and the candidates on ideological spectrum, remembering the names of particular candidates, or knowing where any particular candidate stands on a broad set of issues. According to Campbell et al. (1960, p. 170), "many people know the existence of few, if any of the major issues of policy." Voters simply do not have the incentives to pay the costs associated with becoming informed about politics as the probability that they cast a pivotal vote is extremely low, but this does not imply that they vote randomly (Wittman 1989). Voters may use other processes to obtain information, such as reliance upon a trusted peer (Downs 1957) or reliance upon an interest group (Lupia and McCubbins 1998), although these methods of information procurement may be filled with biased information and may still not be costless. As stated by Popkin et al. "an individual facing a choice situation like voting, where the number of alternatives is limited, need only gather enough information to determine which alternative is preferable" (1976, p. 789). This paper examines yet another possibility for uninformed voters to procure information — pre-election straw polls. If institutions such as straw polls are useful in informing voters about which candidate to select without the voter ever knowing a significant

amount of information about the candidates, then in fact the uninformed voters are likely to choose the candidates who best represent their interests without knowing a great deal about politics (McKelvey and Ordeshook 1986, McKelvey and Ordeshook 1985b). Pre-election straw polls are an institution which affect outcomes by providing additional information (Ordeshook and Palfrey 1988) and one particular advantage of such an instrument is that pre-election polls are available to all voters quickly and easily in all sorts of public media. The extent to which uninformed voters may use this information may assist in producing a more representative democracy, in which all voters have sufficient information to cast informed votes for their preferred candidate and thus elect a candidate whose policy closely resembles that of the median voter.

McKelvey and Ordeshook (1985) first realized that, by giving a set of uninformed voters a historical prior on partisanship (in an experimental setting, defined as which candidate was further left than the other) it would be possible for a series of pre-election polls to reveal the candidate midpoint to the uninformed voter. In their setup, they allow subjects to be informed voters, uninformed voters, and candidates. Voters have symmetric single-peaked preferences, and the distribution of preferences of the electorate is publicly known. Candidates do not know which voters are informed or uninformed, hence they cannot target only the informed voters, but instead base their strategies upon the pre-election straw poll data. McKelvey and Ordeshook provide the candidate locations to the informed voters but permit only the uninformed voters to know the relative candidate location as a proxy for an indirect cue or historical fact, such as an endorsement or a party label. In their setup, an equilibrium to the game is a vector of strategies and beliefs where each voter votes for the candidate whose ideal point is on the same side of the candidate midpoint as the voter's ideal point, where each candidate chooses a platform position to maximizes her payoff based upon her beliefs about the median informed voter, and where informed voters know the true candidate midpoint and uninformed voters use the straw poll data as a measurement of the candidate midpoint. As they state, "uninformed voters initially vote randomly. After observing the first poll, they obtain an estimate of the candidate midpoint on the basis of that poll, assuming that other voters except themselves are informed.

This leads to a new poll result. Uninformed voters revise their beliefs based on this new poll. It can be shown that this process converges to the equilibrium" (McKelvey and Ordeshook 1990, p. 306). In their experimental results they find support for their equilibrium result.

We base our experimental setup upon McKelvey and Ordeshook (1985) but make slight modifications which enable us to test whether or not the uninformed voters are able to have beliefs which converge to the true candidate midpoint after a series of polls. First, we extend the number of polls. McKelvey and Ordeshook conduct only two pre-election polls and then an election in a total of two experimental sessions. By extending the number of polls we increase the ability to determine whether or not the votes cast are converging to the candidate midpoint — we will have additional data in each poll which allows us to test the dynamics across polls. We anticipate that by doing so we will be able to, in fact, observe convergence. Second, we control the informed voters and candidates. This reduces the amount of noise in the experimental setup and allows us to focus more specifically upon the uninformed voters. Finally, we move the candidate midpoint around the one-dimensional space. Because the McKelvey and Ordeshook framework permitted the subjects to be candidates (and in equilibrium the candidates will locate at the median voter), there is very little variation in candidate position across elections. By moving the candidate midpoint around the space we will be able to determine to what extent we observe convergence within each election, and we will not need be concerned about voters realizing the candidate strategies of locating at the median voter across elections, which is a possibility in the McKelvey and Ordeshook results.

We conduct our experiments with the anticipation that it will be possible for the uninformed voters to glean some information from the straw polls. There are often ways the observation of a public signal can prevent individuals from converging towards beliefs and actions based upon the true state the world, however (Feddersen and Pesendorfer 1996, Hung and Plott 1999). We anticipate that as the voters here have no incentive to cast a ballot for anything other than their sincere preference, however, this public signal will work to reveal the true state of the world for all

participants.

These experiments will allow us to test the behavioral assumption that voters will Bayesian update given a noisy, public signal. Experiments are advantageous in that they permit a direct test of a variety of behaviors or theories in a controlled environment. In this case, we isolate the process by which voters will change their beliefs during the pre-election communication phase. This enables us to both understand what voters believe other voters are doing, as well as to test different possibilities about the process by which voters update their beliefs. We find that in fact voters are Bayesian, and are playing weakly dominated strategies; we apply our results to the theoretical model which began this dissertation and confirm that indeed we may assume that voters will update their beliefs based upon voter-to-voter communication.

6.3 Experimental Setup

Each of these experimental sessions were conducted using undergraduates enrolled at the California Institute of Technology during the academic year 2004-2005. Subjects were paid in cash for their participation. Each session took approximately two hours.[1] In the original McKelvey and Ordeshook paper, undergraduates enrolled in the California Institute of Technology were also used as subjects. However, in their experiments, they included a larger number of subjects (between 50-60 as opposed to 10-17) and consequently their experiments ran for approximately four hours.

In this experiment all subjects are uninformed voters. Each voter is characterized by an ideal point on an integer within the interval [0,100] that was drawn from a uniform distribution. Uninformed voters do not know the platform location of either candidate A or B. However, each uninformed voter does have some small amount of information about which to base her decision. She knows the relative locations of the candidates — that is, which candidate is located to the left of the other candidate. She also knows where her ideal point is in relation to the distribution of ideal points across all voters (both uninformed and informed), the total number of voters, and the

[1] In the final section we have included the instructions each subject reads prior to the experiment.

precise number of each type.

In the experiment the behavior of the candidates and the informed voters is completely controlled for by the experimenter. There are two candidates (A and B) with platform positions located on an integer within the interval [0,100]. Candidate A will always have a policy platform that is located to the left of candidate B. The candidate platform locations are determined by a random draw from a uniform distribution on the interval. The informed voters vote sincerely and know both candidate locations.

The experiment begins with all uninformed voters asked to participate in a straw poll. Every voter must participate and vote for either candidate A or B. The voters are asked to vote in the poll as they intend to vote in the final election. After each poll, the results are publicly announced. These results include the votes of both the uninformed and informed voters. The last poll is considered the election — voters' payoffs are primarily determined by this vote — and after the election results are announced, the candidate locations are revealed. The process then repeats with new candidate locations.

The experiment consists of two treatments — one in which there are three straw polls and then one election, and the second in which there are five straw polls and one election. Payoffs to the voters are determined by the distance between the winner of the election and their ideal point. In the second treatment, a small bonus is given for correct voting in the pre-election polls. In the first treatment there are 12 total elections and in the second treatment there are 24 total elections.

6.4 Data

We conducted five sessions, three of the first treatment and two of the second treatment. There are ten informed voters in each treatment, with ideal points at

[32,43,48,52,55,60,64,70,80,82]

in sessions one and two and with ideal points at

[3,4,5,38,47,52,67,69,81,100]

in sessions three, four, and five. The number of uninformed voters in each session varies (based upon availability) somewhere between 10 and 17 subjects. The sessions are described in Table 6.1; the entire experiment produces a total of 5,136 instances where an uninformed voter is asked to "cast a ballot" in either a poll or election. We organize our data into a matrix where each entry describes the behavior of one voter in one particular poll in one election in one session.

Table 6.1 Goes Here

Over all polls and elections, the uninformed voters split almost evenly for candidate A and B (a total of 2762 votes are cast for A and 2373 votes are cast for B). The candidate midpoint ranges from 8.5 to 96.5 (with a mean of 54.29), and the observed split, which the uninformed voters will use to produce their beliefs about the candidate midpoint, ranges from 20 to 90 (with a mean of 52.87). Note here again that the voters' ideal points are drawn from a uniform distribution, with the uninformed voters having ideal points which range from 2 to 99 (with a mean of 46.45).

We observe, for each session, the uninformed voter's vote in each poll and election. We also record the candidate midpoint and the publicly observed outcome of each poll and election. We anticipate that the outcome of the pre-election polls will be used by the uninformed voter to determine her beliefs about the candidate locations. We also observe the treatment used in the session, the number of uninformed voters in the session, and the particular poll number for that particular vote.

We produce several variables which measure distance between the voter and a variable observed to the voter which we anticipate is used in her decision calculus. First, we produce a variable which describes the distance between the voter and the candidate midpoint. Second, we produce a variable which describes the distance between the voter and the point 50. Finally, we produce a variable which describes

the distance between the voter and the previous poll outcome, which we anticipate will be used as her belief. For the first poll of any election we assign a value of "50" to this belief.

Based upon these variables, we produce three measurements of error. First, we produce an indicator variable which describes whether or not the voter cast her ballot in the poll or election for the candidate that was indeed the closest to her ideal point (and thus would yield her the greatest payoff). We will refer to this indicator as a measurement of the "Correct Error". Next, we produce an indicator variable which describes whether or not the voter cast her ballot for the candidate who would be, in expectation, on the same side of the point "50" as the voter herself. We refer to this indicator as "Naive Error". Finally, we produce an indicator variable which describes whether or not the voter used, as though it were the absolute truth, the reported split from the previous poll. We refer to this indicator as "Bayesian Error".

Table 6.2 Goes Here

Table 6.2 describes the number of errors per type of error, and then indicates the rate at which these errors occur over the different polls. It is clear that the Correct Error decreases over time and it appears that the Bayesian Error does as well. The Naive Error, however, appears to be fairly constant.

Our hope is to be able to evaluate two distinct models in comparison to each other. We would like to demonstrate that indeed voters are incorporating the previous poll results into their decision and not voting based simply upon the candidate's expected positions and disregarding all pre-election polls. We refer to the model which produces Bayesian Error as the Bayesian model, and the model which produces Naive Error as the Naive model.

6.5 Results

We pool all sessions together to evaluate the two distinct models— the naive model, that voters make their selections based entirely upon their relative location to the

point 50 and their expectation about the candidate location, and the Bayesian model, that voters make their selections based upon the reported straw polls.[2] We find support for the Bayesian model but do not observe complete convergence to the true candidate midpoint.

We look for an effect on whether or not the voter was able to indeed vote correctly. As the distance between the candidate midpoint and the observed split from the previous poll increases, if the voter is indeed using the previous poll as a way to determine the location of the candidates then the probability that this voter is unable to vote correctly should increase. Similarly, as the distance between the candidate midpoint and the point 50 increases, if the voter is indeed using the point 50 as a way to determine the location of the candidates then the probability that this voter is unable to vote correctly should also increase. Unfortunately, the variable which describes the distance between the candidate midpoint and the observed split and the variable which describes the distance between the candidate midpoint and the point 50 are highly correlated (correlation coefficient of .5). However, we also know that if the voters are indeed updating according to Bayes' Rule then the probability that a voter casts a ballot for the wrong candidate should decrease for each new poll.

Thus, we incorporate both of the distance variables, a variable which describes the particular poll number, an indicator for the treatment (short or long number of polls), an indicator for whether or not this is the first election in any particular session (to control for confusion effects), and an indicator for the number of subjects in the session.

Using the indicator for error from the Correct Error measurement as a dependent variable, we produce logit coefficients as described in Table 6.3.

Table 6.3 Goes Here

The coefficients presented provide a clear case of support for the Bayesian model over the Naive model. A positive and statistically significant coefficient on the variable

[2]These sessions may be pooled so long as an indicator is included for treatment (short vs long), which provides the only significant difference in the setup. The number of subjects is also included as a control.

which describes the distance between the candidate midpoint and the previous poll indicates that as that distance increases, the number of errors is likely to increase. This is consistent with the Bayesian model. Furthermore, the coefficient on the variable which describes the distance between the candidate midpoint and the point 50 is negative and statistically significant. This indicates that as this distance increases the voter is less likely to vote for the wrong candidate — most probably because the candidate locations are easier to distinguish in the pre-election polls. Also, consistent with the Bayesian model, there is a negative and statistically significant coefficient on the variable which describes the poll number. Thus voters are likely to be able to more correctly identify the candidate locations with additional polls.

Interestingly, while the experiment type is not significant, the number of subjects within that particular session has a significant and positive coefficient. This could be due to the fact that an increased number of subjects implies that the ratio between informed and uninformed is lower — thus the quality of the signal observed in the pre-election polls is lower. Finally, as is consistent with our expectations, the indicator for whether or not this is the first election in any particular session has a positive and statistically significant coefficient. The subjects are simply more likely to make errors when they begin the experiment.

We next examine to what extent the ideological distance between the observed poll result and the voter, and between the candidate location and the voter are useful in predicting the voter choice. We also are interested in whether or not the voter is located to the left or to the right of the point 50. We produce three indicator variables — one which describes whether or not the voter is located to the left of the candidate midpoint, one which describes whether or not the voter is to the left of the midpoint as stated in the poll, and one which describes whether or not the voter is to the left of the point 50. Table 6.4 presents logit coefficients where we use an indicator for the subject's choice for candidate (1 if votes for Candidate A, 0 else) as the dependent variable.

Table 6.4 Goes Here

These coefficients are all positive and statistically significant. The largest coefficient is the one for the indicator for whether the voter is to the left of the candidate midpoint as stated in the poll — this suggests that in fact the voters are Bayesians. One problem with this analysis is that these three variables are highly correlated with each other — thus, we rely upon the description of the impact of these variables on the number of total errors produced.

6.6 Conclusion

This analysis demonstrates that uninformed voters do in fact glean information from straw polls. We observe that not only are the number of "errors" associated with voting decreasing over time, but that also the distance between the true candidate midpoint and the observed candidate split is a large factor in determining those errors. The fact that information about candidate location can be determined with such a small number of polls and such a limited set of information suggests that despite the surveys which find extensive lack of information amongst voters, voters may in fact be choosing the candidates whose ideologies are closest to their own.

When examining this analysis, one puzzle which emerges is that the voters did "better" than if they had been voting according to Bayes' Rule and the decision process modeled by McKelvey and Ordeshook (1985). As enumerated in Table 6.2, a total of 471 voters at any point voted for the wrong candidate. Had their decision process been to follow Bayes' Rule precisely as described in this paper, a total of 494 voters failed to do so. How is it that the voters are better than Bayesian? One possibility is that they are using some weighted measure of previous polls when revising their vote as a way to attach a measurement to the amount of error observed in this process. Future research will explore the ways in which voters are able to parse out the noise from these pre-election polls to gauge the candidate locations. If it is possible to produce institutions which can assist voters in doing so, then pre-election polls may in fact help the uninformed voters to have beliefs which converge to the true candidate positions. While this chapter indicates that indeed voters are

incorporating the pre-election polls into their decision, it is also clear that this is not the only factor which goes into determining their vote.

The evidence which suggests that voters are Bayesians is that the Bayesian prediction model has the fewest number of errors and that the coefficient on the Bayesian indicator is the largest (comparing against the other two possible models). This evidence — that voters will indeed update their beliefs about a candidate location from a noisy, public signal — is consistent with the behavioral assumption in the theoretical model.

6.7 Tables

Table 6.1: Experimental Sessions

Session	Total Uninformed	Treatment	Total Observations
One	17	3 Polls, 1 Election	816
Two	15	3 Polls, 1 Election	720
Three	12	3 Polls, 1 Election	576
Four	11	5 Polls, 1 Election	1584
Five	10	5 Polls, 1 Election	1440

Table 6.2: Tabulated Error by Type

Error Type	Total	Poll 1	Poll 2	Poll 3	Poll 4	Poll 5	Poll 6
Correct Error	471	12.59%	11.53%	9.2%	7.65%	5.16%	4.37%
Naive Error	677	9.39%	14.63%	14.43%	13.66%	13.69%	13.88%
Bayesian Error	494	9.4%	14.34%	11.24%	9.1%	4.56%	3.17%

Table 6.3: Logit Coefficients on Error

Variable	Coefficient
Distance between Belief and Cand MP	.04*
	(.006)
Distance between 50 and Cand MP	-.032*
	(.005)
Poll Number	-.065**
	(.038)
Shorter Expt Type	-.22
	(.20)
Total Number of Subjects	.179*
	(.035)
First Election in Session	.96*
	(.15)
Constant	-4.31*
	(.41)
**	$\alpha = .10$
*	$\alpha = .05$
N	5136

Table 6.4: Logit Coefficients on Candidate Vote

Variable	Coefficient
Indicator for Location of CMP and Voter's Ideal Pt	1.08*
	(.091)
Indicator for Observed Poll CMP and Voter's Ideal Pt	1.55*
	(.112)
Indicator if Voter's Ideal Left of Pt 50	.87*
	(.096)
**	$\alpha = .10$
*	$\alpha = .05$
N	5136

6.8 Written Instructions

This experiment studies polling and voting in an election with two candidates. You will be paid for your participation on the basis of the decisions you make. If you are careful and make good decisions, you can make a substantial amount of money.

In this experiment, there are two candidates, labeled A and B, and you are a voter. You will participate in a number of periods, each consisting of three polls and an election.

In each period the candidates will take positions on a number line that goes from 0 to 100. Before each election three polls will be taken in which all voters are asked to indicate their preferred candidate. Then, the poll outcomes will be announced to the group. After the third poll, all voters will vote for a single candidate in the election and the candidate who gets the most votes will be considered the winner. At this point the candidates may move to new locations on the number line and the poll and election process will repeat.

Voters are paid for their participation on the basis of their payoff chart. Please turn to the last page of the instructions to view the sample payoff chart. Remember that this chart is a sample and not your true payoff chart. This chart depicts the line where candidates will take positions and a sample payoff for a voter. The line is simply the set of all numbers between 0 and 100. The experimenters will select candidate positions on the line for each period. Candidates are equally likely to be at the end of the line as they are to be at the middle of the line. Each voter will be paid based on the position of the winning candidate on her payoff chart. For example, suppose candidate A is located at position 20, candidate B is located at position 25, and candidate B wins the election. Then you would earn 400 francs for that period. Note that on the sample payoff the maximum payoff is at position 45.

Voters will also be paid on the basis of their polling choices. For each poll that the voter chooses the candidate that is closest to the voter's maximum point, the voter will receive a small bonus in francs.

In the actual experiment, the payoff charts for each voter may be different. Each

voter will have a payoff chart which has a maximum payoff. Payoffs decrease symmetrically as candidate positions move away from the maximum in either direction as in the sample chart. However, different voters' maximums may be at different points on the number line, and their payoffs may decrease at different rates.

One important rule in the experiment is that the information on your payoff chart is private information. None of the other voters should know the information on your payoff chart. Please do not talk with other participants during the experiment. Are there any questions about the payoff chart?

In the experiment there will be two groups of voters (drawn from a uniform distribution) voting in each poll and election, uninformed and informed. All voters in this room are uninformed. This means that throughout each period the positions of the two candidates will not be made public. You will be given limited information about positions of the candidates and the other voters. Candidate A is always furthest to the left (closest to 0) and candidate B is always furthest to the right (closest to 100). You will also know, from your payoff chart, the percentage of all voters who have maximums to the left and to the right of your most preferred position.

The informed voters will be generated by the experimenters. The informed voters will know candidate positions and thus will always vote for their most preferred candidate. Every poll and election will include the true preferences of these voters. These voters are also included in the information you have about the percentage of voters who have maximums to the left and to the right of your most preferred position.

To review, the sequence of events will be as follows: At the start of each period the first poll will be taken and the results announced. Remember that both uninformed and informed voters will participate in all polls. The second and third polls will be taken and results announced. Then the final election will take place. After the final election the candidate positions from that period will be announced. After the last period the experiment will end. At this point, voters will be paid the sum of their payoffs for the position of the winning candidate in each election. Your monetary payoffs are increasing in the number of francs you earn.

Please take a moment to fill out the quiz below using your true payoff chart and

not the sample payoff.

Quiz:

1. My point of maximum payoff on my payoff chart is: (blank)

2. At this maximum point I will get a payoff of (blank) francs.

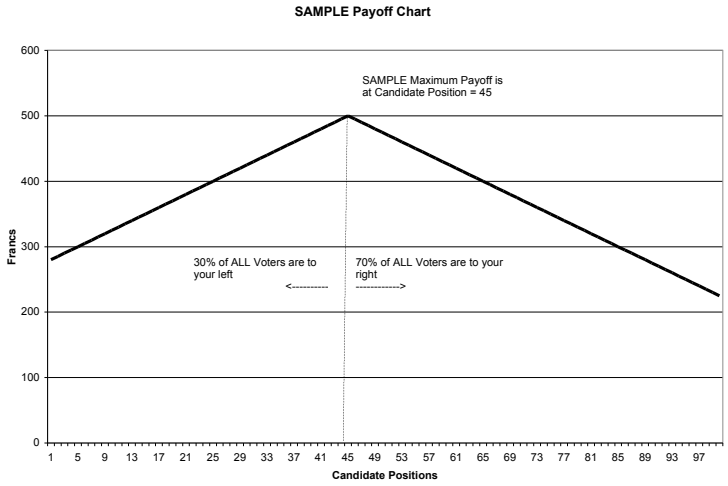
3. There are (blank) percent of voters to the left and (blank) percent of voters to the right of my maximum payoff point.

4. If Candidate A is at position 60, then Candidate B must be located between numbers (blank) and (blank).

5. True or False: By announcing the candidate that is closest to my maximum payoff in poll 1, I will get additional francs regardless of whether or not that candidate wins: (blank)

Are there any questions?

Figure 6.1: Sample Payoff



Chapter 7

Summary of Survey Data and Political Networks

This section examines three distinct data-sets and uses a variety of statistical techniques to examine the potential consequences of political networks on an individual's party identification and presidential vote choice. Conditional upon a series of assumptions about the data-generating processes, political networks play highly influential roles in determining an individual's political behavior.

The first chapter in this section uses the 2000 American National Election Survey to determine whether reported political discussants' presidential choices are likely to influence the respondent. This chapter controls for a large number of covariates which would alleviate biases which might be introduced due to the respondent's selection of particular discussants. The chapter finds statistically significant effects on presidential vote choice and develops a statistical technique to make inferences about the role of discussant presidential choices using the minimal amount of parametric assumptions.

