Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Michael K. Hanley Date

Three Essays on Information and Accountability in American Politics

By

Michael K. Hanley
Doctor of Philosophy

Political Science

John Patty, Ph.D.
Co-Advisor

Elizabeth Maggie Penn, Ph.D.
Co-Advisor

Alan Abramowitz, Ph.D.
Committee Member

Alexander Bolton, Ph.D
Committee Member

Clifford Carrubba, Ph.D.
Committee Member

Accepted:

Kimberly Jacob Arriola, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Three Essays on Information and Accountability in American Politics

By

Michael K. Hanley
B.A., Princeton University, NJ, 2013
M.Ed., Marquette University, WI, 2016

Advisors: John Patty, Ph.D. & Elizabeth Maggie Penn, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Political Science
2022

Abstract

Three Essays on Information and Accountability in American Politics
By Michael K. Hanley

Accountability is central to understanding the distribution, use, and abuse of power in American politics. Information is required to administer accountability. This dissertation examines the relationship between information and accountability in three separate contexts in American politics: (1) In a survey experiment, I demonstrate that, on average, American voters respond to policy-based information, assuaging a legitimate concern that they behave like tribal partisans. Yet paradoxically, party ID performs just as well as having complete information about candidate policy preferences. (2) Through a formal model of credit claiming in an election, I prove that it is incentive-compatible to challenge exaggerated claims less than one might intuitively expect. Further, I show that while media accountability and voter accountability do limit the incentives for exaggerated claiming, increased media accountability diminishes the impact of voter accountability. (3) In the realm of bureaucratic accountability, I develop a machine-learning-based measure of responsiveness, leveraging the information-sharing that occurs during the appropriations process. This measure is validated by comparing it to another measure of accountability, limitation riders. Its usefulness for future research is then explored by using it to arrive at novel results related to agency politicization and responsiveness to Congress.

Overall, this dissertation provides three separate contributions to our understanding of information and accountability, and suggests several broader implications: (1) More information does not equate to more accountability, and (2) the coarseness of information might correspond to the ideal coarseness of the accountability judgement. These suggested implications provide fruitful avenues for further research.

Three Essays on Information and Accountability in American Politics

By

Michael K. Hanley
B.A., Princeton University, NJ, 2013
M.Ed., Marquette University, WI, 2016

Advisors: John Patty, Ph.D. & Elizabeth Maggie Penn, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Political Science
2022

Acknowledgments

2020 was the year that I began thinking longer-term about a dissertation, having just completed coursework and passed the departmental paper requirements. However, most people will rightfully remember 2020 for a very different reason. As the Covid-19 pandemic rages on while I finish this dissertation, I reflect back and am reminded of the many individuals, groups, and institutions that have helped me continue my work through the past few years. I know that in isolation, and not just from Covid, this work would never have completely come together, and I am forever thankful for those that have supported me throughout this process.

I would first like to express my deep gratitude for my thesis advisors, John Patty and Maggie Penn. Their academic expertise has been greatly helpful in informing paths that would or would not be productive for a dissertation, particularly during the summer of 2020 when I felt like I was revving my wheels. They have always been willing to meet to discuss topics, provide suggested readings, and even provide quick feedback on submitted work. But most importantly, I appreciate the grace, understanding, and compassion that they have extended throughout this process. The Covid-19 pandemic affected me more than I was willing to admit, yet they were always there to provide support, guidance, motivation, and occasionally the necessary prodding to ensure that I was making sufficient progress.

Next, I would like to thank my specific paper advisors and co-authors, who are also members of my committee. Alan Abramowitz provided helpful sources and advice for Paper 1, and was even supportive when the experimental results indicated that voters are not tribal partisans when prioritizing selecting on policy preferences. Clifford Carrubba provided continuous support throughout the development of my model for Paper 2, which transformed it from a 1st year RD submission into a project that actually has legs. Alex Bolton has been an ideal co-author for Paper 3, as well as a fantastic personal mentor. From my 1st year of graduate school until present, he has always pushed me to develop my quantitative capabilities, and helped me filter through feedback that was important for

my development. Through writing this project together, Alex has also provided me with invaluable “hands-on" experience in academic writing.

My progress on this academic endeavor did not (entirely) take place in isolation, and I would like to thank the Department of Political Science at Emory for providing the educational, professional, and administrative support necessary to complete this dissertation. I would like to thank Natália Bueno, Zac Peskowitz, Tom Clark, and Pablo Montagnes for specific points of feedback on the papers in this dissertation. Additionally, I would like to thank department staff – Debbie Allen, Kathy Malanoski, Denise Brubaker, and Michelle Brown – for their assistance in all administrative matters. I would also like to thank colleagues and former-colleagues in the department – Kaylyn Jackson-Shiff, Handi Li, Josh McCrain, Allison Cuttner, Dani Villa, and more – for their feedback on various aspects of the dissertation.