The second chapter in this section uses an internet survey conducted in October and November of 2006. Again the respondent provides information about political discussants which is used to examine the possibility that respondent's choice of party identification is influenced by the discussants choices. Controlling for the respondent's characteristics and the nature of the relationship with the discussant, there remains an effect of discussant party identification on respondent's choice of party identification.

The third chapter in this section uses the 2006 American National Election Pilot

Study to determine whether the discussant's partisan identification plays a role in determining the respondent's choice of party identification. Here each respondent is asked twice to identify their party — once in November 2004 and again in November 2006. In the second instance each respondent is also asked to provide a series of discussants and to describe the nature of those relationships. Controlling for the respondent's characteristics and the discussant relationships, there is an increased probability that each respondent will change the party identification towards agreement with the discussants'.

One unique opportunity which emerges from the second and third chapters in this section is the opportunity to examine the responses to identical questions asked on both an internet and an in-person survey. Each survey asked respondents to generate discussants, and both asked respondents about their discussant's political preferences. Both surveys were in the field during identical timeframes. However, the second chapter uses an internet survey (which attempts to gather a national probability sample) while the third chapter is based upon a survey that uses an in-person door-to-door method. We observe a fairly high number of responses who indicate that they have discussion partners who are geographically dispersed, we observe more disagreement in the in-person survey than the internet survey, and in both instances we observe correlation between the respondent's preferences and those of the discussion partners.

One assumption made across these analyses is that voters are able to accurately determine the partisanship of their social connections. Studies of the ability of voters to determine this have found that there is a bias for the respondent to over-report agreement with the discussant. In Huckfeldt and Sprague's 1996 Indianapolis-St. Louis Study they find that the respondents are 80% accurate when asked about the partisanship of their political discussants and that the respondents are more likely to be accurate if they agree with the discussant's true political preferences (Huckfelt, Johnson, and Sprague 2004).

This section provides a rare and limited opportunity to gain some insight into the direction of causality by incorporating additional control variables and new statistical

technologies but, standing alone, does not resolve whether or not the discussants influence the respondent or vice versa. Furthermore, it leaves unaddressed the questions of motivation — do respondents use discussant information to update their priors and thus make better choices, or are they simply motivated by peer pressure to appear like those around them?

This chapter does validate a set of assumptions about voter communication found in the theoretical framework as well as demonstrates findings which are consistent with some of the theoretical chapter's results. In particular, each of these chapters demonstrates that voters are likely to communicate about politics and that there will be correlation between social connection and candidate/party choice, controlling for preferences.

Chapter 8

What Makes a Democrat Vote Like a Republican? Political Networks and Contextual Effects

8.1 Summary

This chapter shows that the politics of a voter's social context influences her vote choice even when controlling for political ideology and other relevant covariates. This analysis which determines the existence of contextual effects is based on the November 2000 American National Election Study (ANES) data, which provides the survey respondent's partisan breakdown of her discussion network, as well as locating the survey respondent within a particular congressional district. The partisanship of the respondent's discussion network and congressional district are observed to have an effect on the respondent's presidential vote choice. The chapter also analyzes two-dimensional factor scores produced using the thermometer score battery and results indicate that respondents' ideology will be distinguishable based upon vote and the majority party of her social context, implying that social context directly impacts political ideology as well. Using a method developed for multi-valued treatments, where the treatment is the interaction of the majority party in the discussion network and the congressional district, a balanced data-set is produced and subsequently analyzed to determine the presence of a significant contextual effect. The contextual effect is robust to model specification and the presence of other control variables,

including ideology. This presents a significant improvement to the literature, enabling the visibility of a contextual effect on vote choice regardless of political or demographic similarities between respondents.

8.2 Introduction

What makes a Democrat vote for a Republican candidate? This paper examines to what extent an individual's social context can impact her vote choice. For example, suppose an individual moved from St. George, Utah, to Santa Monica, California. St. George, Utah, is one of the most conservative cities in the United States. Over 90% of the population in St. George, Utah, is white, 63% of all households are married couples, and over 80% of the voters cast their presidential vote for George W. Bush in November 2000. Santa Monica, California, is one of the most liberal cities in the United States. Not only is their mayor a member of the Green Party, but in November 2000 approximately 80% of the voters in Santa Monica cast their ballot for Al Gore. If a voter were to move from St. George to Santa Monica, likely the political preferences of her social group would change — in particular if she chose her social group based upon shared characteristics (presence of children, income, race, age, marital status) and not based upon politics. This chapter explores whether or not the social group politics have the potential to impact an individual's politics — and if so, to what extent a move such as the one described above has the potential to change an individual's vote choice.

A quick look at the 2000 American National Election Study (ANES) reveals that quite a number of liberals do vote for Republican candidates and vice versa. Using the respondents' reported vote choice for president and ideological self-placement, there are 135 self-identified liberals who voted for the Republican presidential candidate and 341 self-identified conservatives who voted for the Democratic presidential candidate out of a total 1176 reported voters in the 2000 ANES.[1] This paper will

[1]The question used here was the following: "We hear a lot of talk these days about liberals and conservatives. Here is a seven-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place yourself on this

explore motivations for partisans to switch their vote. One possibility is the particular charisma or competence of a candidate motivates a partisan switch. Another possibility is that these voters are in fact misunderstanding the ANES survey question which asks them to self-report their political ideology. The possibility explored in this paper, however, is that the voter's calculus incorporates contextual factors in addition to self-placement.

This paper attempts to examine both the effects of the immediate reported political network on the respondent as well as a broader measurement of her political network — her congressional district. Although neither of these measures are exact replicas of the respondent's true political network, they are the two best available measurements. This analysis incorporates the variables necessary to ensure that the characteristics which drive a respondent to select a particular discussion network or congressional district are controlled for, and thus the consequences of exposure to differing political ideologies may be inferred. Furthermore, this analysis will provide insight into the degree to which geographic averages (such as congressional district returns) are useful in describing the partisan discussion network of any individual. While a voter may reside in a congressional district where she does not agree with the majority of voters, it is possible for her to have a smaller discussion network in which she is exposed to only those opinions with which she agrees.

The political behavior literature has indicated that while it is possible to fit a traditional calculus of voting model which considers only the ideological distance between the voter and the candidate (Downs 1957), an improvement to this model incorporates control variables such as voter's social class, income, or race which are also good predictors of the respondent's vote choice (Alvarez and Nagler 1998). The logic behind these improvements is that the control variables are not random, but instead good proxies for factors which improve upon the measurement of voter's ideology. One other possibility for why these additional variables are key in determining the vote choice is that they represent a second component in the voting calculus. This

scale, or haven't you thought much about this?" The totals are 62 missing, 244 Democrat votes who self-identified as liberals, 135 Democrat voters who self-identified as conservatives, 341 Republican voters who self-identified as liberals, and 394 Republican voters who self-identified as conservatives.

chapter suggests that this component is the influence of a voter's social context on her vote choice — the additional variables are variables which determine her network selection.

The existing literature on this topic has noted correlations between a respondent's reported discussion partner's political ideology and her own, but has not provided a coherent theory describing precisely how a voter's context is able to influence her political behavior (Huckfeldt and Sprague 2004, Mutz 2006). This chapter assumes that voters choose their political networks based upon their characteristics, such as their education level, marital status, income, and whether or not they have children. After selecting a neighborhood network (a larger, neighborhood level network for which this analysis will use a congressional district as a proxy) and an immediate discussion network (close friends which are provided by responses in the 2000 ANES), the respondent will be exposed to political discourse. This chapter uses the available control variables to ensure that the consequences of both the larger, neighborhood-level network and the smaller, discussion-level network are visible without introducing additional correlations from selection into this network. Once controlling for network selection, contextual effects still emerge, suggesting that indeed it is possible for an individual's social network to change her political opinion.

The chapter proceeds with a discussion of the existing literature on individual's networks and contextual effects, and initial tables are presented which document the possibility that indeed context may play an influential role in determining vote choice. A description of the data used and the construction of a latent utility score is then presented. This provides additional encouragement as the latent utility — estimated here via factor analysis — appears to separate into four types based upon an individual's vote and particular partisan context. A simple logit analysis indicates that the contextual variables add explanatory power to the traditional calculus of voting model, which motivates the next section where the effect of the inclusion control variables on the contextual effects is presented both on the entire data-set and upon a balanced sample, with a particular statistical technique developed to handle the multi-valued treatment. It is then possible to conclude that, once controlling for

the selection of the network using the best methodological tools available, contextual effects do impact an individual's vote choice.

8.3 A Respondent and Her Network

Dating as far back as the Columbia School, political scientists have known that individuals use their social networks as resources for political information and to determine their political preference (Berelson, Lazarsfeld, and McPhee 1954; Lazarsfeld, Berelson, and Gaudet 1948; Eulau 1963). However, the traditional calculus of voting considers only the individual's preferences in the vote choice. Despite the inclusion of social network questions in several national surveys, the relationship between voters choices and the influence of their peers is still not well understood.

The 2000 ANES respondents describe their vote choice, characteristics, ideology, and partisan discussion network. I also utilize more information — the partisan congressional returns in their congressional districts, and measurement of their ideology based upon a factor analysis of their responses to the thermometer score battery. This enables a comparison to be drawn: To what extent does a larger neighborhood context (based upon congressional district returns) influence a voter's decision and to what extent does an immediate context (based upon the partisan nature of her discussion network) influence a voter's decision?

The political ideology of discussants with respect to the respondent can prove influential in determining the amount of civic discourse which occurs about politics. Political discourse, and the extent that discussion partners have similar political ideologies, has been studied quite carefully in the literature (Huckfeldt et al. 2004, Huckfeldt et al. 1998), and politically diverse discussion networks do appear to be sustainable. In the analysis of the 2000 ANES by Huckfeldt et al. (2004) comparing the consequences of the discussion network onto a respondent's candidate thermometer score, they find "that strong partisans are not immune to the political messages that are filtered through networks of political communication" (63). That the politics of the discussion network is influenced by its members, however, is not clear: There

may be an additional variable which not only determined whether or not they would become discussants but also which determines whether or not there is agreement or disagreement in the discussion network. However, the Huckfeldt et al. (2004) results are clear — there is correlation within discussion groups' political preferences. The immediate consequence of a politically diverse discussion network on an individual voter's ideological preferences is not, however, well understood.

A quick look at the breakdown of data available in the 2000 ANES demonstrates that indeed the respondents do live in politically diverse communities; that is, there are many individuals who voted for a Democratic presidential candidate but who live in a Republican congressional district, or who likewise voted for a Democratic presidential candidate but whose principle discussion network is composed of Republicans.[2] Table 8.1 describes the percentages and raw numbers available in the data-set which separate the majority political party in either the respondent's congressional district or discussion network and the vote into all four possible categories. The presence of individuals whose vote is different from the majority party in their network ensures that there are enough of these cases to study.

Table 8.1 Goes Here

The puzzle begins when the data is tabulated both by vote, by respondent ideology, and by social context. Table 8.2 describes the number of instances where a self-identified liberal voter who lives in a majority Republican congressional district has chosen to vote for a Republican candidate or where a self-identified conservative voter who has in a majority Democratic discussion network has chosen to vote for a Democratic candidate. These percentages are significantly higher than if the individual lived in, or had a network full of, individuals with similar partisanship. Looking at these percentages alone it seems possible that in fact voters are influenced by their context. Note that, for example, while only 1% of those voters who live in majority Democratic congressional districts and and who self-identified as liberals voted for

[2]Individuals are removed from the data-set if their discussion network is evenly split between Republicans and Democrats as it is then impossible to determine the majority party. Thus 126 individuals were dropped from the analyses.

the Republican presidential candidate, 6.55% of those voters who live in a majority Republican congressional district and who self-identified as liberals voted for a Republican presidential candidate. To what extent are voters influenced by others within their congressional district? Within their immediate social circle? Are their friends in fact imbedded within their congressional district?

Table 8.2 Goes Here

These results are presented in Figure 8.1 for closer inspection. The height of each of these bars indicates the percentage of individuals who fell into this category who voted for Bush. The categories are described at the bottom of each bar — the ideology of the respondent is listed first (either Liberal or Conservative) followed by the majority party category (Democratic or Republican). The first bar in each category denotes the congressional district effect, and the second bar denotes the discussion network effect. Note here that if there were no effect of congressional district or discussion network, we would anticipate the height of the bars on the far left, which indicate the percentage of individuals who voted for Bush but who are self-identified as liberal and live in a majority-Democratic congressional district (light blue) or have a majority-Democratic discussion network (purple) should be the same height as those immediately to their right, who are also self-identified as liberal but live in a majority-Republican congressional district or who have a majority-Republican discussion network. However, we observe a large shift in the percentage of Bush voters once the congressional district or discussion network changes majority party. This effect is identical for respondents who are self-identified conservative voters.

Figure 8.1 Goes Here

These results are next tabulated by Bush vote while examining only the majority party discussion network in Figure 8.2. Here again the height of each bar denotes the percentage of individuals who fall into that category who voted for Bush. Note that the greatest percentage of Bush votes come from individuals who report having an all-Republican discussion network. Individuals who reported having a mixed-party

discussion network fall into the middle category, and individuals who report having an all-Democratic discussion network fall into the lowest category.

Figure 8.2 Goes Here

All of this suggests that individuals are increasingly voting like those around them, and that the social context explanation for vote choice is potentially appropriate given the tabulations of the data seen above. One possibility is that voters are increasingly able to seek out like-minded individuals as it becomes increasingly easy to maintain connections across larger geographic distances. Additionally, if it is the case that a voter's network is persuasive, then it is important to understand the role of diversification in networks for ideological stability.

In February 2004 the Pew Internet and American Life Project conducted a survey on social ties in America. In the report on these survey results they find that "traditional orientation to neighborhood- and village-based groups is moving towards communities that are oriented around geographically dispersed social networks" (Pew Internet and American Life Project 2006, 2). The Pew results also indicate that respondents were likely to get advice from people online. Many claim that Americans are increasingly "bowling alone" (Putnam 2000) and replacing neighborhood social interactions with online social interactions (Sunstein 2001). If the internet and other technological changes (such as nationwide long distance cellular phone plans) begin to change with whom voters interact, this is likely to affect their voting choices. In particular, if respondents are more likely to have discussion partners who are geographically distant, then knowledge of the respondent's geography is not sufficient to control for the information she would receive from other voters. It is then crucial to know something about the structure of her particular social network if we believe that voter-to-voter communication impacts voters choices. Most survey data fails to include questions regarding the respondent's social ties. However, this seems like a likely place that a voter may actually be influenced.

Huckfeldt and Sprague (1987, 1988) and Huckfeldt, Sprague, and Levine (2000) did conduct surveys on political discussion partners. These surveys asked an initial

respondent a series of questions and also asked each respondent for the contact information of their discussion partners. The researchers then contacted the discussion partners and administered the same survey. This type of survey is referred to as a "snowball" survey. The existing snowball surveys on political discussion partners are limited, both in terms of geographic scope and by being conducted for only a few specific election cycles. Furthermore, only non-relatives are considered in the scope of political discussion partners. The General Social Survey (GSS) did complete a social network battery in 1987 and 1988, but there is little to identify the network political ideology or the respondent's political ideology. Finally, the 2004 ANES asked respondents if they ever discuss politics with their family or friends, and the frequency with which that discussion occurs. However, these questions fail to include the partisan nature of the discussants which plays a crucial role in determining the respondents' preferences and information set.

Both the fields of political science and sociology have benefited from inclusion of questions attempting to analyze details of the respondent's political network. The GSS social network battery produced innovations in the analysis of network survey data (Marsden 1990), an increased understanding of the structure of political discussion networks (Marsden 1987), and an increased understanding of the relationship between organizational affiliations and network density (Liedka 1991). It provided data from which to analyze with whom people discuss politics, the frequency of that discussion, and the impact of political discussion on political participation (Straits 1991; Knoke 1990). The ANES social network battery lead to an increased understanding of the role between disagreement and social ties (Huckfeldt, Johnson, and Sprague 2004).

8.4 Data and Variable Construction

8.4.1 Voters

Data in this analysis comes from the 2000 ANES and is augmented by partisan congressional district election return data.[3] The congressional district returns from 2000 are used as proxies for the partisan nature of a respondent's congressional district. Although congressional districts are clearly too large in some cases (in states such as Montana, for example) to serve as adequate proxies for an individual's neighborhood political experience, they are the smallest level of geographic identification available in the data-set. Thus, they will be used to control for the effect of the larger context within which the respondent receives political information. Each respondent is asked a series of questions about the people with whom which she discusses important matters, and the respondent is asked to identify how she believes her discussion partners voted in the 2000 election.[4] These variables describe the smaller context within which the respondent receives political information. This chapter will examine to what degree both the discussion network and the congressional district affect the respondent's vote choice. There is a great deal of variance in response for the number and partisanship of discussion partners as well as vote choice — this variance provides leverage within which to estimate the contextual variable effects. Of the respondents, only 18% could name four discussants. Over 40% of all discussants are relatives, 25% are coworkers, 10% attend the same church, and 20% are neighbors. The average respondent speaks to the discussant a couple of times per week and believes the discussant is slightly informed about politics.

Each respondent is asked a series of questions which are used as control variables

[3]The 2000 ANES data is provided by ICPSR. Congressional district returns for the 2000 election were provided by Gary Jacobson.

[4]The questions used here are a name generator and then a question as to how the respondent believes that individual voted. They are "From time to time, people discuss government, elections and politics with other people. I'd like to ask you about the people with whom you discuss these matters. These people might or might not be relatives. Can you think of anyone?" and then using the names generated by this question, "How do you think NAME1 voted in the election? Do you think he/she voted for Al Gore, George Bush, some other candidate, or do you think NAME1 didn't vote?"

in the analysis. These questions include the respondent's marital status, income, age, race, gender, highest level of education achieved, current employment status, length of residence in current community, whether or not the respondent has children, and whether the respondent has a good current financial situation. These questions resemble most closely the vector of characteristics on which a voter will select a neighborhood and friendship group and thus these particular characteristics will be important to balance the data-set as they are most likely to be correlated with the particular contextual treatment. The average individual in this data-set is married, has some college education, has children, is female, is in good financial status, is white, is ideologically moderate, has lived in her community for 27 years, has a moderate income, and is 47 years old.

8.4.2 Latent Utility and Thermometer Scores

As described in the introduction, there are many self-identified partisans who voted for the opposite party in the ANES. In order to be able to distinguish a contextual difference across ideologies, a more nuanced measurement is needed than the seven-point scale self-reported ideology. For this analysis, a two-dimensional latent utility score is constructed for each respondent to be used as an ideology proxy.[5] This variable will capture more of the ideological preferences of each respondent than party identification, and it is highly correlated to party identification. This variable is constructed using the thermometer score questions on the ANES and three supplemental variables. Each respondent is asked how they feel, where 0 is the coldest and 100 is the warmest, about a series of particular groups. There are 32 total thermometer score questions which range from questions evaluating feminists to welfare recipients.[6] Also included in estimating the respondent's ideology are the strength of support for

[5]Future work will include using Bayesian IRT to estimate these scores as well, so as to incorporate variance in the estimate as a measurement of certainty.

[6]The questions include evaluating Asians, feminists, Protestants, Jews, Catholics, the Christian Coalition, homosexuals, the environment, older people, women, fundamentalist Christians, Hispanics, welfare recipients, the poor, big business, labor unions, liberals, conservatives, whites, blacks, the federal government, the military, Congress, the Supreme Court, George Bush, Jesse Jackson, Ralph Nader, Bill Clinton, Republicans in the House, Democrats in the House, Al Gore, and George W. Bush.

the president, whether or not there was Democratic Party contact, and whether or not there was Republican party contact.

The two-dimensional factor scores, by individual, are plotted below. Each individual is assigned a color in the plot based upon both whether or not she resides in a majority Democratic congressional district, as calculated using the 2000 congressional district election returns, as well as upon her presidential vote choice. Initially factor scores are plotted only for those who live in a majority Democratic congressional district and who voted for the Democratic presidential candidate (blue), as well as for those who live in a majority Republican congressional district and who voted for the Republican presidential candidate (red). Figure 8.3 indicates that there is a clear split between these groups, which gives the factor scores some credence in actually describing ideology.

Figure 8.3 Goes Here

Next all four categories are plotted by color, so that individuals who reside in a majority Democratic congressional district and who voted for the Republican candidate (green) and who reside in a majority Republican congressional district and who voted for the Democratic candidate (orange) are included as well. Note that in Figure 8.4 the individuals sort into four categories, blue/green/orange/red, based roughly upon their contextual demographics, with the DD group at the top left and the RR group at the bottom right. The split along both vote and context is more clearly seen when separated out using the majority party within an individual's discussion network as opposed to congressional district. The figure using discussion network context is produced in Figure 8.5 and indeed the colors again split blue/green/orange/red. Ideology appears to be influenced by both the majority party of the voter's discussion network as well as by her congressional district.

Figure 8.4 and Figure 8.5 Go Here

In order to examine whether or not these scatterplots are indeed splitting into four groups based upon a respondent's vote and social context, each category is collapsed

into a single mean and 95% CI. That is, the mean factor score for all the respondents who both live in a Democratic congressional district and voted for a Democratic presidential candidate is calculated, and likewise for the other three possible pairings. Then the means and 95% confidence intervals of the scores by category (DD,DR, RD, RR) are plotted. Figure 8.6 describes the case where the category incorporates the congressional district contextual effect. This plot indicates that visually the means of the four categories are distinguishable. The two categories of Republican voters, however, are much closer together than are those for Democratic voters. Figure 8.7 is an identical plot, with the exception that the categories are described using the discussion network as the contextual effect. Here, again the means for the Democratic voters are visually distinguishable (when broken down by context), but the Republican voters are not.[7] These figures demonstrate that there is indeed separation based upon context for individuals who made the same vote choice — for Figure 8.6 there are four distinct means, where both Republican and Democratic voters are separated from each other by their congressional district partisan breakdown. In Figure 8.7 there is less separation, with the Democratic voters still being slightly distinguishable but the Republican voters no longer distinguishable by partisan breakdown of the discussion network. Both of these graphs are particularly interesting because they suggest that in fact context does appear to be correlated with ideology, despite similar candidate choices. However, note that there should be no significant substantive interpretations made here as no control variables are included, which might account for correlation in preferences and context.

Figure 8.6 and Figure 8.7 Go Here

Table 8.3 and Table 8.4 compare the means of all the factor scores to evaluate whether or not they are different. This comparison will yield a a precise measurement of the difference in means across treatments which corresponds to the intuition from looking at the previous figures. An examination of whether or not any of these

[7]Note the scale of the axes are not symmetric, so that in fact these graphs have been drawn to highlight the fact that there are still very small differences in the first factor.

differences in means are statistically significant in Table 8.3 finds there is no statistically significant difference between RD and RR using the congressional district as a context for the first factor, but that DD and DR are indeed different. All other comparisons have statistically different means, including those for the second factor. Table 8.4 demonstrates similar differences. The only pair of means here which are not statistically different are the means DR and RR in the first factor. These t-tests suggest that context is indeed playing an influential role in determining ideology, which should then influence vote choice.

Table 8.3 and Table 8.4 Go Here

Simply plotting the factor scores and associating them with a particular color does not imply any causation — the respondents' scores could be sorted due to something other than contextual effects, such as the influence of income. In order to examine whether or not the scatterplots in fact reveal information about the respondent's social context, the scores are then used, individually, as dependent variables in a linear regression with Democratic vote share and presidential party choice as the two independent variables. Because the dependent variables here are estimated quantities, robust estimation is appropriate.[8] The robust linear regression coefficients also are not proof that the voter's context determines her ideology — but if the coefficients for context are not large and statistically significant in this analysis, then context is unlikely to impact voter ideology.

Table 8.5 Goes Here

In Table 8.5, however, it is clear that the percent Democratic of the congressional district and percent Democratic of the discussion network is not determining the ideological sorting. The contextual variables have small coefficients, and only the percent Democratic of the congressional district is statistically significant for the second factor. Given that ideology (as measured using the thermometer scores) does

[8]The additional error introduced from the estimation of the dependent variable implies that either FGLS or robust estimation is necessary to eliminate the effects of heteroskedasticity. In this case robust estimation of the standard errors corrects for this, and I use Huber-White standard errors.

initially appear to sort itself by some measurement of context, the analysis proceeds to more complicated methods in an attempt to determine to what extent a voter's context will influence her vote choice. Additional variables will be included, such as race, income, and marital status, which are likely to influence an individual's contextual choices.

8.5 Methods

8.5.1 Calculus of Voting with Contextual Effects

In order to test that both the congressional district partisan breakdown and the discussion network are influential on the voter's candidate choice, a simple set of logit coefficients are produced. Table 8.6 presents coefficients from both the partisan congressional district variable as well as the partisan discussion network variable. Each of the covariates are included in the analysis, and the full results can be seen in Table 8.9.

Table 8.6 Goes Here

The dependent variable in the analysis in Table 8.6 is an indicator as to whether or not the respondent voted for the Republican presidential candidate. Ideological distance from the Republican candidate is established using the respondent's self-placement and placement of the Republican presidential candidate. Congressional district Democratic vote share and the number of Democrats the respondent reported in her discussion network are also included as independent variables. The coefficients produced here are sensible — as distance increases, the voter is less likely to vote for the Republican. Also, the coefficients indicate that context does in fact impact the voter's choice. As the percent of Democrats increases in the voter's congressional district, the voter is more likely to vote for the Democratic presidential candidate, and the same is true as the number of Democrats in the discussion network increases. Thus, preliminary indications are that the voter's context is important. This analysis

is very similar to Huckfeldt, Johnson, and Sprague's work (2004), with the exception of the inclusion of the congressional district context.

Not only do the 2000 ANES survey respondents provide a great deal of information on their vote choice, ideology, and discussion network, but they also respond to a series of questions which describe their own characteristics. Respondents are asked to provide their race, age, gender, marital status, and a large number of other variables which are likely to be correlated with their ideology and thus affect their vote choice. Inclusion of these variables has improved upon the ability of political scientists to predict an individual's vote choice beyond the simple calculus of voting (Alvarez and Nagler 1998), and should be included as control variables in any analysis. The section below details why results are presented both on the existing data from the 2000 ANES where vote choice is explained by ideology, context, and controls, as well as on a balanced data-set.

8.5.2 Matching with Multi-valued Treatment

There has been great concern in the political methodology literature about empirical analysis on data that does not derive from a randomized experiment. Without the ability to compare the effects of a treatment on similar subjects, it is difficult to know whether or not the observed differences are the result of the treatment or instead the consequence of differences in the subjects. Matching to produce a balanced data-set ensures that there are no observable differences between the subjects who have received any particular treatment, and then permits an analysis of a treatment effect that does not rely upon a particular model specification.

In the study the treatment assignment is not random but instead the respondent has selected a particular treatment. Thus there is valid concern that the treatment may be related to an observed characteristic of the voter. Specifically, the voter will select a neighborhood and a social group based upon her vector of characteristics (children, marriage, gender, race, etc.) and these variables may also be associated with particular political ideologies. While including these characteristics as control

variables in a regression does ensure that their influence on a respondent's vote choice is incorporated, the disadvantage of including these variables as controls is two-fold. First, it requires a specific linear functional form which may not be applicable. Second, it requires an assumption that the group that received a particular treatment did not produce an outcome which is correlated with any of that group's characteristics (and thus would be different for any other group which had received the treatment). Eliminating the effects of these characteristics in the analysis would greatly improve an estimation of the effects of contextual ideology on voters.

In this analysis the treatment variable is the pair (majority party of the discussion network, majority party of the congressional district). This means that there are four possible treatments for each respondent — (Democratic discussion network, Democratic congressional district), (Democratic discussion network, Republican congressional district), (Republican discussion network, Democratic congressional district) and (Republican discussion network, Republican congressional district) — hereafter labeled DD, DR, RD, and RR.[9] The quantity of interest is the effect of each treatment (context) on each respondent's vote choice, with the interesting comparison of how any one particular treatment affects votes compared to any other. However, most of the existing literature that examines treatment effects deals only with a binary treatment case — where the subjects are either treated or not.[10] In this case there are four treatment variables so the binary treatment processes which produce a balanced data-set are not applicable. Therefore, two existing processes are modified slightly to produce a new algorithm to incorporate multiple treatments.[11] The intuition behind this process is to produce something similar to matching based upon a single propen-

[9]Here clearly the partisan breakdown of the congressional district is a continuous variable which has been broken into two discrete cases. Although examination of a continuous treatment would be an improvement on this analysis, the analytic techniques have not been developed to handle a continuous treatment. Thus, although the discrete cases are a simplification, they are a first step towards balancing the data by contextual treatment, and future work will hopefully address this problem.

[10]For a good review of the literature, please see Alexis Diamond and Jasjeet Sekhon, "Genetic Matching for Estimating Causal Effects", (unpublished manuscript) http://sekhon.polisci.berkeley.edu/matching/.

[11]Two studies for multi-valued treatments include one using balancing (Imbens 2000) and one using sub-classification instead of matching for multi-valued treatments (Imai and Van Dyk 2004).

sity score; use of the propensity score is discussed for sub-classification with the binary case in Rosenbaum and Rubin (1984) and introduced with matching in Rosenbaum and Rubin (1983). The methodology developed here is an extension of the generalized propensity score literature where instead of matching based upon a single propensity score, the matching algorithm will be based upon a multi-valued propensity score — a vector of propensity scores for each individual which represent the probability of receiving each treatment. Using this vector, the matching algorithm will produce a new data-set, where observations are selected into the new data-set based upon the similarities of their propensity vectors (a multi-valued propensity score). This process should reduce any bias in the estimates that would have occurred from the lack of a randomized experiment. As it is simply not possible to randomly assign voters to congressional districts and discussion networks, this process will help to compensate for the lack of random assignment by producing a new data-set which attempts to mimic the data-generating process that would have resulted from a randomized experiment.

Again, note that the number of treatments (contexts) is four — these treatments reflect the possible combinations of congressional district and discussion network possibilities. Before producing a balanced data-set, there are not an equal number of observations across treatments — this is because voters select their treatment. While the sample is from the ANES (and is thus a random probability sample from across the nation), the sample was chosen to be a random sample of the set of eligible voters in the U.S. and was not chosen so that the sample would be based upon a voter's context.

Using the sample, the first step is to produce a propensity vector — a vector of the conditional probability of assignment for each particular treatment given the vector of characteristics (self-placement ideology, the 2-dimensional factor scores, gender, age, race, marital status, children, income, length of residence in the community, current financial status, highest education achieved, and employment status). Treatment is used as the dependent variable and the covariates included above are used as independent variables. The data is fit using multinomial logit and the predicted probabilities are calculated for all observations. These predicted probabilities are then

used as a vector with the same number of values as there are treatments. This vector represents a multi-valued propensity score. While in the binary case the propensity score would reduce the dimensionality of the covariates into a single value, this process to analyze multi-treatment data is also effective at reducing dimensionality (from covariates to treatment vector) so long as there are fewer treatments than there are covariates. Each entry in the vector is the probability of getting that treatment given the observed covariates (i.e., the first entry in the vector is the probability of getting the first treatment). Rosenbaum and Rubin (1983) prove that matching observations based upon their propensity score is equivalent to matching on all observed covariates — the matching process will here also consist of matching based upon the propensity score in an effort to achieve balance in the covariates.

The second step is to sort the observations into different treatment groups. Note that each observation will have a propensity vector associated with it. Once the data is organized by treatment and every observation has a propensity vector, the matching process begins.

First, a single observation is randomly chosen from the data in the first treatment. Then the best observation in the second treatment is found to minimize the distance between the two observations in terms of the propensity vectors. Let x_1 and x_2 be defined as two propensity vectors. The two observations are selected because they minimize the *pair distance* between the two observations compared to any other observation in the second treatment. Formally, this is defined as:
$d_{PAIR}(x_1, x_2) = \sum_{i=1}^{c}(x_{1i} - x_{2i})^2$, where c is the number of values in the propensity vector.

This notion of the distance between two pairs is intuitive; the distance between the two propensity vectors is the sum of squared distances between each entry in the propensity vector.[12]

Next an observation in the third treatment is found that is also the closest (in terms of the propensity vector) to both of the two observations already selected (from

[12]Note here that these calculations below were also examined using absolute distance as well and that this notion of paired distance produced similar results. In general any notion of distance would be appropriate.

the first and second treatment) in the group. In order to find the "closest" observation, it must be the observation with the smallest group distance. Formally this is defined as:

$d_{GROUP}(x_1, x_2, \ldots, x_g) = max[(d_{PAIR}(x_i, x_j) \forall (x_i, x_j) \in (x_1, x_2, \ldots, x_g)$ where $i \neq j]$.

The observation in treatment three is found that minimizes group distance. This process is repeated for each additional treatment so that an observation is selected from each treatment; once observations from each treatment are selected, this group is considered matched. This group is then discarded from the larger data-set and the process is repeated, drawing one observation randomly from the observations which received treatment one and finding the next set of observations in the remaining treatments which minimize group distance.

Observations are selected until the paired-t-tests on the means of the covariates begin to produce values which indicate that there are statistically significant differences. Note here that it is possible that these tests may be misleading — for example, it is conceivable that randomly discarding data might result in acceptable t-statistics while yet not improving balance (Imai, King, and Stuart 2005). Therefore, another method of evaluation is necessary to compare whether or not the new sample has indeed improved balance and will help in eliminating bias.

Thus two inspectations of the data will take place to evaluate whether or not the new sample has improved balance. First, differences in means across treatments are compared so as to observe improvement in the balanced data-set. Then each covariate is examined visually, where the mean, plus or minus 1/2 of the standard deviation, are plotted for a single covariate for each treatment in both the original sample and the balanced sample.

Figure 8.8 Goes Here

This is a particularly simple summary of the data. Note, however, that the mean and 1/2 standard deviation interval of the covariates in the original sample often have less overlap than in the balanced sample.[13] This indicates that there are significant

[13]Graphs of the remaining covariate means and 1/2 standard deviation intervals are included in the Appendix in Figure 8.9 and Figure 8.10.

differences across particular treatments which may result in biased coefficients. Examining the balanced covariates, the means are often grouped closely together and generally the 1/2 standard deviation intervals overlap across treatments (Ho et al. 2005). According to Ho et al.,

> "One particularly simple low dimensional summary compares the mean of each variable in X (the set of covariates) for the treated group with the mean of each variable in the control group. The smaller these differences are the better. One rule of thumb that has been offered is if one or more of these differ by more than half a standard deviation of the respective X variable, then better balance is needed, but finding "small" imbalance in the original units is the real goal."

However, the means are not identical across treatments, they are simply closer than they were before the matching algorithm was applied to the data. This implies that it may be necessary to not only examine the treatment effects but also to examine a parametric model in order to examine the contextual effect.

8.6 Results

As Table 8.6 demonstrated, there is reason to believe that a voter's particular context will be influential in determining her vote choice. However, the simple logit required a particular function form to incorporate the control variables. That is, the work done to this point has not been able to distinguish to what extent a voter's ideology is similar to her social context because she has chosen people who are similar in terms of characteristics (and hence similar in terms of ideology) and to what extent a voter has been influenced by her peers.[14] Logit coefficients are produced in Table 8.7 which incorporate all the variables available to control for the selection of a context by the

[14]Future work will include estimating a structural model for the vote choice which incorporated the consequences of the respondent's ideology on her immediate discussion network, as well as the contextual effect observed on the factor scores. For now, the analysis simply assumes that voters have not chosen their congressional district or discussion network because of their partisan preferences.

voter on the unbalanced data and balanced data.[15]

Table 8.7 Goes Here

Again the dependent variable is an indicator for whether or not the respondent voted for the Republican presidential candidate. For the unbalanced data, several of the control variables have statistically significant coefficients at traditional levels — employment status, financial status, and age. The only contextual variable which has a statistically significant coefficient is the percent of Democrats within the respondent's discussion network — and this coefficient is positive as expected. The coefficient on the Democratic percentage in the congressional district is also positive but not significant. The coefficient on self-placement is negative as expected (liberal is closer to 1, conservative is closer to 7) but is also not significant. The two factor scores are both significant with one positive and the other negative. It is difficult to know how to interpret the directionality of these coefficients — looking back to the robust linear regressions in Table 8.5, note that both of these variables are highly correlated with vote choice, and the second factor score is indeed correlated with the percent of the congressional district that is Democratic. It is possible that the factor scores are measuring both the voter's ideology and the voter's congressional district partisan breakdown.

Using the algorithm described earlier for multi-valued treatment data, a new data-set is produced which has better balance within the covariates. An identical model is fit using logit on this balanced data. Here the disadvantage is clearly the smaller sample size. However, these new coefficients present two significant improvements. First, these are observations which are similar that have received each treatment so that the outcome (vote choice) is not driven by any of the control variables. Thus the observed coefficients on the contextual outcomes are not correlated with the variables which determine the selection of the social context — an improvement over all existing work. Second, producing a balanced data-set permits the examination of

[15]Future work will incorporate multiple imputation for those respondents who failed to answer all questions from which the control variables are drawn so as to increase the available sample size.

a treatment effect without fitting a particular model specification. The results from using the balanced data are that the impact of the discussion network is much *smaller* and that the impact of the congressional district is slightly *larger* than seen using the unbalanced data. Note that the 95% confidence intervals around the congressional district values do include zero, but that the discussion network coefficient is statistically significant. These results are surprising given that it has generally been thought that the selection of a particular geographic context would produce the correlation in vote choice. However, it is clear that discussion network plays a role in candidate selection.

The trend for the coefficients from the analysis on the balanced data-set are almost identical to the unbalanced coefficients. Again financial status and income are both statistically significant, as are the two factor scores. This is not surprising — it implies that when the data were balanced, they were not balanced perfectly so that there is still residual variance in these variables which helps to explain vote choice. But visually examining the mean and standard deviation covariate plots on the balanced data it was clear that the data were not identically balanced for all covariates. As before, the percentage of Democrats within the respondent's discussion network is significant and has the expected sign.

The conclusion from both of these tables of coefficients is difficult to summarize by simply looking at the coefficients. Looking at only the magnitude of the coefficients, it appears that the partisan breakdown of a congressional district does not seem to directly impact a voter's candidate choice, but the partisan breakdown of immediate friends does. As congressional districts are enormous entities (upwards of 600,000 people) it seems likely that a voter, even once ensconced within a highly partisan district, could have a friendship network that was mostly composed of the opposite party. The percentage of Democrats within a respondent's discussion network is only weakly positively correlated to the percentage of Democrats within a respondent's congressional district.

To clarify the consequences of the coefficients for the contextual variables, first consider the mean respondent. Her discussion network and congressional district are

approximately each half Democratic. She herself is slightly conservative when asked to self-identify her partisanship and her ideological factor scores place her in the middle of the distribution. She is likely to be married, to have some college education, to be employed, and to have lived her community for over twenty-six years. She is likely to be female, to have children, to be in a medium financial situation, to be white, and to be approximately 48 years old. Suppose she were to move from a situation where she lived in a congressional district that was 20% Republican to a congressional district that was 80% Republican. Using the unbalanced data, this would decrease the probability that she voted for the Republican presidential candidate by 4% and using the balanced data, this would increase the probability that she voted for the Republican presidential candidate by 8%. Again, however, note that these results are not statistically significant at conventional levels. Suppose that instead she were to move from a discussion network that was 20% Republican to a discussion network that was 80% Republican. Using the unbalanced data, this would increase the probability that she voted for the Republican presidential candidate by 18% and using the balanced data this would only increase the probability that she voted for the Republican presidential candidate by 15%.[16]

Using a balanced data-set does yield different conclusions than the unbalanced data but it does still appear that the contextual effect from the voter's immediate discussion network exists, while the effects of congressional district are never statistically significant at traditional levels. We do observe a treatment effect for context, but only at the friendship level. Thus, suppose we were to consider treatment only one of three categories — having an all-Republican discussion network, having a mixed-party discussion network, or having an all-Democratic discussion network. It is then possible to replicate the analysis, producing a new balanced data-set using these treatment categories. The results from this analysis for the discussion network coefficient are presented in Table 8.8.

Table 8.8 Goes Here

[16]First differences were calculated using Clarify (Tomz, Wittenberg, and King 2001).

Here we observe a much larger effect of discussion network. Again were an average respondent in this data-set move from a situation where she had a 20% Republican discussion network to one where she had an 80% Republican discussion network, the probability that she would vote for the Republican presidential candidate would increase by 25%. This individual's characteristics are included as control variables in this analysis, variables which are traditionally incorporated into the calculus of voting, so that this effect is truly based upon the individual's partisan context. Thus this effect can truly be viewed as the marginal effect of social interaction on an individual's choice of presidential candidate.

8.7 Conclusions

So what makes a Democrat vote like a Republican? The answer, from this analysis, appears to be an individual's social context. Were an individual to move from St. George, Utah, to Santa Monica, California — and make friends with a randomly selected group of members of the community — the probability that that individual would vote for the Republican candidate in this analysis would increase by 25%. The results above have demonstrated that a voter's discussion network does influence her partisan choice for president while controlling for variables which influence the choice of discussion partners. Using a new technique developed for multi-treatment data, the results are consistent between the initial analysis (which requires a particular model specification) and the analysis on the balanced data. The voter's ideology appears to sort by contextual effect when using a factor score based upon the thermometer score battery in the ANES, but this is not necessarily due to discussion network or congressional district influence. There is, however, a congressional-district and discussion-network effect in determining presidential vote choice. This effect is observed when controlling for ideology and variables which may influence a respondents' choice of congressional district or discussion partners — her social network affects her vote. That this effect is robust to the inclusion of control variables and model specification gives credence to the previous work which has found correlation in discussion group

vote choice and ideology, but also begins to find evidence that indeed this correlation is caused by influence from the other actors within the voter's social network.

While these results demonstrate that a voter's social context will influence her decision, they are unsatisfying in that the process by which this influence occurs is still not observed. The variables provided in the ANES are proxies for a voter's social network and do not provide information on the strength of the ties between them, for example, or for whether the individuals within the network are known to each other as well as to the respondent. Also, these results make a compelling case for additional data to be collected on the partisan nature of individuals' discussion networks. It is clear that geographic proxies (such as the partisan breakdown for congressional district) may not be sufficient, especially in this technologically advanced age. Additional survey questions will be included on the ANES 2006 pilot study which will help to address these questions.

Previous research indicated that there was correlation between the partisanship of the respondent and her discussion partners. This chapter goes further and indicates that indeed there is a relationship between both the respondent's ideology and vote choice and her contextual environment and that this relationship exists even when controlling for the respondent's choice of context. This research has enormous implications for the causal process by which individuals may change their preferences. However, it does rely heavily upon the assumption that individuals choose their contextual networks for reasons other than partisanship. It seems unlikely that individuals choose their neighborhoods or friends because of their voting patterns, however. More likely, individuals choose their social context for many of the other variables which have been controlled for in this analysis, and that some of them are in fact correlated, but not completely determined by, partisanship.

This chapter additionally develops a new process for producing a balanced dataset when treatment is multi-valued. This new methodology has applications beyond that of discussion-network treatments and would be applicable to instances of either a treatment which was administered in multiple levels (so as not to be considered simply categorical) or a treatment which is, like the one considered here, multiple

valued. This methodology allows the researcher to draw simultaneous comparisons about the implications of each treatment while making the fewest possible parametric assumptions.

Future research will examine additional data-sets to determine whether or not the results presented here are consistent with other survey respondents at other points of time to determine whether or not partisan contextual effects are key variables in determining all voters' choices. Data-sets of particular interest include the 1987 and 1988 General Social Survey, the South Bend data-set accumulated by Huckfeldt and Sprague, and the Niemi-Jennings Youth-Parent Socialization Study.

8.8 Tables and Figures

Table 8.1: Respondents' Partisan Majority Network and Vote

Majority Party Congressional District, Vote	Percent	Number
DD	29.35	344
RD	22.28	262
DR	24.40	287
RR	24.06	283
Total	100	1176
Majority Party Discussion Network, Vote	Percent	Number
DD	30.10	354
RD	22.79	268
DR	13.69	161
RR	33.42	393
Total	100	1176

Cell entries represent the percentage and raw count of respondents who had the row combination of majority party contextual influence and presidential vote.

Table 8.2: Respondents' Presidential Votes by Ideology and Context

Self-Placement, Majority Party Cong. Dist	Liberal, Dem	Liberal, Rep	Conservative, Dem	Conservative, Rep
Presidential Vote	Percent	Percent	Percent	Percent
Rep	.89	6.55	21.18	18.85
Dem	12.48	15.89	18.40	5.75
Self-placement, Majority Party Discuss Net				
Rep	.72	7.54	11.49	27.72
Dem	12.92	15.98	18.85	4.76

Cell entries represent the percentage of respondents who had the column combination of self-reported ideology and majority party contextual influence, and row measure of presidential vote.

Table 8.3: Paired T-Tests On Factor Scores Grouped by Vote and Majority Party Congressional District Pair

Types	Difference in Means (Factor one)	Difference in Means (Factor Two)	N
DD, RD	.05	-.13*	606
DD, DR	.52*	-1.23*	631
DD, RR	.69*	-1.41*	627
RD, DR	.47*	-1.10*	549
RD, RR	.4*	-1.28*	545
DR, RR	.17*	-.18*	570
* = α = .05			

Table 8.4: Paired T-Tests On Factor Scores Grouped by Vote and Majority Party Discussion Network Pair

Types	Difference in Means (Factor one)	Difference in Means (Factor Two)	N
DD, RD	-.009	-.13*	622
DD, DR	.54*	-.96*	515
DD, RR	.59*	-1.5*	747
RD, DR	.55*	-.83*	429
RD, RR	.60*	-1.37*	661
DR, RR	.06	-.54*	552
* = α = .05			

Table 8.5: Robust Linear Regression Coefficients, Dependent Variable: Factor Scores

	Factor One	Factor Two	Factor One	Factor Two
Democratic Vote	.56*	-1.19*	.59*	-1.25*
	(.058)	(.05)	(.05)	(.042)
Percent Democratic CD	.003	-.005*	—	—
	(.002)	(.001)	—	—
Percent Democratic DN	—	—	-.0001	-.0001
	—	—	(.0001)	(.0001)
Constant	-.44*	.83*	-.2749	.64*
	(.089)	(.068)	(.04)	(.035)
N	1001	1001	1176	1176
*	$\alpha = .05$	$\alpha = .05$	$\alpha = .05$	$\alpha = .05$

Table 8.6: Logit Coefficients, Dependent Variable: Vote for Republican Presidential Candidate

Variable	Coefficient	First Differences (20% to 80%)
Percent Republican Discussion Network	1.25	.18
	(.36)	(.05)
Percent Democratic Cong. District	-.003	-.04
	(.008)	(.12)
Constant	2.72	—
	(1.49)	—
N	834	—
Pseudo R2	.72	—
Percent Correctly Predicted	80%	

Covariates: Ideological self-placement; 2-dimensional factor scores based upon thermometer score questions: marriage, education, employment status, length of residence in the community, presence of children, gender, financial status, income, race, and age.

Table 8.7: Logit Coefficients, Dependent Variable: Vote for Republican Presidential Candidate

Variable	Coefficient (Original)	First Differences (20% to 80%)	Coefficient (Balanced)	First Differences (20% to 80%)
Percent Republican Discussion Network	1.25 (.36)	.18 (.05)	1.01 (.48)	.15 (.07)
Percent Democratic Cong. District	-.003 (.008)	-.04 (.12)	.01 (.13)	.08 (.18)
Constant	2.72 (1.49)	— —	2.48 (2.05)	— —
N Percent Correctly Predicted	834 80%	—	273 81%	—

Treatment is the pair (majority party within discussion network; majority party congressional district). All other covariates are included (Congressional district returns; ideological self-placement; 2-dimensional factor scores based upon thermometer score questions; marriage, education, employment status, length of residence in the community, presence of children, gender, financial status, income, race, and age) but their coefficients are not presented here.

Table 8.8: Logit Coefficients on Balanced Data, Dependent Variable: Vote for Republican Presidential Candidate

Variable	Coefficient	First Differences (20% to 80%)
Percent Republican Discussion Network	2.18	.25
	(.61)	(.07)
Percent Democratic Cong. District	-.003	—
	(.01)	—
Constant	2.49	—
	(2.26)	—
N	258	—
Pseudo R2	.64	—
Percent Correctly Predicted	82%	

Treatment is only based upon discussion network categories. All other covariates are included (Congressional district returns; ideological self-placement; 2-dimensional factor scores based upon thermometer score questions; marriage, education, employment status, length of residence in the community, presence of children, gender, financial status, income, race, and age) but their coefficients are not presented here.

Figure 8.1: Self-Identified Partisanship Combined with Network or Contextual Effects

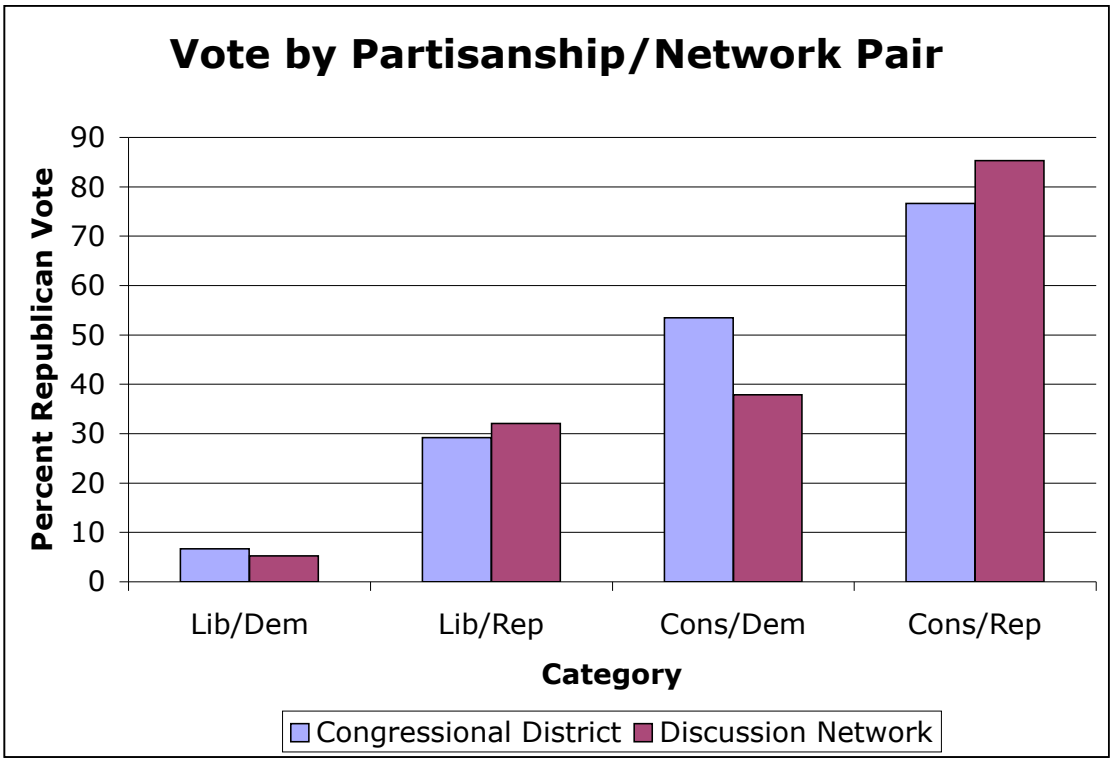


Figure 8.2: ANES: Percent Republican Vote by Discussion Network

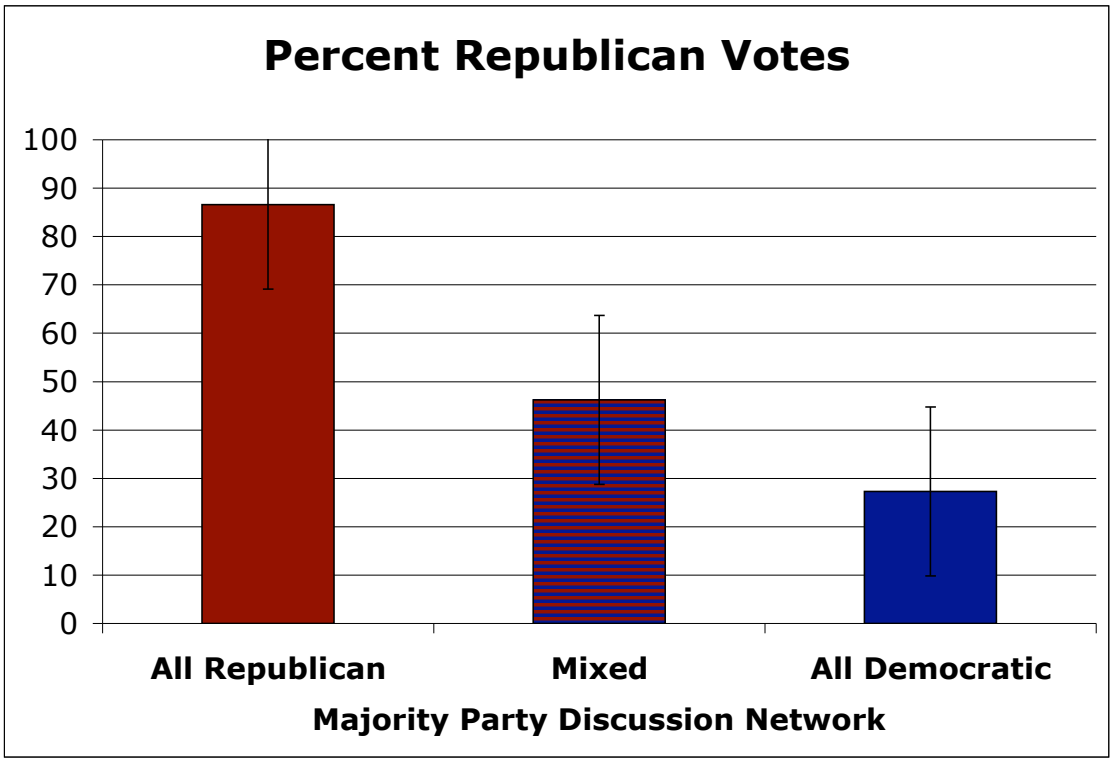


Figure 8.3: Pure Democrat and Republican Respondents' Two-Dimensional Factor Scores

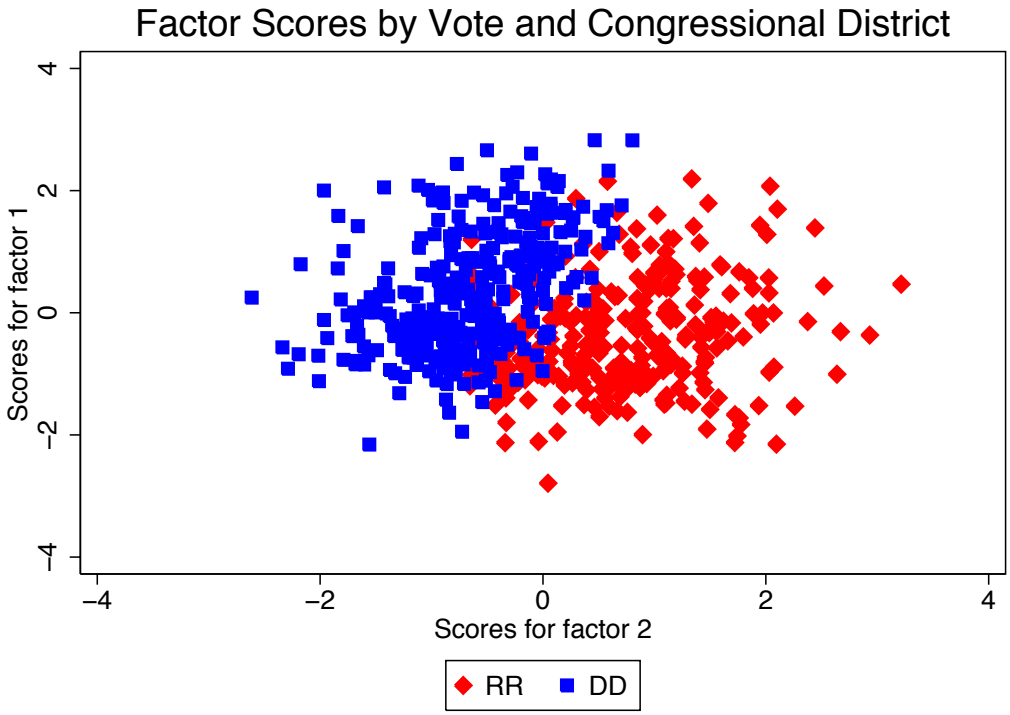


Figure 8.4: All Respondents' Two-Dimensional Factor Scores by Vote and Majority Party Congressional District

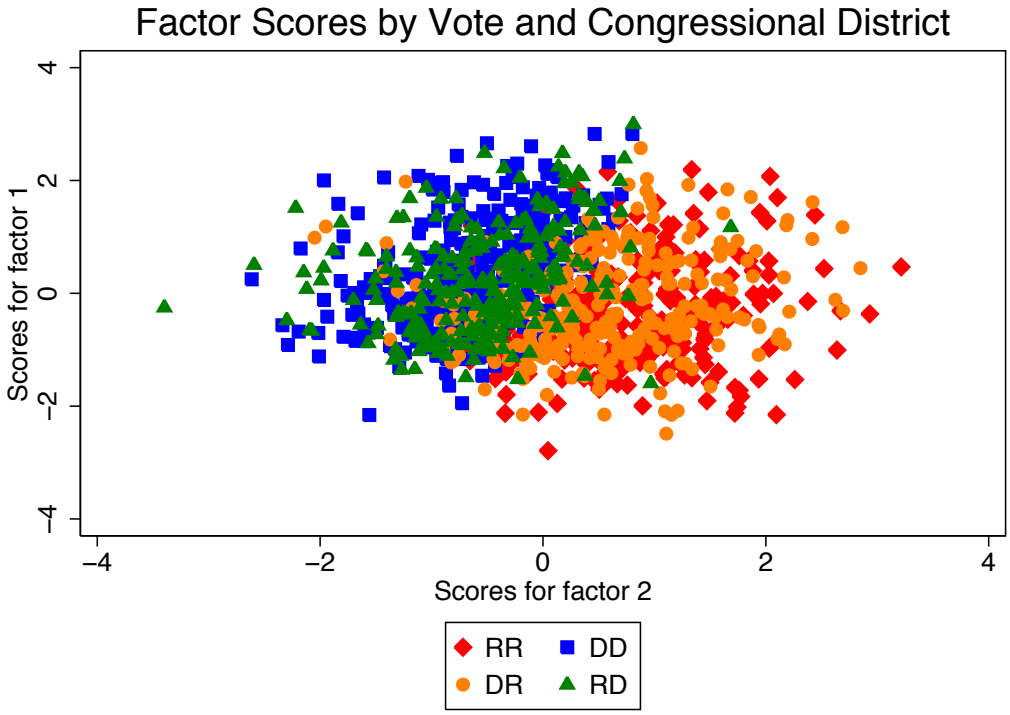


Figure 8.5: All Respondents' Two-Dimensional Factor Scores by Vote and Majority Party Discussion Network

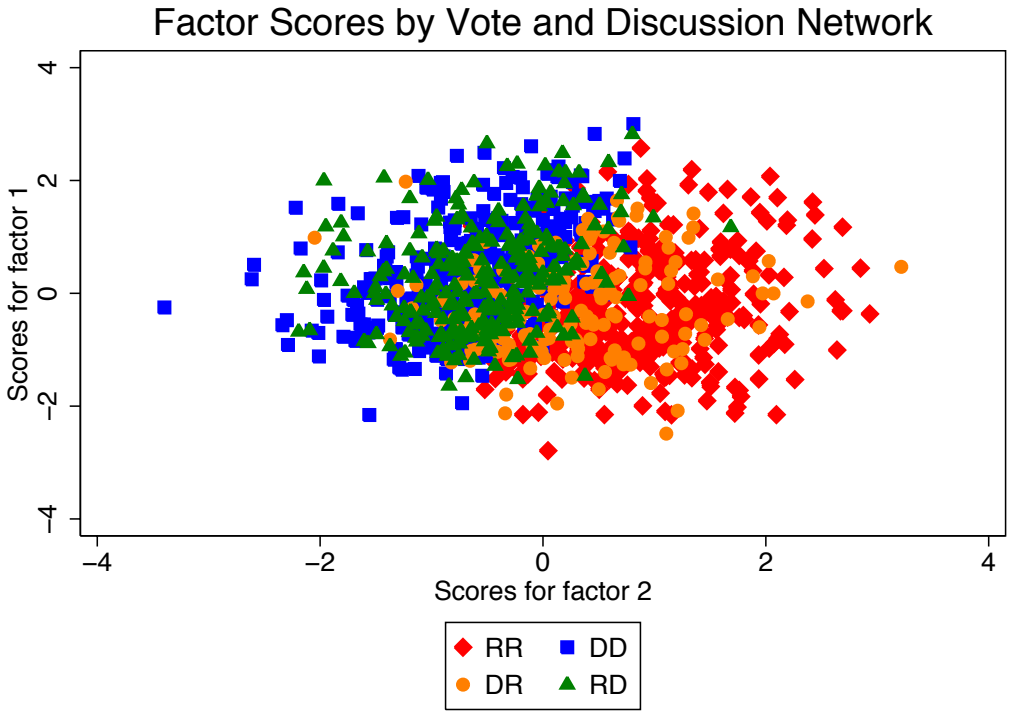


Figure 8.6: Two-Dimensional Mean Respondent Factor Score and 95% CI, Mean Calculated For Vote and Majority Party Congressional District Pair (from left-to-right, RR, DR, RD, DD, respectively)

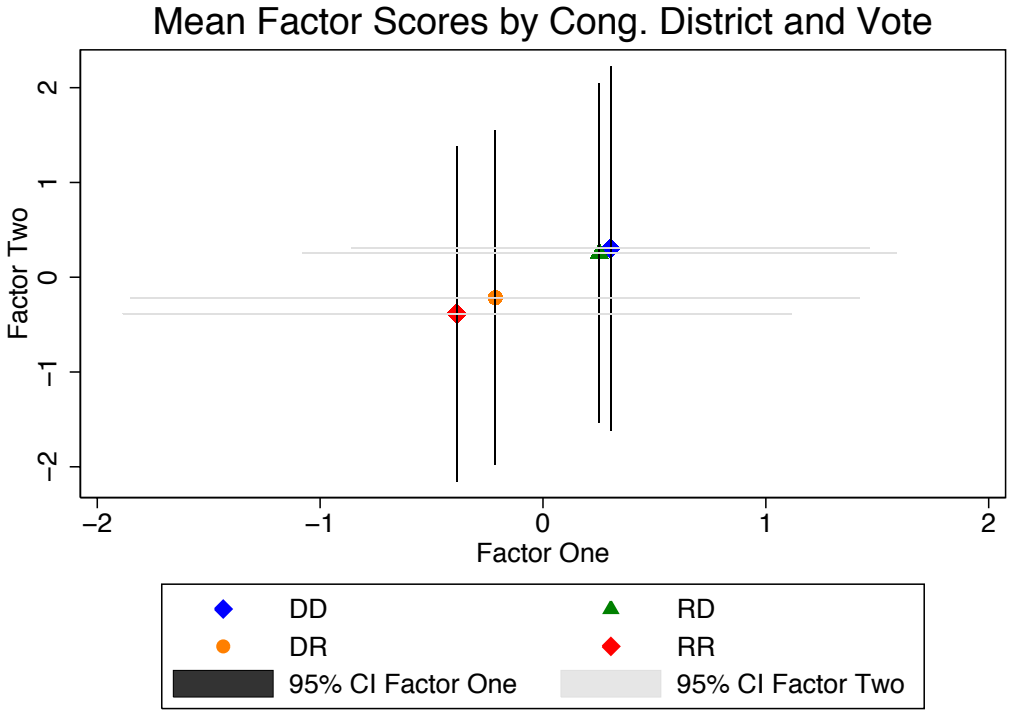
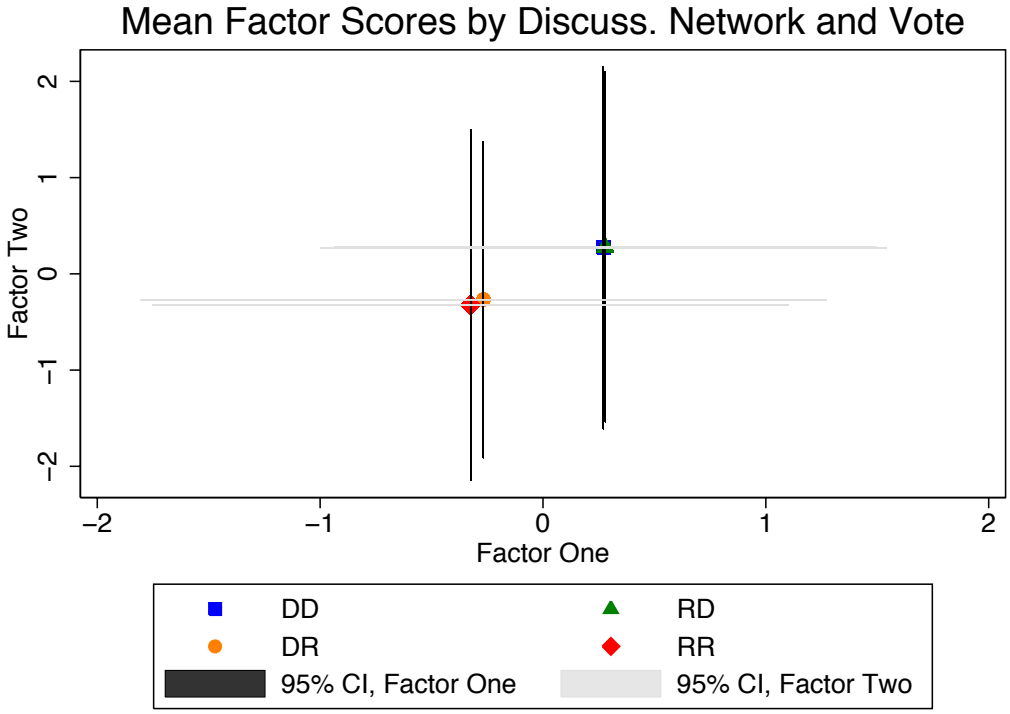


Figure 8.7: Two-Dimensional Mean Respondent Factor Score and 95% CI, Mean Calculated For Vote and Majority Party Discussion Network Pair (from left-to-right, RR, DR, RD, DD, respectively)



8.9 Appendix

Table 8.9: Original Logit Coefficients, Dependent Variable: Vote for Republican Presidential Candidate

Variable	Coefficient	SE
Percent Republican Discussion Network	1.25	(.37)
Percent Democratic Cong. District	-.003	(.008)
Factor 1	-.89	(.18)
Factor 2	4.91	(.45)
Self-Placement	.003	(.13)
Married	-.22	(.)
Education	.001	(.01)
Employed	-.25	(.34)
Length of Residence	.003	(.005)
Kids	-.57	(.37)
Female	-.04	(.30)
Financial Status	-.31	(.17)
Income	.13	(.16)
White	-.58	(.59)
Age	-.02	(.01)
Constant	2.72	(1.49)
N	834	
Pseudo R2	.72	

Table 8.10: Balanced Logit Coefficients, Dependent Variable: Vote for Republican Presidential Candidate

Variable	Coefficient	SE
Percent Republican Discussion Network	1.01	(.48)
Percent Democratic Cong. District	.01	(.01)
Factor 1	-.73	(.25)
Factor 2	4.78	(.72)
Self-Placement	.11	(.18)
Married	-.26	(.46)
Education	.001	(.01)
Employed	-1	(.51)
Length of Residence	.004	(.007)
Kids	-1.35	(.55)
Female	-.15	(.45)
Financial Status	-.49	(.28)
Income	.47	(.25)
White	-1.33	(.89)
Age	-.02	(.02)
Constant	2.48	(2.05)
N	273	
Pseudo R2	.59	

Figure 8.8: Unbalanced vs Balanced Data

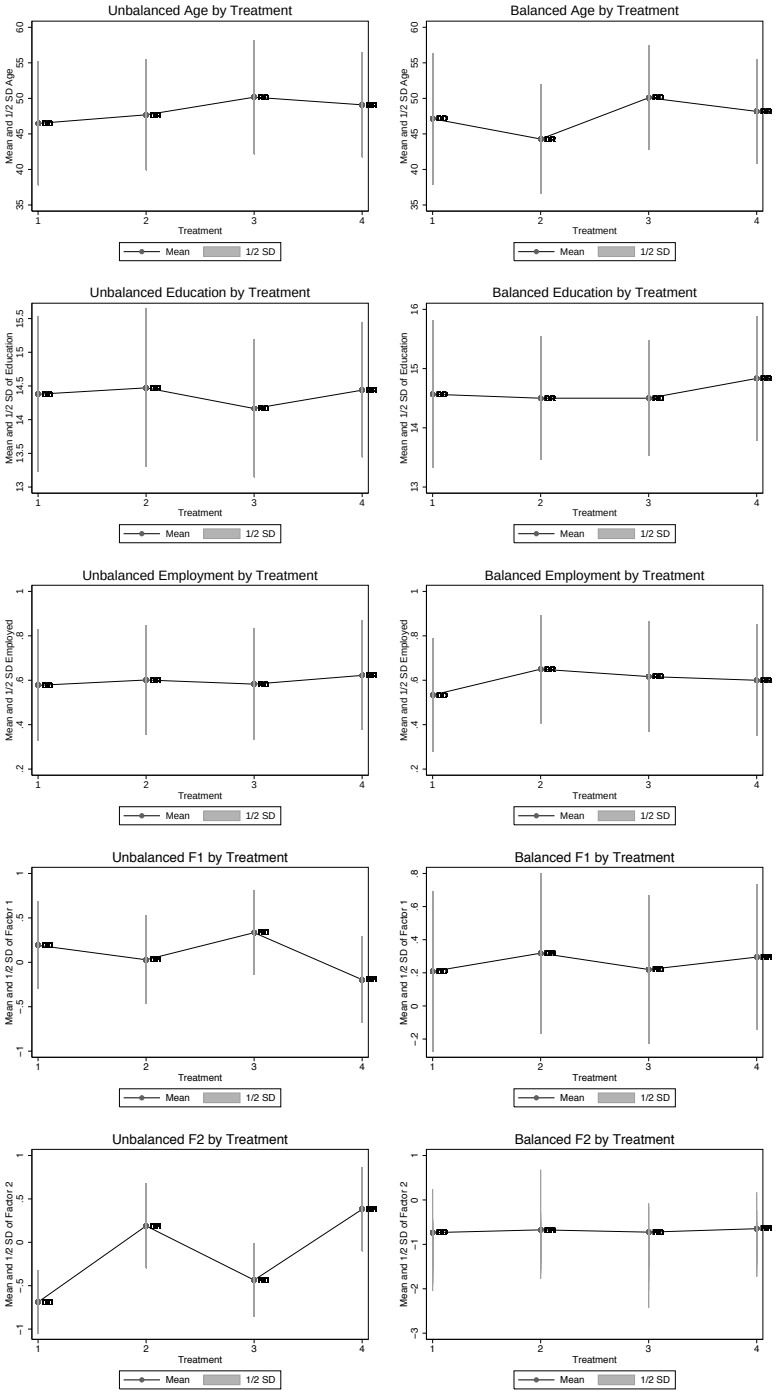


Figure 8.9: Unbalanced vs Balanced Data

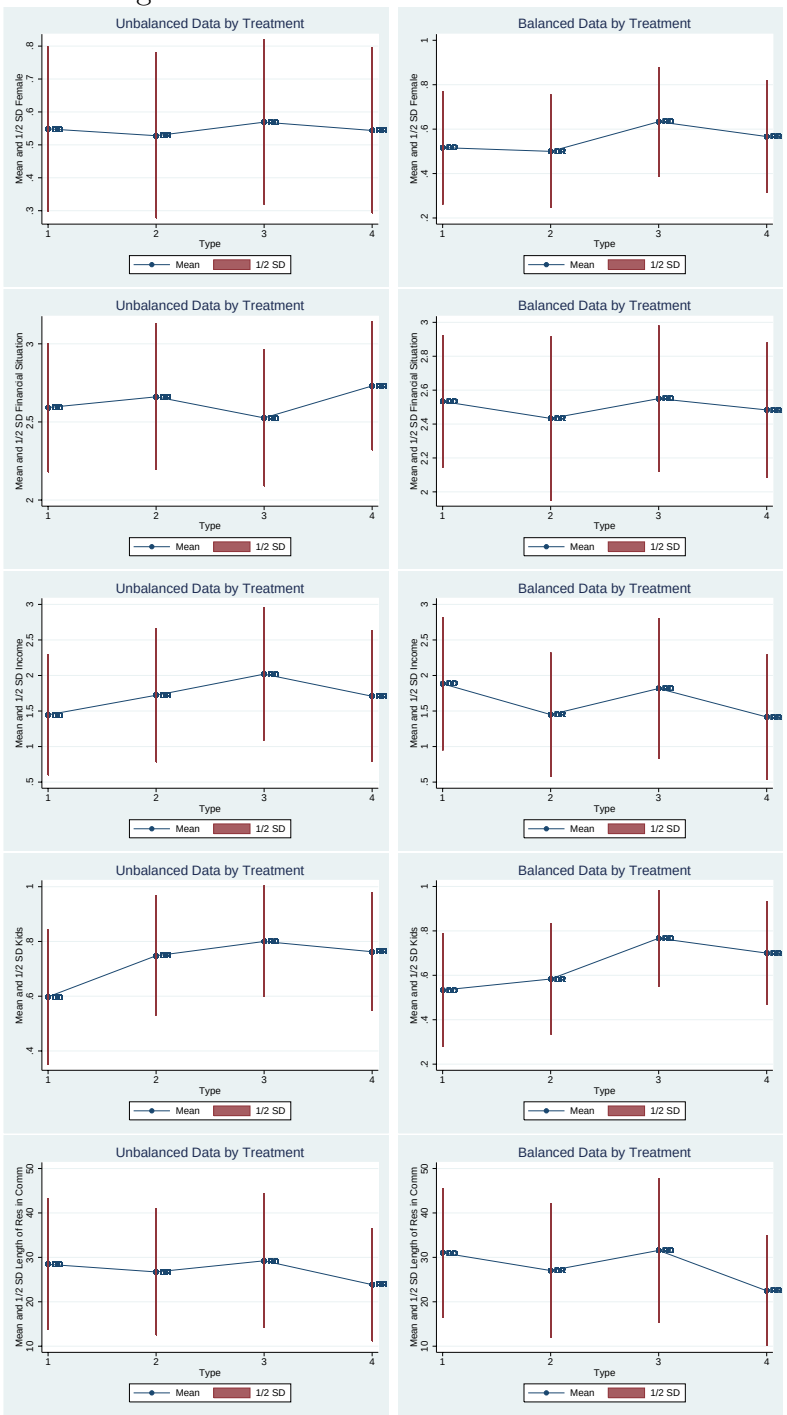
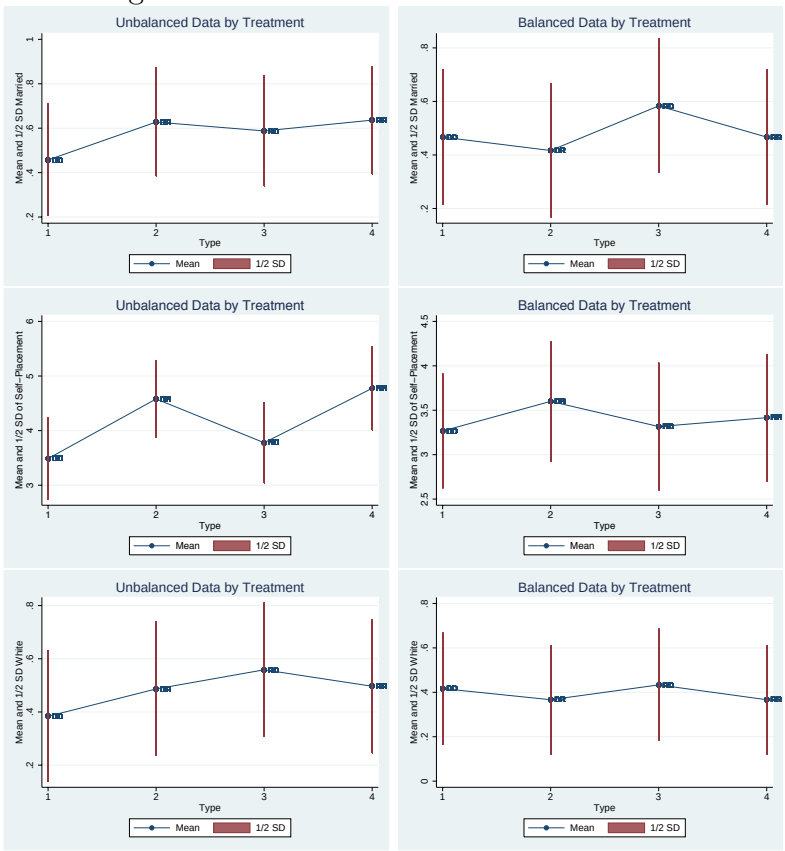


Figure 8.10: Unbalanced vs Balanced Data



Chapter 9

Peer-Pressured Party Identification

9.1 Summary

This chapter analyzes the effect of the partisanship of a social network on an individual's partisan identification. Using survey data, each respondent is asked to identify two discussants with whom she has political conversations. Controlling for the respondent's characteristics and the nature of the relationship with the discussant using propensity score matching, each respondent's own partisanship is influenced by the discussants' party identification choices. Respondents are also seen to be most influenced by discussants with whom they have frequent conversations and who are geographically proximate.

9.2 Introduction

People's choices are affected by many of their friends' and family's choices. A casual observer of human behavior would observe this phenomena amongst teenagers in their choice of slang, jeans, or cell phones — and recent economic field experiments have demonstrated that in fact people's purchasing choices, vaccination choices, and retirement plan decisions are correlated with those in their social networks (Mobius, Niehaus and Rosenblat 2005; Rao, Mobius and Rosenblat 2006). This chapter explores the possibility that there is a peer-pressure component to the adoption of party identification. Party identification appears to be the best predictor of an indi-

vidual's vote choice (Campbell et al. 1952), but there has been little research done to clarify the process of the adoption of a particular party identification. Some have speculated a genetic component (Alford, Funk, and Hibbing 2005) while others have argued that party identification is in fact inherited from parental influence or other childhood socialization (Jennings and Niemi 1968; Searing, Schwartz and Lind 1973; Iyengar 1976). This chapter argues that there is a social component to the choice of party identification and that, like the purchasing choices studied above, people's party identity is influenced by their social network.

In this study each individual describes their own partisanship and the partisanship of their political discussants. It is then possible to examine the correlation which exists between the two — which in the data-set used here is extremely high — and to posit a relationship which allows the discussants to influence the partisan identification choice of the respondent. The survey includes a series of questions to ascertain the details of the respondent's relationship with the discussants as well as to determine many of the respondent's characteristics. These variables are used to help control for the selection of a discussant — done using propensity score matching — which then allows for the discussant effect to be determined.

That voters often appear to choose candidates based upon their partisan identification, and that the voters' affiliations with particular partisan identifications appear to remain fairly static, makes predicting voter's party identification an important task for political scientists. Father's party identification and other social characteristics or policy preferences have been used to predict party identification with some success (Goldberg 1966; Cassel 1982; Knoke and Hout 1974; Jennings and Niemi 1974) with the father's party identification the best predictor. However, the explanation that an individual inherits her party identification is a fairly unsatisfactory explanation — in particular when thinking of an individual as a rational agent capable of forming political allegiances which align directly with her own preferences. It seems more likely that the individuals have a set of shared characteristics with their family, but that were the individual to form new, close ties, the strength of the father's party identification as a predictor would decrease. This is in fact observed in the Jennings-Niemi parent-

socialization study: The high-school seniors have decreasing rates of agreement with their father's party identification as time goes on. Furthermore, Abramowitz (1983) demonstrates that the partisan identification for a group of college students thrust into a liberal campus environment has the ability to change away from the father's party identification. While there do appear to be political changes which can trigger a shift in party identification (Fiorina 1981; Franklin and Jackson 1983; Franklin 1984), the role of the father's partisan identification continues to be positive and significant in any explanation of the adoption of party identification.

Political scientists have long been interested in party identification, not only because voters seem to be particularly attached to their party and reluctant to change, but also because party identification is a powerful predictor of vote choice (Campbell et al. 1960, Belknap and Campbell 1952, Keith et al. 1992, Miller and Shanks 1996, Stokes 1966, Converse 1966, Green, Palmquist, and Schickler 2002). In this literature, party identification is seen to be stable and often inherited. Franklin and Johnson (1983, 957) say "party identification was seen as developing in youth, largely as a result of one's association with parents' partisan preference, and was ubiquitous in its influence over the other aspects of electoral behavior." Campbell et al. argue that party identification is an inherently social characteristic which ties individuals to particular social groups, in particular families, and which voters are unlikely to change. Achen (1992) argues partisanship stability is still consistent with a model in which voters update their beliefs about party location. Partisanship is seen to be more stable than "beliefs about equal opportunity, limited government, and moral tolerance" (Goren 2005, 881). Green, Palmquist, and Schickler (2002, 74) report a correlation over 13 panel surveys of .97 for individual respondent party identifications. Across the 1974 and 1976 CPS American National Election Study Panel, Keith et al. (1986, p. 244) find party stability of 78%. Reported party stability is very high, even in data-sets which do not require the respondents to recall their previous party identifications (Niemi, Katz, and Newman 1980, Jennings and Markus 1984, Green and Palmquist 1990). There is a debate, however, about whether or not party identification is flexible or static, and if flexible, what factors which would provoke

a particular switch. Were party identification completely static and pre-determined, for example, and simultaneously such a good predictor of vote choice, then it seems unlikely there would ever be significant change in electoral outcomes.

Another set of political scientists have considered party identification to be a direct product of an individual's opinions about the current political framework (Downs 1957, Fiorina 1981, Green and Gerber 1998, Meier 1975). This argument states that the change in party identification can be explained by updates in information, so that "partisanship may undergo considerable change as events alter public confidence in each party or as changes in party leadership alter the perceptions of the parties' policy stances" (Green and Gerber 1998, 795). Many scholars have found that partisanship may indeed change (Brody and Rothenberg 1988, Franklin and Jackson 1983, Jackson 1975, Page and Jones 1979, Markus and Converse 1979, Dobson and Meeter 1974, Dobson and St Angelo 1975, Dryer 1973). Franklin (1984) finds that an individual's issue preferences are indeed components of partisanship, even when controlling for parental party identification, which suggests that individual partisanship is not completely inflexible.

The debate of flexible or static partisanship plays a role in understanding the very principles which guide voting behavior. If indeed individuals are making candidate choices based upon partisanship then the process of adoption of partisanship is crucial in determining the relationship between individual's political values, political preferences and candidate choice as well as other political behaviors (Alvarez and Brehm 2002, Feldman 1988, Martinez and Gant 1990). Underlying each political behavior is an assumption about the voter's decision-making process. This chapter will maintain that one component of this process heavily relies upon with what political party others in the individual's social group identify. Dating back to Campbell et al. (1960), political scientists have known that an individual's social group was influential in a voter's decision-making process. Campbell et al. says about partisanship change, that it was attributable to change such as "a marriage, a new job, or a change in neighborhood" (150). In this data-set we observe reported social context and infer the level of influence the social context will have on partisanship.

One solution which has the potential to explain the deviations in partisanship from the father's partisan identification is that the individual's social network has changed. Individuals are observed to have correlations in their voting and political preferences with their political discussants across surveys (Huckfeldt and Sprague 2004). Furthermore, Mutz (2006) has documented that in fact partisan identification is less likely to differ across individuals who are in close relationships, such as family members. Mutz documents patterns of deliberation and finds that disagreement is most sustainable in relationships which are described by sociologists as "weak ties" — relationships which do not require high levels of regular interaction. Thus there is sufficient likelihood to hypothesize that socialization is still a reasonable explanation of partisan identification but that simply using the father's partisan identification is not sufficient, as an individual's social network will change over time.

As noted by economist Manski (1993), drawing causal inferences while using data from an individual's social network is particularly difficult because it is possible to observe what appears to be influence from discussion partners but is in fact influence from a common and shared environment (for example: observing a group of students, who are all friends, doing better in school and then realizing that they share a teacher) or instead is in fact influence from a sorting procedure (for example: observing a group of students, who are all friends, doing better in school and then realizing that they became friends because they were all interested in reading). The problem is further complicated by the fact that it is likely that each network member exerts influence on each other. The data in this study is drawn to mirror a national probability sample and thus the individuals are unlikely to be exposed to a common environment. However, selection of discussants and the direction of influence are both of concern in drawing conclusions from these results.

This data-set is particularly problematic because is difficult to draw causal inferences in observational data with potential selection biases. Thus the results in this paper rely upon several assumptions. The first of such assumptions is that the problem of selection into a particular discussion network is controlled for by incorporating a series of characteristics about the respondent and the nature of the relationship

with the discussants. These include a series of covariates about the respondent which are likely to be highly correlated with partisan discussion selection — race, age, income, educational status, political ideology, marital status, and gender — and serve as proxies for the variables which determined why each respondent selected particular discussants. Furthermore, variables which describe the nature of the relationship with the discussant and the frequency of communication should also alleviate the selection problem. The second of such assumptions regards the direction of influence. The statistical model is set up such that the discussants will influence the respondent's choice of party identification, and not vice versa. This assumption is supported by the findings of Huckfeldt and Sprague (2004), who document extremely low levels of reciprocity for discussants naming the respondent as an individual with whom they discuss political matters in their 1996 Indianapolis-St. Louis Study. The nature of political discourse is such that it is likely asymmetrical with some individuals who have a particular taste for accumulating political information or particularly strong preferences for communicating political information to others. The final assumption is that the respondents are able to accurately discern the partisanship of the discussants. Studies of the ability of voters to determine this have found that there is a bias for the respondent to over-report agreement with the discussant. In Huckfeldt and Sprague's 1996 Indianapolis-St. Louis Study they find that the respondents are 80% accurate when asked about the partisanship of their political discussants and that the respondents are more likely to be accurate if they agree with the discussant's true political preferences (Huckfelt and Sprague 2004).

The chapter proceeds with a detailed discussion of the data available, including a summary of each respondent's characteristics. The following section describes the methods used to analyze the results of social disagreement on an individual's choice of party identification, with includes describing the consequences of applying a matching algorithm to select observations which are essentially identical in terms of covariates. Standard parametric procedures are applied to both the matched and unmatched data-sets, and the large effect of discussant party identification supports Campbell's original hypothesis — that social ties play a role in determining partisanship. This

links back in to the initial theoretical model, which demonstrated that individuals who were socially connected would make correlated political choices after sharing information.

9.3 Data

The data for this study were collected as part of an internet panel which was in the field shortly before the congressional midterm elections in November 2006.[1] Each respondent was asked a battery of questions designed to provide a standard set of covariates and then a set of questions particular to the study of peer influence on politics. One advantage of this data-set is that it is unlikely that the respondents were at all influenced by interviewer effects, as they were able to complete the survey in the privacy of their own home. One disadvantage of the survey, however, is that despite the fact that their covariates were matched to a randomly drawn survey population, it is still possible that the internet-user population is significantly different in some unobserved way than the average survey respondent. In this study in particular, it seems possible that respondents who are internet users would have the opportunity to engage discussants who are not geographically proximate. However, given that covariates such as income, age, and education have been matched directly, this chapter assumes that the responses to the network battery would not be influenced by the fact that this is an internet population.

Each survey respondent was asked to generate up to two names of individuals with whom the respondent discusses politics.[2] Of the 1000 individuals, 155 were unable or unwilling to name any discussants whatsoever and they were excluded from further analysis. Of the individuals who could name discussants, 64 individuals were also excluded as they were either unable to identify their own party identification (12

[1]Polimetrix surveyed approximately 133,000 individuals from the last week of September 2006 through the first Tuesday in November and provided a random sample of 1000 individuals who answered this particular battery, matched by covariates to the 2004 American Community Survey (ACS). Although the survey population by itself would likely be different than a random sample, as internet users tend to have higher income and education levels, the comparison with the 2004 ACS via one-to-one matching does provide the equivalent of a random sample.

[2]See the final section in this chapter for detailed question wording.

individuals) or were unable to identify the party identification of their discussants (53 individuals). For the remaining 781 individuals in the sample it is possible to observe both their own party identification as well as that of their discussion partners.

Each respondent is asked to classify their party identification on a seven-point scale ranging from "Strongly Democratic" to "Strongly Republican". When asked to describe the partisanship of their discussants, each respondent may choose Democratic, Republican, Independent, Other or Don't Know. The tabulated responses in Table 9.1 categorize each respondent into their stated party identification and tabulate the party identification of those respondents' discussants. These categories demonstrate that each respondent is more likely to have discussants who agree with their own party identification. Of the individuals who identified as "Strongly Democratic", 125 of their discussants (68%) were also classified as Democratic, whereas for the individuals who identified as "Strongly Republican", only 39 of their discussants (18%) were identified as Democratic. Based upon each respondent's description of the partisanship of their discussants, each respondent is assigned a binary partisan discussion treatment variable. This variable takes on a value of 1 if the respondent has at least one Democratic discussion partner and is 0 otherwise.[3] Of the 781 respondents, 420 have at least one Democratic discussion partner (54%) while the other 361 do not (46%).

Table 9.1 Goes Here

The central question this chapter will address is whether or not the party identification of the political discussants will influence the party identification of the survey respondents. Each respondent, however, is likely to identify with the particular party based upon a number of personal characteristics which are included in later analysis as control variables. These variables include the respondent's ideological self-placement (on a 0 to 100 scale), income, marital status, gender, age, education level, race, and

[3]This classification by definition treats partisanship unequally, weighing more heavily Democratic discussion partners (suppose that an individual had one Republican discussion partner and 1 Democratic discussion partner, for example). Future research will consider multiple classifications for partisan discussion treatments.

homeownership. Also, each respondent is likely to have established a relationship with the discussants due to a number of shared characteristics. The particular content of this survey allows for many of these characteristics to be included — the geographic distance between discussant and respondent, the frequency and method of conversation, the nature of the relationship (i.e.,friend, family, etc.), and whether or not the discussant is likely to vote in national elections.

Table 9.2 Goes Here

Each of the variables is described in Table 9.2. The average respondent is Independent, has some college, is 50 years old and is in a fairly high income bracket. Most of the individuals in this data are home owners (74%), 63% are married, and 14% of the respondents are nonwhite. Most of the respondents have discussants with whom they are geographically proximate, although 36% of the discussants do not live in the same city as the respondent. This statistic alone has implications for the need for additional survey research on the social networks of respondents as it is clear that with 37% of the discussants living outside of the respondent's city that geographic proxies for social relationships are insufficient. The primary method of communication with discussants was in person (66%), followed by phone (22%), and 56% of the discussants spoke to the respondent daily. Over 25% of the discussants shared a home with the respondent and the majority of the relationships between the discussant and respondent could be classified as family or friend, as seen in Table 9.3.

Table 9.3 Goes Here

This data-set provides a unique opportunity to identify the effect of the discussants' partisanship on the respondent because it incorporates much of the information about the relationship between the respondent and discussant while allowing for the respondent's covariates to serve as control variables in the choice of party identification.

9.4 Method

In order to determine whether or not each respondent is affected by the discussants' partisanship, it is necessary to include a series of control variables along with the partisanship of the respondent and the discussant. Individual respondents will have selected their discussants because of the types of their relationships — for example, it is possible to choose to have political discussions with friends but not necessarily with family members. Furthermore, respondents will undoubtedly have selected discussants who are similar in terms of their shared socioeconomic characteristics — income, age, race, education, etc. Controlling for the respondent's characteristics and the relationship characteristics with the discussant will help to eliminate any biases which would be introduced by respondents choosing discussants with similar characteristics or having discussions within particular types of relationships. Thus all covariates described in the previous section will be included in the analyses.

Several of the covariates have missing observations, in particular the income category.[4] These missing values are imputed using the R package *Amelia* (Honaker et al. 2001) and five completed data-sets are produced.[5] Each of the data-sets will be analyzed independently and the final statistics will be computed after averaging across all five data-sets.

Suppose the following experiment were to occur — a randomly selected group of individuals were assigned a "treatment" of either a Democratic discussant or a non-Democratic discussant. If in fact this experiment had occurred in this data-set, it would then be possible to observe the effect of the Democratic discussant "treatment" on party identification. However, the data-set in this study is not experimental —

[4]The following variables are incomplete from the set of 781 observations: ideological self-placement (747 total), income (631 total), education level (780 total), geographic distance of first discussant (765 total), geographic distance of second discussant (737 total), the frequency of conversation with the first discussant (769 total), the frequency of conversation with the second discussant (738 total), the method of communication with the first discussant (768 total), the method of communication with the second discussant (738 total), and the frequency with which the discussants participate in national elections (743 total).

[5]The missing variables are likely missing at random (MAR) — they are variables which are correlated with other observations in the data-set — and thus *Amelia* is appropriate. Producing multiple data-sets helps to eliminate any bias which would be introduced as a consequence of the imputation (King et al. 2001).

respondents have selected their discussants, and, consequently, the "treatment" of Democratic discussant is not random in this data-set. Assuming that it would be possible to predict the type of discussant that the respondent would select based upon the respondent's characteristics and the nature of the relationship with the discussant, however, it is the possible to control for the discussant selection.

Because the quantity of interest here is to examine the effect of discussant partisanship on the respondent's party identification, a subset of observations is drawn from each treatment group (Democratic discussants or non-Democratic discussants) who, in terms of their characteristics and relationships with their discussants, are practically identical. Essentially the goal is to find two individuals who both, for example, are married, have some college education, are white, have moderate income, named two family members as discussants, and who are moderately liberal, but who have different treatments — that is, one respondent who has non-Democratic discussants and the other respondent with Democratic discussants. It is then possible to compare the effect of Democratic discussants on the respondents' party identification because the only variable which is different between the two respondents is in fact the treatment variable. The process of drawing a subset of respondents from the data whose covariates are identical across the two treatments is called matching. The matching process on this data-set uses all the covariates and runs on each imputed data-set separately using the program MATCHIT (Ho et al. 2007). Using treatment as the dependent variable and all covariates as independent variables, a propensity score is estimated for each observation using a logistic model. Based upon that propensity score, a genetic-matching algorithm is used to determine the optimal number of observations to be included in each matched data-set (Diamond and Sekhon 2004).

The quality of the match is presented in Table 9.6. Here the percent improvement for each data-set, by variable, is presented. The first row demonstrates the percent overall improvement in covariates from matching for each data-set and these values are all between 85% and 93% — the covariates are, on average, looking more identical for the two types of discussant groups. Many of the covariates themselves, however, show

enormous percentage decreases in improvement, in particular the geographic distance from the first discussant, if the discussants are group members, and the principle method of communication with the first discussant. Because not every covariate demonstrates improvement in distance across treatment groups, the covariates will still be included as control variables and the matching will be used as a preprocessing technique for the data (Ho et al. 2007).

Table 9.6 Goes Here

To determine the average treatment effect of being exposed to Democratic discussants, the matched data is combined together into a single data-set and fit using both a linear model as well as an ordered logit.[6] The linear model does not capture the categorical nature of the respondent's party identification that the ordered logit will — however, the party identification is fit to a seven-point-scale (with 1 being strong Democrat and 7 being strong Republican) and the advantage of the linear model is to permit the incorporation of all the data into a single estimate of the mean effect of a Democratic discussant. However, the ordered logit is in fact more appropriate, as the respondent's party identification is in fact not a continuous variable. The seven-point party identification variable is used instead of the three-point party identification variable because not only were more respondents able to categorize themselves using the seven-point variable, but also because it allows for greater precision in understanding an individual voter's party identification choice. Analysis computed using the 3-point party identification (Democratic, Republican, and Independent), however, produces similar results.

9.5 Results

The effect of having a Democratic discussant is presented in Table 9.4. Here the dependent variable is the seven-point party identification measurement and the quantity

[6]The standard errors are adjusted appropriately as the models are fit using Zelig (Imai et al.).

of interest is the coefficient on the partisan discussant treatment variable.[7] The covariates are included as controls in the analysis.[8] In the linear model, the first differences are presented using the covariates at their means and comparing the expected value of having a Democratic discussant against the expected value of having a non-Democratic discussant. This produces an effect of almost one point and is statistically significant at traditional values. The difference in probabilities from the ordered logit supports this finding as well. Here the probability of falling into a particular partisan category is computed conditional upon having Democratic discussants or not while holding all covariates at their means. If there is an effect of discussants then the pattern of probabilities should appear as it does in Table 9.4, with the differences in probabilities for the Democratic identification positive and the Republican identification negative. If there is a discussant effect then the probability of being a Democrat conditional upon having Democratic discussants should go up, while the probability of being a Democratic conditional upon having non-Democratic discussants should go down. Thus the difference between these two should be positive. Similarly, if there is a discussant effect then the probability of being a strong Republican conditional upon having Democratic discussants should go down while the probability of being a strong Republican conditional upon having non-Democratic discussants should go up. Thus the difference between these two should be negative. Interestingly the magnitude of the differences also change appropriately throughout the table.

Table 9.4 Goes Here

The interpretation of these results as causal relies upon a series of assumptions discussed in the introduction. The results appear to be of such high magnitude that it seems possible that part of the selection problem has not been solved by inclusion of covariates or that a respondent will over-report their discussant's partisanship in line with their own. However, the results are of such large magnitude that these

[7]The equivalent result for OLS using the three-point party identification measurement would have a mean of -.337, a standard deviation of .059, and an N of 2934. Equivalent results for Democratic, Independent, and Republican identifications for the ordered logit would have means of .177 (sd .09), -.02 (sd .1551), and -.155 (.09), respectively.

[8]Coefficients for the controls are included in the final section of this chapter.

explanations are insufficient to account for such a large effect of discussant party identification, at least at the rate of over-reporting which is measured in the existing literature (Huckfeldt and Sprague 2004). There does appear to be an effect of discussant party identification on the respondent's choice. This effect is the marginal effect of discussant party identification alone, as many of the respondent's covariates are included in the analysis, including variables which describe the nature of the respondent's relationship with the discussant.

To document disagreement in terms of partisan identification and to get a better intuition as to what sorts of relationships sustain it, a variable is created which calculates the average of the discussant party identifications in a three-point party identification scale. This variable ranges from 1 to 3 with an average of 1.98. A variable to document the amount of disagreement in the relationship is then calculated by determining the distance between the respondent's three-point party identification and that of the average discussant. This variable ranges from 0 to 2 and has an average of .45.

It is then possible to simply look at the correlations between the disagreement variable and several of the other variables in the data-set. Correlation between disagreement and the average geographic distance between discussants and respondent demonstrates that these variables are very slightly positively correlated — there is more disagreement in relationships which have a greater geographic distance. The intuition behind this correlation might be that respondents who have more regular conversations with discussants are less likely to disagree, as respondents speak more frequently to respondents who are more geographically proximate, and both of these intuitions are supported by the data. Conversations occur most frequently with discussants who are geographically proximate and furthermore, discussants who agree with the respondent have more frequent conversations.

These correlations have implications for the interpretations of the results. If the discussants' partisan identifications have real effects on the respondent's choices, then one reasonable interpretation might be that the more frequent the interaction, the more likely the agreement.

9.6 Conclusions

The results here have demonstrated the peer-pressure component in the adoption of party identification. The existing literature has drawn the conclusion that there are influences from within families (either genetic or parental influence). This paper does not speak to the possibility that party identification is genetically inherited, but does demonstrate that party identification is influenced by the respondent's social connections — thus the previous parental studies may now be extended to larger social units of influence. The magnitude of the results here is quite large and suggests that peer influence plays an enormous role in the respondent's choice of party identification.

These findings speak to the literature on political disagreement within social networks directly. Mutz (2006) and Huckfeldt, Johnson, and Sprague (2004) have documented correlations in party identification and candidate choice within voter's social networks. Mutz (2006) suggests that voters should be encouraged to sustain more "weak ties" as those are the relationships within which political disagreement can occur. These findings support the conclusion that in fact there will be more disagreement within relationships that have less frequent conversation and more geographic distance, but suggests instead that it is in fact the frequency of those conversations which indeed plays a role in influencing agreement in party identification.

This in turn suggests that there is a peer-pressure component to party identification which could derive from a rational choice framework. Suppose, in fact, that voters update their information about which party best represents their own personal preferences from communication within their social network. If this is indeed the case, then the combination of finding influence from discussants and finding that more frequent discussants have greater influence give additional credence to this possibility. Another possibility to explain the correlations in behavior, however, is that voters simply want to behave like those around them because they find political disagreement distasteful. The empirical results demonstrated in this paper cannot discern between these two potential hypotheses. However, the presence of correlation in party identification, controlling for the respondent's characteristics and the nature

of the discussant relationship, improve the understanding of how social groups play a role in determining an individual's political choices.

9.7 Future Research

Future research for this particular data-set includes several different options for modeling and for methodological techniques. One alternative would be to use the respondent's ideological self-placement as the outcome variable and to change the treatment variable to whether or not the respondent's party identification agreed with that of the discussant. One concern with the method above is that it is possible that the respondent's statements regarding the discussant are to a large degree endogenous — using the respondent's ideological placement as opposed to party identification might help alleviate this concern.

An option for additional methodological sophistication would be to use the multi-valued treatment algorithm described in the previous chapter to account for the variety of party identifications present amongst the discussants — or at least to account for whether or not the respondent has one or two Democratic discussants. This would enable the intensity of the disagreement to be documented in the analysis.

9.8 Appendix

1. From time to time, people discuss government, elections and politics with other people. I'd like to know two people you talk with about these matters. These people might or might not be relatives. Just tell me their first names or initials.[9]

2. There are many ways in which people are connected to each other. Some people can be connected to you in more than one way. For these relationships, please indicate

[9]This name-generator is similar to that used by Huckfelt and Sprague in their St. Louis-Indianapolis survey (1996) and was later applied to the 2000 American National Elections Study. Huckfeldt and Sprague demonstrated that there was no difference in responses between this question and an equivalent question which asked people about "important" matters. The questions which follow in this survey are unique.

all the ways (NAME1) is connected to you on this list:

Family member,

Coworker,

Member of a group to which you belong,

Neighbor,

Friend,

Professional advisor or consultant,

E-contact via the internet,

Other.

3. There are many ways in which people are connected to each other. Some people can be connected to you in more than one way. For these relationships, please indicate all the ways (NAME2) is connected to you on this list:

Family member,

Coworker,

Member of a group to which you belong,

Neighbor,

Friend,

Professional advisor or consultant,

E-contact via the internet,

Other.

4. Sometimes the people we know live close to us and sometimes they live far away. Please tell me where (NAME1) lives compared to where you live.

In your home,

In your neighborhood,

In your city,

In your state,

In a different state,

In a different country.

5. Sometimes the people we know live close to us and sometimes they live far away. Please tell me where (NAME2) lives compared to where you live.
In your home,
In your neighborhood,
In your city,
In your state,
In a different state,
In a different country.

6. There are many ways to communicate with people. Most of the time, is the predominant way in which you communicate with (NAME1) by:
Phone,
Email,
In person,
By instant-message,
By mail.

7. There are many ways to communicate with people. Most of the time, is the predominant way in which you communicate with (NAME2) by:
Phone,
Email,
In person,
By instant-message,
By mail.

8. Thinking about how often you usually talk to (NAME1), on average, do you talk to (him/her):

Almost every day,
At least once a week,
At least once a month,
Less than once a month?

9. Thinking about how often you usually talk to (NAME2), on average, do you talk to (him/her):
Almost every day,
At least once a week,
At least once a month,
Less than once a month?

10. In general, would you say (NAME1) is a:
Democrat,
Republican,
Independent,
Something else.

11. In general, would you say (NAME2) is a:
Democrat,
Republican,
Independent,
Something else.

12. As far as you know, would you say that (NAME1) is a:
Strong (Democrat/Republican/Independent),
Moderate (Democrat/Republican/Independent),
Weak (Democrat/Republican/Independent)?

13. As far as you know, would you say that (NAME2) is a:

Strong (Democrat/Republican/Independent),

Moderate (Democrat/Republican/Independent),

Weak (Democrat/Republican/Independent)?

14. As far as you know, does (NAME1) vote in national elections:

Almost always,

Sometimes,

Almost never,

Never.

15. As far as you know, does (NAME2) vote in national elections:

Almost always,

Sometimes,

Almost never,

Never.

9.9 Tables

Table 9.1: Respondent Party Identification and Discussant Party Identification

Respondent's Party Identification	Discussant Democratic	Discussant Independent	Discussant Republican	Discussant Other	Discussant Don't Know
Strongly Democratic	67.57%	15.68%	11.89%	2.16%	2.7%
Weakly Democratic	48.25%	21.68%	21.68%	2.1%	6.3%
Lean Democratic	41.23%	35.55%	16.59%	1.9%	4.74%
Independent	27.87%	36.01%	20.49%	4.1%	11.48%
Lean Republican	22.38%	24.48%	41.26%	3.5%	8.39%
Weakly Republican	27.42%	8.87%	58.06%	1.61%	4.03%
Strongly Republican	17.89%	11.01%	66.51%	1.38%	3.21%

Row percentages indicate the fraction of partisan discussants reported by each of the seven-point partisan categories.

Table 9.2: Summary of Variables

Variable	Mean	Min	Max
Party Identification	4.03	1	7
Democratic Discussant	.54	0	1
Home Owner	.74	0	1
Female	.54	0	1
Nonwhite	.14	0	1
Education Level	3.52	1	6
Marital Status	.63	0	1
Age	50.79	18	106
Income Bracket	8.24	1	13
Ideological Self-placement	54.62	0	100
Discussant — Family	.92	0	2
Discussant — Coworker	.36	0	2
Discussant — Group member	.22	0	2
Discussant — Neighbor	.16	0	2
Discussant — Friend	.84	0	2
Discussant — Advisor	.05	0	2
Discussant — Internet	.16	0	2
Discussant — Other	.08	0	2
Geographic Distance, First Discussant	1.58	0	5
Geographic Distance, Second Discussant	2.17	0	5
Frequency of Conversation, First Discuss	3.51	1	4
Frequency of Conversation, Second Discuss	3.31	1	4
Method of Communication, First Discuss	3.23	1	5
Method of Communication, Second Discuss	3.01	1	5
Frequency of Discussants' Voting	7.6	2	8

Table 9.3: Respondent Relationship with Discussants

Relationship	Number of Discussants	Percentage
Family	734	33.01%
Friend	672	30.28%
Coworker	281	12.66%
Group Member	171	7.71%
Neighbor	126	5.68%
Internet Contact	129	5.81%
Advisor	40	1.8%
Other	66	2.97%
Total	2219	100%

Note here that it is possible for an individual to classify a discussant into more than one relationship category, and thus the column percentages indicate the percentage of relationships which fall into that category out of all the relationships listed, not of total discussants.

Table 9.4: Treatment Effects

OLS Quantity	Mean	Std. Dev.
E[Party | Dem Discuss] - E[Party | No Dem Discuss]	-.9172	.1471
Ordered Logit Quantity	Mean	Std. Dev.
P(Strong Dem | Dem Discuss) - P(Strong Dem | No Dem Discuss)	.122	.092
P(Weak Dem | Dem Discuss) - P(Weak Dem | No Dem Discuss)	.038	.056
P(Lean Dem | Dem Discuss) - P(Lean Dem | No Dem Discuss)	.016	.089
P(Indep | Dem Discuss) - P(Indep | No Dem Discuss)	-.012	.05
P(Lean Rep | Dem Discuss) - P(Lean Rep | No Dem Discuss)	-.026	.053
P(Weak Rep | Dem Discuss) - P(Weak Rep | No Dem Discuss)	-.037	.048
P(Strong Rep | Dem Discuss) - P(Strong Rep | No Dem Discuss)	-.1	.09
N	2968	

These estimates are calculated after pooling all five imputed and matched datasets. Only the coefficient calculated with OLS is statistically significant at traditional levels.

9.10 Additional Tables

Table 9.5: Respondent Party Identification and Discussant Party Identification Raw Numbers

Respondent's Party Identification	Discussant Democratic	Discussant Independent	Discussant Republican	Discussant Other	Discussant Don't Know	Total
Strongly Democratic	125	29	22	4	5	185
Weakly Democratic	69	31	31	3	9	143
Lean Democratic	87	75	35	4	10	211
Independent	34	44	25	5	14	122
Lean Republican	32	35	59	5	12	143
Weakly Republican	34	11	72	2	5	124
Strongly Republican	39	24	145	3	7	218
Total	420	249	389	26	62	1146

Table 9.6: Percent Improvement in Covariate Means From Matching

Variable	Data1	Data 2	Data 3	Data 4	Data 5
Overall Improvement	85.76	89.61	88.27	93.37	89.53
Home Owner	66.71	91.12	37.87	80.03	66.72
Female	-67.46	28.23	40.19	28.23	-55.5
Nonwhite	84.59	96.15	88.44	88.44	76.88
Education Level	19.66	47.69	-6.62	18.13	-8.36
Marital Status	83.79	88.65	83.79	95.14	98.38
Age	-52.13	-26.38	-76.81	36.27	54.08
Income Bracket	-10.38	73.49	11.77	-46.03	-425.87
Ideological Self-placement	85.8	86.59	85.6	91.06	86.36
Family Member	96.21	88.64	96.21	84.86	90.54
Coworker	87.35	57.83	45.18	49.4	11.45
Group Member	-159.04	-19.56	-72.7	-59.41	-52.77
Neighbor	4.24	55.83	23.4	29.78	87.233
Friend	80.89	61.78	72.21	68.73	51.34
Advisor	47.01	-45.73	47.01	60.26	73.5
Internet Friend	-299.45	-299.45	-336.9	-249.52	-411.79
Other Relationship	63.56	91.90	79.76	95.95	75.71
Geo Distance, Discuss 1	-308.04	-396.55	-1642.58	-887.28	-2279.733
Geo Distance, Discuss 2	91.74	-13.21	45.01	.4	-78.92
Frequency of Communication, Discuss 1	-55.04	98.37	7.16	-71.19	-56.67
Frequency of Communication, Discuss 2	3.39	-69.24	34.53	71.94	-69.01
Voting Propensity, Both Discussants	-12.57	29.63	91.67	49.33	31.51
Method of Communication, Discuss 1	-72.83	-101.17	-58.6	19.62	-269.74
Method of Communication, Discuss 2	58.8	90.24	87.86	84.55	81.37
Matched Sample Size	599	595	588	590	596

Each entry represents the percentage change in the difference between the treated and untreated data before and after matching. Data is matched within each imputed dataset, so that the columns represent change for each dataset.

Chapter 10

Dynamic Party Identification: A Social Network Explanation

10.1 Summary

This chapter uses the 2006 American National Election Pilot Study (ANES) to determine whether the discussant's partisan identification plays a role in determining the respondent's choice of party identification. Each respondent is asked twice to identify their party — once in November 2004 and again in November 2006. In the second instance each respondent is also asked to identify a group of political discussants and to describe the nature of those relationships. Controlling for the respondent's characteristics and the discussant relationships, there is an increased probability that each respondent will change party identification towards agreement with the discussants' party identification. These findings have implications for the study of party identification, with particular emphasis on the adoption of new party identifications.

10.2 Introduction

Party identification is a powerful predictor of vote choice, and the adoption of party identification is a process which is influenced by an individual's social network. This paper demonstrates that individuals who change their party identification do so as a consequence of political disagreement within their social group, and furthermore that the party to which they are most likely to change is one which is consistent with that

of their social context. This result contributes to the literature on the stability and adoption of party identification, and helps tie party identification as a component of social group membership. Individual party identification does change, but in this setting it changes not only due to changes in the political arena but also due to social pressure.

Political scientists have long been interested in party identification, not only because voters seem to be particularly attached to their party and reluctant to change, but also because party identification is a powerful predictor of vote choice (Campbell et al. 1960, Belknap and Campbell 1952, Keith et al. 1992, Miller and Shanks 1996, Stokes 1966, Converse 1966, Green, Palmquist, and Schickler 2002). In this literature, party identification is seen to be stable and often inherited. Franklin and Johnson (1983, 957) say "party identification was seen as developing in youth, largely as a result of one's association with parents' partisan preference, and was ubiquitous in its influence over the other aspects of electoral behavior." Campbell et al. argue that party identification is an inherently social characteristic which ties individuals to particular social groups, in particular families, and which voters are unlikely to change. Achen (1992) argues partisanship stability is still consistent with a model in which voters update their beliefs about party location. Partisanship is seen to be more stable than "beliefs about equal opportunity, limited government, and moral tolerance" (Goren 2005, 881). Green, Palmquist, and Schickler (2002, 74) report a correlation over 13 panel surveys of .97 for individual respondent party identifications. Across the 1974 and 1976 CPS American National Election Study Panel, Keith et al. (1986, p. 244) find party stability of 78%. Reported party stability is very high, even in data-sets which do not require the respondents to recall their previous party identifications (Niemi, Katz, and Newman 1980, Jennings and Markus 1984, Green and Palmquist 1990). There is a debate, however, about whether or not party identification is flexible or static, and if flexible, what factors which would provoke a particular switch. Were party identification completely static and pre-determined, for example, and simultaneously such a good predictor of vote choice, then it seems unlikely there would ever be significant change in electoral outcomes.

Another set of political scientists have considered party identification to be a direct product of an individual's opinions about the current political framework (Downs 1957, Fiorina 1981, Green and Gerber 1998, Meier 1975). This argument states that the change in party identification can be explained by updates in information, so that "partisanship may undergo considerable change as events alter public confidence in each party or as changes in party leadership alter the perceptions of the parties' policy stances" (Green and Gerber 1998, 795). Many scholars have found that partisanship may indeed change (Brody and Rothenberg 1988, Franklin and Jackson 1983, Jackson 1975, Page and Jones 1979, Markus and Converse 1979, Dobson and Meeter 1974, Dobson and St Angelo 1975, Dryer 1973). Franklin (1984) finds that an individual's issue preferences are indeed components of partisanship, even when controlling for parental party identification, which suggests that individual partisanship is not completely inflexible.

The debate of flexible or static partisanship plays a role in understanding the very principles which guide voting behavior. If indeed individuals are making candidate choices based upon partisanship then the process of adoption of partisanship is crucial in determining the relationship between individual's political values, political preferences and candidate choice as well as other political behaviors (Alvarez and Brehm 2002, Feldman 1988, Martinez and Gant 1990). Underlying each political behavior is an assumption about the voter's decision-making process. This paper will maintain that one component of this process heavily relies upon with what political party others in the individual's social group identify. Dating back to Campbell et al. (1960), political scientists have known that an individual's social group was influential in a voter's decision-making process. Campbell et al. says, about partisanship change, that it was attributable to change such as "a marriage, a new job, or a change in neighborhood" (150). In this data-set we observe reported social context and infer the level of influence the social context will have on partisanship.

This chapter contributes to this literature by analyzing party identification change on a new data-set which incorporates group pressures. Next the data-set is presented in detail, with emphasis on describing the characteristics of the partisan discussants.

The results of the consequences of political disagreement within a discussion network are presented in the section which follows, which incorporates the use of a matching technique to provide a balanced sample. The results are consistent for both unbalanced and balanced, which implies that conditional upon the assumptions that the respondent is able to accurately report her party identification and that of her discussion network, there is a large effect of political disagreement on the stability of party identification and the direction of change.

10.3 Data

The data in this paper were generated by face-to-face interviews as part of the American National Election Study Pilot Study (ANES) in the fall of 2006. A total of 665 individuals were interviewed — each of whom had been interviewed in the fall of 2004. Each respondent was asked a series of questions to identify their own personal characteristics during the 2004 survey. During the pilot survey in 2006, each respondent was asked to describe some of the characteristics of their discussion networks and to provide additional details about the relationship with each discussant. This data is of particular use for two reasons, both of which this paper uses to its advantage. First, this is a rare opportunity to glean insight into the characteristics of respondent-discussant relationships. Further knowledge of the respondent's discussion network is key to understanding how the discussants might influence the respondent's political behavior and these variables will be used as control variables in order to determine general discussant effects.

Second, the split timing of this survey provides a rare opportunity to explain change in party identification. Each respondent is asked for her party identification in both the 2004 and 2006 survey instruments. Thus it is possible to produce an indicator variable which is a 1 if the respondent's answer has changed from 2004 to 2006 and 0 otherwise. Amongst the group of respondents who answered both party identification questions (661), a total of 198 individuals reported a different party identification in 2006 then they had reported in 2004. The first component of

the analysis will deal with the factors which may influence a respondent's change in party identification. A quick glance at Table 10.1 describes the party identification switch for those individuals who changed party from 2004 to 2006. Most of the individuals who switched parties initially indicated that they were Independent in 2004 — 1 of these individuals reported Democratic party identification in 2006, 28 reported Republican party identification in 2006, 10 reported no preference in party identification in 2006, and 8 reported having other party identification in 2006.

Table 10.1 Goes Here

Each of the variables included in the analysis are incorporated in Table 10.2. The table is broken down into two categories — the characteristics of the respondent and the average characteristics of the discussion network. Each respondent names approximately two discussants and almost a third of the respondents report a different party identification. The average respondent is female, in a middle income bracket, is almost fifty years old, is married, has some college education, is employed, is white, and is a home owner. Only 20% of the respondents report being parents. The average respondent is ideologically located in the middle of a 0–10 scale and has some interest in politics.

Discussants are perceived by the respondent to have more interest in politics and the average discussion network reports a moderate amount of disagreement and a low level of relationship intensity. The average discussion network communicates with the respondent 50 days out of the year and is identified as having many strong partisans. Interestingly the average discussant is reported to be 320 minutes away from the respondent — this fact alone has implications for using geographic statistics as proxies for network measurements — and approximately one pair of the discussants communicate regularly with each other.

Table 10.2 Goes Here

Two other variables are produced to evaluate the amount of disagreement to which the respondent has been exposed. This disagreement variable is calculated by

dividing the total number of discussants whose partisanship disagrees with that of the respondent in November 2004 by the total number of discussants provided by the respondent. Thus as the respondent's political discussion network becomes more homogeneous, the disagreement variable will decrease. Each of the variables are calculated using the respondent's party identification in November 2004 and November 2006, respectively, and future discussion of these variables will refer to them as percent disagree in 04 and percent disagree in 06. The percent of disagreement within each network is documented for this data-set in Table 10.3. Note that there is significantly more reported disagreement using the party identification from 2004 then there is using the party identification from 2006.

Table 10.3 Goes Here

One assumption made here is that voters are able to accurately determine the partisanship of their social connections. Studies of the ability of voters to determine this have found that there is a bias for the respondent to over-report agreement with the discussant. In Huckfeldt and Sprague's 1996 Indianapolis-St. Louis Study they find that the respondents are 80% accurate when asked about the partisanship of their political discussants and that the respondents are more likely to be accurate if they agree with the discussant's true political preferences (Huckfelt, Johnson and Sprague 2004). We also anticipate that individual's discussion networks may have changed during this time span, so that the characteristics of the discussants who they identify in 2006 may in fact be different then those of their principle discussants in 2004, which complicates the identification of the causal mechanism. However, this data-set here provides a rare opportunity to both observe individual partisan reports and responses regarding their discussant characteristics.

10.4 Methods and Hypotheses

The advantage of this particular survey is that it is possible to observe the partisanship of the respondent at two distinct intervals of time and to simultaneously observe the

reported partisanship of her discussion network. This paper examines two hypotheses, outlined below, which are testable due to the opportunity to observe the respondents' change of party identification in the two-year window November 2004 to November 2006.

The first hypothesis is that if the party identification of the respondent in November 2004 is different then the party identification of the discussants, the probability that the respondent has changed party identifications in November 2006 will increase. It is possible to test this hypothesis by documenting whether or not the respondent has changed her party identification and to use the percent disagree in 04 as an explanatory variable. If the impact of percent disagree in 04 is positive, controlling for the selection of the discussants, then indeed the increased level of disagreement is a good predictor that the individual respondent will change party identification. To evaluate this hypothesis the statistical model will use an indicator for whether or not the individual has changed parties as a function of percent disagree in 04, the respondent's characteristics, and the discussant characteristics.

The second hypothesis is that, controlling for the change in party identification, if the party identification of the respondent in November 2004 is different then the party identification of the discussants, controlling for the selection of the discussants and for whether the respondent has changed parties, the probability that the respondent disagrees with the discussants in November 2006 will decrease. That is, if the respondent switches parties, it will be towards the direction of the discussants party identification. To evaluate this hypothesis the statistical model will describe percent disagree in 06 as a function of whether or not the individual has changed parties, percent disagree in 04, the respondent's characteristics, and the discussant characteristics.

Here both hypotheses are tested in three separate analyses — the first includes only the main variables of interest; the second incorporates a series of covariates about the respondent which are likely correlated to how she has selected discussants; the third incorporates those covariates and additionally includes a series of measurements to describe the discussants themselves. Given the small number of observations,

one disadvantage of incorporating the discussant characteristics is the presence of missing data which further decreases the number of observations available for analysis. Fortunately, however, all models produce similar results.

In the first model only the variables which are the focus of this analysis are included. The second model incorporates the respondent characteristics as control variables — these include ideological self-placement, home ownership, race, employment status, education level, marital status, age, income, gender, and whether or not the respondent is a parent. The third model adds an additional layer, incorporating not only the respondent-specific covariates but also those of the discussion network — these include the respondent's interest in discussing politics, the closeness of the relationship with the discussants, the frequency of conversation with the discussants, the geographic proximity of the discussants, and the discussant network density.

10.5 Results

Coefficients on the variables which are the focus of this analysis from each statistical model are presented in Table 10.4 and Table 10.5. Note here that all quantities of interest are statistically significant at traditional levels and have signs which support both hypotheses. Table 10.4 presents each of the coefficients on the variable percent disagree 04. Each of these coefficients are positive, with an interpretation that as disagreement between the respondent's party identification and the discussants' party identification increases, the respondent is increasingly likely to report a different party identification in November 2006. Table 10.5 presents each of the coefficients on both the percent disagree 04 as well as the partisan change indicator. For respondents who had disagreement in their discussion network in 2004, they are likely to have disagreement in their discussion network in 2006. However, controlling for disagreement in the discussion network in 2004, as the probability that the respondent changes party identifications increases, the amount of disagreement in the discussion network in 2006 will decrease. That is, as change increases, the amount of disagreement in 2006 decreases.

Table 10.4 and Table 10.5 Go Here

10.6 Conclusions

The results above provide support for the hypothesis that a voter's discussion network plays an enormous role in determining that individual's choice of party identification. The partisanship of an individual's discussion network appear to affect not only the probability that the individual will change party identification but also to predict to which party the individual will change.

One alternative explanation to the results above would be that the respondents are all influenced by policy changes in the time-span between surveys. Given the combination of both the outcome of the congressional midterms and the electorate's overall disapproval of the war in Iraq — as well as the fact that there are more voters who switch to Democratic or Independent party identifications then who switch to identify themselves as Republicans, it seems possible that the voters in this study are all simultaneously influenced by a change in party location. However, the analysis attempts to incorporate this possibility by controlling for the respondent's ideological self-placement, so that a change in party identification may not be entirely explained by political events. One alternative in this analysis would be to examine the interaction between the indicator for partisan change and the amount of disagreement present when the respondent is first asked for party identification. This would add additional support to the idea that individuals are most likely to change party identification when exposed to political disagreement within their social context.

Throughout the literature, regardless of whether party identification is seen as stable or flexible, party identification is clearly a powerful predictor of vote choice. This analysis demonstrates that party identification is a component of more than a transitive presidential approval, but rather a component of social group membership. Individuals identify with political parties in much the same fashion that they identify with other social identifying characteristics, and thus political identification shifts occur only when the individual has begun to reexamine her own group memberships.

These findings have important implications for polarization in America and the role of heterogeneous political networks in providing moderate political preferences. Party identification, says Campbell et al., "raises a perceptual screen through which the individual tends to see what is favorable to his partisan orientation. The stronger the bond, the more exaggerated the process of selection and perceptual distortion will be" (1960, p. 133). If individuals are influenced by their peers to choose party identifications similar to those around them, and if indeed party identification will bias an individual's opinions in any particular direction, then it becomes increasingly necessary to encourage debate and friendship across parties to maintain a politically balanced electorate.

10.7 Appendix

ANES Pilot 2006 Questions

Q: During the last six months, did you talk with anyone face-to-face, on the phone, by email, or in any other way about [things that were important to you / government and elections], or did you not do this with anyone during the last six months?

Yes, did talk

No, did not talk

Don't know

Refused

Q: What are the initials of the people who you talked with face-to-face, on the phone, by email, or in any other way during the past six months, about [things that were important to you / government and elections]?

Q: For each name, what is that person's gender?

Male

Female

Q: How close do you feel to [NAME 1/ NAME 2/ NAME 3]?
Extremely close
Very close
Moderately close
Slightly close
Not close at all

Q: During the last six months, about how many days did you talk to [NAME 1/ NAME 2/ NAME 3]?

Q: First think about [NAME 1] and [NAME 2]. During the last six months, about how many days would you guess they talked to each other?

Q: Now think about [NAME 1] and [NAME 3]. During the last six months, about how many days would you guess they talked to each other?

Q: Now think about [NAME 2] and [NAME 3]. During the last six months, about how many days would you guess they talked to each other?

Q: In general, how different are [NAME 1/ NAME 2/ NAME 3]'s opinions about government and elections from your own views?
Extremely different
Very different
Moderately different
Slightly different
Not different at all

Q: Generally speaking, does [NAME 1/ NAME 2/ NAME 3] probably think of [himself/herself] as a Democrat, Republican, Independent, or what?
Democrat

Republican

Independent

Other

No preference

Q: Would [he/she] call [himself/herself] a strong [Democrat/Republican] or a not very strong [Democrat/Republican]?

Strong

Not very strong

Don't know

Refused

Q: Does [he/she] think of [himself/herself] as closer to the Democratic Party or the Republican Party?

Closer to Democratic

Closer to Republican

Neither

Q: How interested is [NAME 1/NAME 2/ NAME 3] in information about what's going on in government and politics?

Extremely interested

Very interested

Moderately interested

Slightly interested

Not interested at all

Q: How much time would it take to drive from your home to [NAME 1/NAME 2/NAME 3]'s home?

Time given

Can't drive there

Lives with [NAME]

10.8 Tables

Table 10.1: Respondent Party Change, November 04 to November 06

	Democrat 06	Independent 06	Republican 06	No Pref 06	Other 06
Democrat 04	0	13	4	2	2
Independent 04	61	0	28	10	8
Republican 04	18	23	0	0	5
No Pref 04	9	5	4	0	1
Other 04	2	0	0	0	0

The values in this table indicate the number of individuals who reported changing their party identification from the row category (their 2004 response) to the column category (their 2006 response).

Table 10.2: Summary of Variables

Variable	Mean	Min	Max	N
Respondent Characteristics				
Party Change	.3	0	1	661
Total Discussants	1.86	0	3	661
Percent Disagree in 04	.35	0	1	661
Percent Disagree in 06	.30	0	1	661
Female	.54	0	1	661
Parent	.21	0	1	661
Income Bracket	12.07	1	23	661
Age	49.75	18	90	661
Marital Status	.56	0	1	661
Education Level	14.2	4	17	661
Employment Status	.67	0	1	661
Nonwhite	.18	0	1	661
Homeowner	.78	0	1	661
Ideological Self-placement	5.12	0	10	632
Interest in Politics	2.21	0	5	661
Discussant Averages				
Interest in Politics	2.41	0	5	468
Perception of Disagreement	3.62	1	5	467
Intensity of Relationship	1.94	1	5	468
Frequency of Conversation	49.67	0	182	314
Identity as Strong Partisan	.66	0	1	439
Geographic Distance (in minutes)	320.28	0	18000	416
Network Density	1.02	0	3	661

Table 10.3: Discussion Network Disagreement: November 04 and November 06

Percent Disagreement	November 04	November 06
0	325	343
.33	92	112
.5	29	30
.66	95	89
1	120	87
N	661	661

Respondents could name up to three discussants — thus the percentage of an individual's discussion network which had a different partisan label is discrete and falls into the row categories. The columns indicate the fraction of disagreement reported in 2004 and in 2006, respectively.

Table 10.4: Logit Coefficients, Effect of Disagreement in 04 on Change

Variable	Coefficient	Controls	N
Percent Disagree 04	1.28*	—	661
	(.22)		
Percent Disagree 04	1.49*	Respondent characteristics	584
	(.25)		
Percent Disagree 04	3.39*	Respondent and network characteristics	265
	(.58)		
$* = \alpha = .05$			

The dependent variable is an indicator for whether or not the respondent had changed partisan affiliations between 2004 and 2006. The independent variable of interest is the percent of a respondent's discussion network which had a different partisan label than herself in 2004. Three separate models are presented here; one with only the percent disagreement, one with percent disagreement and respondent characteristics, and one with percent disagreement, respondent characteristics and network characteristics.

Table 10.5: Ordered Logit Coefficients, Effect of Change on Disagreement in 06

Variable	Coefficient	Controls	N
Change	-1.76*	—	661
	(.26)		
Percent Disagree 04	6.85*		
	(.37)		
Change	-1.63*	Respondent characteristics	584
	(.27)		
Percent Disagree 04	6.97*		
	(.40)		
Change	-1.48*	Respondent and network characteristics	265
	(.37)		
Percent Disagree 04	7.25*		
	(.62)		
$* = \alpha = .05$			

The dependent variable is the fraction of disagreement within a respondent's discussion network in 2006. The independent variables of interest are an indicator for whether or not the respondent had changed partisan affiliations between 2004 and 2006 as well as a measurement of the partisan disagreement within a respondent's discussion network in 2004. Three models are considered, one with simply the variables of interest, one which adds the respondent characteristics and a final which includes both respondent and network characteristics.

Chapter 11

Conclusion

The empirical chapters which follow from the theory chapter all relate an individual's discussion network to a particular political behavior. In the June 2006 and November 2006 elections in South Los Angeles, it is clear that individuals who received a door-to-door mobilization treatment were more likely to turn out to vote. Additionally, they were increasingly likely to vote if that contact was also a resident of their precinct. Precincts are small geographically compact units of no more than 1000 individuals, and thus it is likely that the home-turf canvassers were often speaking to people they knew. This peer-pressured component of voter mobilization not only appears to be effective but also suggests that individuals are influenced in a particular way from others in their social framework. A significant question remains from this analysis: Is the increase in turnout due to the desire of the treatment individuals to conform to social pressures, or is the increase in turnout due to the weight that the treatment individuals associate with communication from a trusted source? An answer to this question underlies the fundamental behavioral assumptions associated with an individual's voting behavior. In the theoretical model, voters are assumed to transmit information to each other and to base their vote choice upon that information. If their discussants are too ideologically different from themselves — and their communication is sincere about their most-preferred candidate — then voters are most likely to adopt new positions after communicating with individuals who have similar ideological preferences. This is consistent with the empirical finding that individuals are most likely to turn out to vote if contacted by someone who is known to be

similar — a precinct member — which results in a 7% marginal increase if contacted by a precinct member. Although the field experiments deals with mobilization and persuasion, it is a good example of how social members can influence others. One advantage of the field experiment is that there is a clear causal mechanism — the canvasser increases turnout by communicating the message to the voter.

The laboratory experiment chapter investigates the extent to which individual votes will rely upon Bayes Rule to update their priors about candidates when the voters are themselves uninformed about the candidate locations. This links nicely into the theoretical argument that individual voters will rely upon the communication phase to update their beliefs about which candidate is relatively more competent. Voters' behavior and choices fit Bayes Rule better than any of the other behavioral possibilities explored in the paper. Again, this paper is not a direct test of the theoretical setup, as it is exploring voter beliefs when the voters are uninformed about the candidate location, whereas the theoretical paper investigates voter beliefs about a common, competency dimension. However, given the results presented here and the quality of the convergence towards the true candidate midpoint, the assumption that the voters will use Bayes Rule to update their priors even though there is no additional increase in payoff from sincere communication, is consistent with the empirical data and thus may be used in the formal model. Furthermore, this underlies assumptions about what voters will do during communication. So long as it is not in their strategic best interest to misrepresent their own preferences, but instead they weakly prefer to communicate sincerely (as is the case in the laboratory experiment setup), voters do appear to communicate sincerely. This also ties nicely into the theoretical framework.

The greatest contribution to the literature on political behavior is in the survey data section. Here, three chapters each discuss a different data-set but the results are remarkably consistent across all of them — when voters are asked about the political preferences of their discussions, the political preferences of their discussions can be seen to be excellent predictors of the respondent's preferences, controlling for socioeconomic and demographic characteristics. Here voters are seen to have chosen similar candidates or partisan identifications consistent with those within their

discussion framework. Each of these results ties nicely back into the theoretical argument that there will be correlation in vote choice across discussion networks.

With all of this empirical data, it is then possible to assume that the candidates, who are assumed to be strategic in their choice of policy platforms, would consider the discussion network ramifications in their choice of policy. Thus as the voters themselves were increasingly likely to have conversations with individuals who are similar to themselves, there would be an increasing effect in the correlation of their behaviors. We observe in the empirical survey section that were a voter to move from a majority Republican discussion network to a majority Democratic discussion network, this switch would have the possibility of changing the probability that she voted for a Democrat by 25%. As more and more individuals are able to select into particular discussion networks, it becomes increasingly likely to observe bubbles of similar voting within social frameworks. The theoretical model argues that these correlations will allow candidates to move away from the preferences of the median voter and become increasingly ideologically polarized. This suggests that one potential causal mechanism for the polarization observed amongst America's elite representatives could be due to the changes in communication structures. As voters are increasingly able to maintain geographically distant friendships (which the survey data validates — over 40% of individuals have close relationships with partisan discussants who do not live in the same city), voters are more able to select discussants who are more similar to them along socioeconomic and demographic characteristics — variables which are typically associated with political preferences. Thus individuals are less likely to be exposed to a politically diverse community. The limited exposure has the potential to produce polarized candidates.

The implications of this finding can be observed in the current presidential campaign. Campaigns are increasingly attempting to incorporate social pressures in support for particular candidates, such as the "MyBarakObama.com" website which allows people to form friendship groups or to invite friends to participate in online social networking in conjunction with the Barak Obama presidential campaign. If peer-to-peer communication encourages Obama support, then voters are significantly

more likely to vote for Obama even if they procure no new additional information, but simply exchange information within a politically homogeneous social framework. Future research of campaigns and elections will need to incorporate measurements of how campaigns have tapped into individual social networks.

Researchers have long observed that individuals behave similarly to others like them in terms of socioeconomic and demographic characteristics. This dissertation demonstrates one explanation of this fact, which is that individuals are likely to form friendships with those who are like them and to exchange information within those friendship networks. This exchange then prompts individuals to update their beliefs with similar information and then to have political behaviors which are similar. While studying small changes in partisanship or vote choice based upon political network communication may not initially seem like a significant cause for alarm, it is the case that other extremely negative outcomes could be supported from the lack of exposure to political heterogeneous ideals. Furthermore, polarized representatives decrease the average utility of all voters — and normatively provide low-quality representation. This dissertation suggests that institutions which would increase dialogue across ideological lines would provide for better representation and for a more informed electorate.

Chapter 12

Future Work

One potential direction for future work is to examine the role of social ties in explaining participation in collective action movements. In particular, it is interesting to examine whether if it is the case that there was an existing social network which collectively mobilized for a particular cause once the cause was realized or if there were a group of elites which convinced several social group members to mobilize their social networks. If collective action movements could be initiated by radicalizing only a few individuals, who could then enforce their social network to participate, it would have the potential to explain individual participation in large group movements.

Bibliography

[1] Abramowitz, Alan I. and Kyle L. Saunders. (1998) "Ideological Realignment in the U.S. Electorate." *The Journal of Politics*, Vol. 60, No. 3.

[2] Abramowitz, Alan. (1983) "Social Determinism, Rationality, and Partisanship Among College Students." *Political Behavior*, Vol. 5, No. 4.

[3] Achen, Christopher H. (2002) "Parental Socialization and Rational Party Identification." *Political Behavior*, Vol. 24, No. 2.

[4] Achen, Christopher. (1992) "Social Psychology, Demographic Variables, and Linear Regression: Breaking the Iron Triangle in Voting Research." *Political Behavior*, Vol. 14.

[5] Achen, Christopher. (1975) "Mass Political Attitudes and the Survey Response." *The American Political Science Review*, Vol. 69, No. 4.

[6] Almond, Gabriel A., and Sidney Verba. (1963) *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Princeton: Princeton University Press.

[7] Alford, John, Carolyn Funk, and John R. Hibbing. (2005) "Are Political Orientations Genetically Transmitted?" *American Journal of Political Science*, Vol. 99, No. 2, May.

[8] Alvarez, R. Michael, and Jonathan Nagler. (2000) "A New Approach for Modeling Strategic Voting in Multiparty Elections." *British Journal of Political Science*, Vol. 30, No. 1, 57-75.

[9] Alvarez, R. Michael, and Jonathan Nagler. (1998) "When Politics and Models Collide: Estimating Models of Multi-Party Elections." *American Journal of Political Science*, Vol. 42, No. 1, p. 55-96.

[10] Alvarez, R. Michael, and John Brehm. (2002) *Hard Choices, Easy Answers: Values, Information, and American Public Opinion.* Princeton: Princeton University Press.

[11] Alvarez, R. Michael. (1998) *Information and Elections.* University of Michigan Press.

[12] Austen-Smith, David, and Jeffrey S. Banks. (1996) "Information Aggregation, Rationality, and the Condorcet Jury Theorem." *American Political Science Review*, Vol. 90, No. 1.

[13] Bartels, Larry M. (1992) "Beyond the Running Tally: Partisan Bias in Political Perceptions." *Political Behavior*, Vol. 24, No. 2.

[14] Belknap, George, and Angus Campbell. (1952) "Political Party Identification and Attitudes Toward Foreign Policy." *Public Opinion Quarterly*, Vol. 15.

[15] Berelson, Bernard R., Paul F. Lazarsfeld, and William N. McPhee. (1954) *Voting: A Study of Opinion Formation in a Presidential Campaign.* Chicago: Chicago University Press.

[16] Berelson, Bernard, Paul F. Lazarsfeld, and William N. McPhee. (1954) *Voting*. Chicago: University of Chicago Press.

[17] Brody, Richard A., and Lawrence Rothenberg. (1988) "The Instability of Partisanship: An Analysis of the 1980 Presidential Election." *British Journal of Political Science*. Vol. 18.

[18] Campbell, Angus, Philip E. Converse, Warren Miller and Donald Stokes. (1960) *The American Voter*. London: University of Chicago Press.

[19] Cassel, Carol A. (1983) "Predicting Party Identification, 1956-80: Who Are the Republicans and Who Are the Democrats?." *Political Behavior*, Vol. 4, No. 3.

[20] Cillizza, Chris. (2006) "The Fix: RNC Doles Out Cash, DNC Goes Door to Door", *Washington Post*.

[21] Collie, Melissa, and John Mason. (1999) "The Electoral Connection Between Party and Constituency Reconsidered." *Continuity and Change in House Elections*. David Brady, John Cogan, and Morris Fiorina, editors. Stanford: Stanford University Press.

[22] Coleman, James. (1988) "Free Riders and Zealots: The Role of Social Networks." *Sociological Theory*, Vol. 6, No. 1, 52-57.

[23] Conover, Pamela Johnston, and Stanley Feldman. (1982) "Projection and the Perception of Candidates." *Western Political Quarterly*, Vol. 35.

[24] Conover, Pamela Johnston, and Stanley Feldman. (1989) "Candidate Perception in an Ambiguous World: Campaigns, Cues, and Inference Processes." *American Journal of Political Science*, Vol. 33, No. 4.

[25] Converse, Philip. (1962) "Information Flow and the Stability of Partisan Attitudes." *Public Opinion Quarterly*, Vol. 26, p. 578-599.

[26] Converse, Philip E. (1964) "The Nature of Belief Systems in Mass Publics", *Ideology and Discontent*. David E. Apter, editor, New York: Free Press.

[27] Converse, Philip E. (1966) "Information Flow and the Stability of Partisan Attitudes." *Elections and the Political Order*. Angus Campbell et al., editors, New York: Wiley and Sons.

[28] Cornfield, Michael. (2004) "The Internet and Campaign 2004: A Look Back at the Campaigners". *Pew Internet and American Life Project*, (working paper), http://www.pewinternet.org/pdfs/Cornfield.commentary.pdf.

[29] Diamond, Alexis and Jasjeet Sekhon, (2007) "Genetic Matching for Estimating Causal Effects." (unpublished manuscript), http://sekhon.polisci.berkeley.edu/matching/.

[30] Dobson, Douglas and Duane Meeter. (1974) "Alternative Markov Models for Describing Change in Party Identification. *American Journal of Political Science*, Vol. 18.

[31] Dobson, Douglas, and Douglas St Angelo. (1975) "Party Identification and the Floating Vote: Some Dynamics." *American Political Science Review*, Vol. 69.

[32] Downs, Anthony. (1957) *An Economic Theory of Democracy*. Addison Wesley.

[33] Dreyer, Edward. (1983) "Change and Stability in Party Identification." *Journal of Politics*, Vol. 35.

[34] Eulau, Heinz. (1963) *The Behavioral Persuasion in Politics*. New York: Random House.

[35] Feddersen, Timothy J and Pesendorfer, Wolfgang. (1996) "The Swing Voter's Curse," *American Economic Review*, Vol. 86, No. 3, p. 408-424.

[36] Feldman, Stanley. (1988) "Structure and Consistency in Public Opinion: The Role of Core Beliefs and Values." *American Journal of Political Science*, Vol. 32, No. 2.

[37] Fiorina, Morris. (2004) *Culture War? The Myth of a Polarized America*. Longman.

[38] Fiorina, Morris P. (1981) *Retrospective Voting in American National Elections.* New Haven: Yale University Press.

[39] Franklin, Charles H. (1984) "Issue Preferences, Socialization, and the Evolution of Party Identification." *American Journal of Political Science*, Vol. 28, No. 3.

[40] Franklin, Charles H., and John E. Jackson. (1983) "The Dynamics of Party Identification." *American Political Science Review*, Vol. 77.

[41] Gerber, Alan, and Donald P. Green. (1998) "Rational Learning and Partisan Attitudes." *American Journal of Political Science*, Vol. 42, No. 3.

[42] Gerber, Alan and Donald Green. (2001) "Do Phone Calls Increase Turnout? A Field Experiment." *Public Opinion Quarterly*, Vol. 65, No. 1.

[43] Gerber, Alan, and Donald Green. (2000) "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *American Political Science Review*, Vol. 94, No. 3.

[44] Green, Donald, Alan Gerber, and David Nickerson. (2003) "Getting out the vote in local elections: Results from six door-to-door canvassing experiments." *Journal of Politics*, Vol. 65.

[45] Green, Donald, Bradley Palmquist, and Eric Schickler. (2002) *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. New Haven: Yale University Press.

[46] Green, Donald, and Bradley Palmquist. "Of Artifacts and Partisan Instability." *American Journal of Political Science*, Vol. 34, No. 3.

[47] Goldbert, Arthur S. (1966) "Discerning a Causal Pattern among Data on Voting Behavior." *American Political Science Review*. Vol. 50.

[48] Goren, Paul. (2005) "Party Identification and Core Political Values." *American Journal of Political Science*, Vol. 49, No. 4.

[49] Hamilton, David. (1981) "Cognitive Representation of Persons." *Social Cognition*. E. Tory Higgins, Peter Herman and Mark Zanna, editors, Hillsdale: Erlbaum.

[50] Higgins, E. Tory, and Gillian King. (1981) "Accessibility of Social Constructs: Information Processing Consequences of Individual and Contextual Variability." *Personality, Cognition and Social Interaction*. Nancy Cantor and John Kihlstrom, editors, Hillsdale: Erlbaum.

[51] Ho, Daniel E., Kosuke Imai, Gary King and Elizabeth Stuart. (2005) "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." (unpublished manuscript), http://gking.harvard.edu/projects/cause.shtml.

[52] Ho, Daniel E., Kosuke Imai, Gary King and Elizabeth A. Stuart. (2005) "MATCHIT: Nonparametric Preprocessing for Parametric Causal Inference." http://gking.harvard.edu/matchit/docs/matchit.html.

[53] Honaker, James, Gary King and Matthew Blackwell. (2001) "Amelia II: A Program for Missing Data." http://gking.harvard.edu/amelia.

[54] Huckfeldt, Robert, Paul E. Johnson, and John Sprague. (2004) *Political Disagreement*. Cambridge: Cambridge University Press.

[55] Huckfeldt, Robert, and John Sprague. (1995) *Citizens, Politics, and Social Communication*. New York: Cambridge University Press.

[56] Huckfeldt, Robert, Paul Allen Beck, Russell J. Dalton, Jeffrey Levine, and William Morgan. (1998) "Ambiguity, Distorted Messages, and Nested Environmental Effects on Political Communication." *Journal of Politics*, Vol. 60: 996-1040.

[57] Huckfeldt, Robert, John Sprague, and Jeffrey Levine. (2000) "The Dynamics of Collective Deliberation in the 1996 Election: Campaign Effects on Accessibility, Certainty, and Accuracy." *American Political Science Review*, Vol. 94: 641-51.

[58] Huckfeldt, Robert, Levine, Jeffrey, Morgan, William and Sprague, John. (1998) "Election Campaigns, Social Communication, and the Accessibility of Perceived Discussant Preference." *Political Behavior*, Vol. 20, No. 4.

[59] Huckfeldt, Robert, and Sprague, John. (1988) "Choice, Social Structure, and Political Information: The Information Coercion of Minorities." *American Journal of Political Science*, Vol. 32, No. 2, 467-482.

[60] Huckfeldt, Robert, and Sprague, John. (1987) "Networks in Context: The Social Flow of Political Information." *The American Political Science Review*, Vol. 81, No. 4, 1197-1216.

[61] Hung, Angela, and Charles R. Plott. (1999) Information Cascades: A Replication and Extension to Majority Rule and Grand Jury Instructions, *American Economic Review*.

[62] Imai, Kosuke, and Van Dyk, David. (2004) "Causal Inference with General Treatment Regimes: Generalizing the Propensity Score", *Journal of the American Statistical Association*, Theory and Methods. September, Vol. 99, No. 467.

[63] Imai, Kosuke, Gary King, and Elizabeth Stuart. (2005) "The Balance Test Fallacy." http://gking.harvard.edu/files/balfal.pdf.

[64] Imai, Kosuke, Gary King, and Olivia Lau. "Zelig: Everyone's Statistical Software." http://gking.harvard.edu/zelig.

[65] Imbens, Guido. (2000) "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*, Vol. 87. No. 3, Sep, p. 706-710.

[66] Iyengar, Shanto. (1976) "Childhood Learning of Partisanship in a New Nation: The Case of Andhra Pradesh." *American Journal of Political Science*. Vol. 20, No. 3, August.

[67] Jackson, John. "Issues, Party Choices, and Presidential Votes." *American Journal of Political Science*. Vol. 19.

[68] Jennings, M. Kent, and Richard Niemi. (1981) *Generations and Politics: A Panel Study of Young Adults and their Parents*. Princeton University Press.

[69] Jennings, M. Kent, and Richard G. Niemi. (1968) "The Transmission of Political Values from Parent to Child." *American Journal of Political Science*. Vol. 62, March.

[70] Jennings, M Kent and Gregory Markus. (1984) "Partisan Orientations Over the Long Haul." *American Political Science Review*, Vol. 77.

[71] Kartik, Navin and McAfee, R. Preston. (2006) "Signaling Character in Electoral Competition." (Caltech working paper.) http://ssrn.com/abstract=873404.

[72] Keith, Bruce E., et al. (1992) *The Myth of the Independent Voter*. Berkeley: University of California Press.

[73] Keith, Bruce, et al.. (1986) "The Partisan Affinities of Independent Leaners." *British Journal of Political Science*, Vol. 16.

[74] Kinder, Donald R. and David O. Sears. (1985) "Public Opinion and Political Action", *Handbook of Social Psychology*. Gardner Lindzey and Elliot Aronson, editors, New York: Random House.

[75] Kinder, Donald R. (1978) "Political Person Perception: The Asymmetrical Influence of Sentiment and Choice on Perceptions of Presidential Candidates." *Journal of Personality and Social Psychology*, Vol. 36.

[76] King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. (2001) "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review*, Vol. 95, No. 1, March.

[77] Knoke, David. (1990) "Networks of Political Action: Toward Theory Construction." *Social Forces*, Vol. 68, No. 4, p.1041-1063.

[78] Knoke, David and Michael Hout. (1974) "Social and Demographic Factors in American Political Party Affiliation, 1952-72." *American Sociological Review*. Vol. 39.

[79] Lazarsfeld, Paul F., Bernard Berelson and Hazel Gaudet. (1948) *The People's Choice*. New York: Columbia University Press.

[80] Liedka, Raymond V. (1991) "Who Do You Know in the Group?: Location of Organizations in Interpersonal Networks." *Social Forces*, Vol. 70, No. 2, p. 455-474.

[81] Lupia, Arthur and Mathew McCubbins. (1998) *The Democratic Dilemma*. Cambridge University Press.

[82] Markus, G. B., and P. Converse. (1979) "A Dynamic Simultaneous Equation Model of Electoral Choice." *American Political Science Review*, Vol. 73.

[83] Marsden, Peter V. (1990) "Network Data and Measurement." *Annual Review of Sociology*, Vol. 16, p. 435-463.

[84] Marsden, Peter V. (1987) "Core Discussion Networks of Americans." *American Sociological Review*, Vol. 52, No. 1, p. 122-131.

[85] Manski, Charles F. (1993) "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, Vol. 60, No. 3.

[86] Martinez, Michael D. (1990) "Partisan Issue Preferences and Partisan Change." *Political Behavior*, Vol. 12, No. 3.

[87] McKelvey, Richard and Peter Ordeshook. (1985) "Elections with Limited Information: A Fulfilled Expectations Model Using Contemporaneous Poll and

Endorsement Data as Information Sources." *Journal of Economic Theory*, Vol. 36, 55-85.

[88] McKelvey, Richard D. and Peter C. Ordeshook. (1985) "Sequential Elections with Limited Information." *American Journal of Political Science*, Vol. 29, No. 3, pg 480-512.

[89] McKelvey, Richard and Peter Ordeshook. (1990) "Information and Elections: Retrospective Voting and Rational Expectations." *Information and the Democratic Processes*. Ed. John A. Ferejohn and James H. Kuklinski, Chicago: University of Chicago.

[90] McKelvey, Richard D. and Peter C. Ordeshook. (1986) "Information, Electoral Equilibria, and the Democratic Ideal." *Journal of Politics*, Vol. 48, No. 4, pg 909-937.

[91] Meier, Kenneth J. (1975) "Party Identification and Vote Choice: The Causal Relationship." *Western Political Quarterly*, Vol. 28.

[92] Michelson, Melissa R. (2003) "Getting Out the Latino Vote: How Door-to-Door Canvassing Influences Turnout in Rural Central California", *Political Behavior*, Vol. 25, No. 3, p. 247-263.

[93] Milgram, Stanley. (1967) "The Small World Problem", *Psychology Today*, p. 60-67.

[94] Miller, Warren E. and J. Merrill Shanks. (1966) *The New American Voter*. Cambridge: Harvard University Press.

[95] Miller, Arthur, Warren Miller, Alden Raine and Thad Brown. (1976) "A Majority Party in Disarray: Policy Polarization in the 1972 Election." *American Political Science Review*, Vol. 70.

[96] Mobius, Markus, Paul Nichaus and Tanya S. Rosenblat. (2005) "Social Learning and Consumer Demand." (Working Paper) http://trosenblat.web.wesleyan.edu/home

[97] Mutz, Diana C. (2006) *Hearing the Other Side*. Cambridge University Press.

[98] Nickerson, David. (2006) "Volunteer Phone Calls Can Increase Turnout: Evidence from Eight Field Experiments." *American Politics Research*, Vol. 34.

[99] Niemi, Richard, Richard Katz and Donald Newman. "Reconstructing Past Partisanship: The Failure of the Party Identification Recall Questions." *American Journal of Political Science*, Vol. 24.

[100] Niven, David. (2004) "The Mobilization Solution? Face-to-Face Contact and Voter Turnout in a Municipal Election." *The Journal of Politics*, Vol. 66, No. 3.

[101] Ordeshook, Peter C. and Thomas R. Palfrey. (1988) "Agendas, Strategic Voting, and Signaling with Incomplete Information." *American Journal of Political Science*, Vol. 32, No. 2, p. 441-446.

[102] Page, Benjamin and Calvin Jones. (1979) "Reciprocal Effects of Policy Preferences, Party Loyalties, and the Vote." *American Political Science Review*, Vol. 73.

[103] Page, Benjamin and Richard Brody. (1972) "Reciprocal Effects of Policy Preferences, Party Loyalties, and the Vote." *American Political Science Review*. Vol. 73.

[104] Palfrey, Thomas and Keith Poole. (1987) "The Relationship between Information, Ideology and Voting Behavior". *American Journal of Political Science*, Vol. 31, No. 3.

[105] Popkin, S., J.S. Gorman, C. Phillips, and J.A. Smith. (1976) "What Have you Done for Me Lately: Toward an Investment Theory of Voting." *American Political Science Review*, Vol. 70, p. 779-805.

[106] Poole, Keith and Howard Rosenthal. (1984) "The Polarization of American Politics." *Journal of Politics*, Vol. 46.

[107] Putnam, Robert. (1966) "Political Attitudes and the Local Community." *The American Political Science Review*, Vol. 60, No. 3, p. 640-654.

[108] Putnam, Robert D. (2000) *Bowling Alone*. Simon and Schuster: New York.

[109] Rao, Neel, Markus Mobius and Ranya Rosenblat. (2006) "Social Networks and Vaccination Decisions." (Working Paper) http://trosenblat.web.wesleyan.edu/home.

[110] Rogers, Brian and Matthew Jackson. (2005) "The Economics of Small Worlds", *Journal of the European Economic Association*, Vol. 3, No. 23, p. 607-627.

[111] Rosenbaum, Paul and Rubin, Donald. (1984) "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score", *Journal of the American Statistical Association*, Vol. 79, No. 387 (Sep).

[112] Rosenbaum, Paul and Donald Rubin. (1983) "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, Vol. 70, p. 41-55.

[113] Schofield, Norman. "Valence Competition in the Spatial Stochastic Model." *Journal of Theoretical Politics*, Vol. 15, No. 4.

[114] Schofield, Norman. "Equilibrium on the Spatial 'Valence' Model of Politics". *Journal of Theoretical Politics*, Vol. 16, No. 4.

[115] Sekhon, J.S. (2004), "Multivariate and Propensity Score Matching Software." http://jsekhon.fas.harvard.edu/matching.

[116] Sheingold, Carl. (1973) "Social Networks and Voting: The Resurrection of a Research Agenda." *American Sociological Review,* Vol. 38, No. 6, p. 712-720.

[117] Straits, Bruce C. (1991) "Bringing Strong Ties Back in Interpersonal Gateways to Political Information and Influence." *Public Opinion Quarterly*, Vol. 55, No. 3, p. 432-448.

[118] Stokes, Donald. (1966) "Party Loyalty and the Likelihood of Deviating Elections." *Elections and the Political Order*, Angus Campbell et al., editors, New York: John Wiley and Sons.

[119] Sunstein, Cass. (2001) *Republic.com.* Princeton University Press: Princeton.

[120] Todorov, Alexander, Anesu Mandisodza, Amir Goren, Crystal Hall. (2005) "Inference of Competence from Faces Predict Election Outcomes." *Science*, Vol. 203, No. 5728.

[121] Tomz, Michael, Jason Wittenberg and Gary King. (2001) "Clarify: Software for Interpreting and Presenting Statistical Results". http://gking.harvard.edu/clarify/docs/clarify.html.

[122] Wearing, Donald, D. J. Schwartz, and A. E. Lind. (1973) "The Structuring Principle: Political Socialization and Belief Systems." *American Political Science Review*. Vol. 67, June.

[123] Weatherford, M. Stephen. (1982) "Interpersonal Networks and Political Behavior." *American Journal of Political Science*, Vol. 26, No. 1, p. 117-143.

[124] Wittman, Donald. (1989) "Why Democracies Produce Efficient Results." *Journal of Political Economy*, Vol. 97, No. 6, p. 1395-1424.

[125] Wolfinger, Raymond E. and Steven J. Rosenstone. (1980) *Who Votes?*. New Haven: Yale University Press.

[126] Verba, Sidney, Kay Lehman Schlozman and Henry E. Brady. (1995) *Voice and Equality*. Cambridge: Harvard University Press.

[127] Zaller, John, Stanley Feldman. (1992) "A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences." *American Journal of Political Science*, Vol. 36, No. 3.