A dissertation is never completed in an academic vacuum (even in 2020!), and there are specific colleagues that I would like to acknowledge for both their feedback and (more importantly) their morale support. Marcella Morris was with me throughout my entire graduate career, whether I needed someone to bounce ideas off of, give feedback on specific submissions, or even just vent. During the pandemic, we got into the habit of running together, and that time has been helpful to re-center me, and allow me to vent existential frustrations in a healthy way. I am forever thankful for that time (and healthy outlet). Devon Thurman has always been a colleague that I enjoyed working with, and spending dedicated “non-work" times with. Finally Gray Barrett has been extremely helpful as I navigated career options throughout the pandemic. It was his guidance that allowed me to best present myself for a successful application to the State Department, and I am thankful for his support.

Finally, I would like to thank my lifetime partner, Ian Pavelich, for his love and support as we both navigated graduate school while maintaining a household. We have supported each other through the best and worst of times, and I am extremely grateful for his support

through this difficult chapter of my life. I have and will continue to grow while together with him, and am excited for the journey. Additionally, I would like to thank our corgi, Elton John. While an impulsive Covid-19 decision, raising Elton has required consistency that has been hard to find during the pandemic. And while he has not been the most helpful in terms of productivity, his personality has brought a lot of joy to my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

[A] body of men, holding them accountable to nobody, ought not to be trusted by anybody.

– Thomas Paine, *The Rights of Man*

Politics, defined most broadly, is concerned with the allocation and maintenance of power. However, to maintain the legitimacy of their power, those allocated power are to be held accountable for the manner in which they wield it. Yet accountability requires information regarding those who hold power, how they have used it, and how that comports with the original intent when such power was allocated in the first place.

This dissertation is concerned with the relationship between information and accountability in two contexts: electoral politics and bureaucratic politics. Its central aim is to provide unique findings regarding information and accountability in these sub-fields, while suggesting broader implications that could form the basis for further research. A survey experiment examines the types of information that voters prioritize when selecting a candidate. Formally modelling credit claiming behavior provides insights into when expectations of accountability might affect the presence of exaggerated claims. And developing a measure of bureaucratic responsiveness using information-sharing during appropriations paves a direct path for further study of bureaucratic accountability.

The studies in this dissertation build upon decades of prior academic work in this area. The general model for accountability, the principal-agent model, has been widely (but never exhaustively) studied over the past century (e.g. Moe 1984; Lupia et al. 1998; Achen and Bartels 2004; Healy, Malhotra et al. 2010; Huber, Hill and Lenz 2012; Gailmard 2012). In the electoral setting, on the voter's side, a large body of work has been focused on whether the voter can hold elected officials accountable (e.g. Campbell et al. 1980; Lupia 1994; Carpini and Keeter 1996; Converse 2006), or whether they must rely on heuristics to make a decision (Downs et al. 1957; Tversky and Kahneman 1974; Rahn 1993; Kirkland and Coppock 2018, e.g.). From the candidate's perspective, prior work has examined their incentive to credit claim (e.g. Mayhew 1974; Fenno 1978; Grimmer, Messing and Westwood 2012; Cain, Ferejohn and Fiorina 2013), and to a lesser extent the ability of challengers to rebut these claims (e.g. Shotts and Ashworth 2011; Ogden and Medina 2020; Dziuda and Howell 2021). In the bureaucratic context, a significant amount of scholarly work has been dedicated to conceptualizing responsiveness (e.g. Moe 1984; Saltzstein 1992; Durant 1995; Carpenter 2001; Krause 1999), as well as measuring it in certain domains (e.g. Weingast and Moran 1983; Wood and Waterman 1991; 1994; Lowande 2018*b*).

While the findings of each paper will be most directly felt within their respective domains, a couple of broader findings emerge through these studies. One consistent finding throughout this dissertation is that, while information is a prerequisite to accountability, more information does not equate to more accountability. Another broad, yet not specifically tested, result is that the ideal of coarseness of information should correspond to the coarseness of the accountability judgement. These broader findings present opportunities for further research into information and accountability in American politics.

1.1 Accountability Defined

In broadest terms, “accountability" is a function of two components: (i) the performance of some agent, and (ii) the ability of principals to judge that performance (Lupia et al. 1998; Achen and Bartels 2004; Healy, Malhotra et al. 2010; Huber, Hill and Lenz 2012; Gailmard 2012). The principal is an actor that delegates some political action to an “agent," and then judges that performance according to some criteria. This simple model can be applied to two areas that are the focus of this dissertation: electoral accountability and bureaucratic accountability.

In an electoral setting, the voters are the principal and the candidates are the agents. The most foundational models of this environment depict the incumbent as having exerted some effort or implemented some policy, and the voter either judges the act of the incumbent positively by re-electing them, or negatively by electing someone else (Fearon 1999). Focus of study in this environment typically centers on the candidate’s performance (or signal thereof), or the ability of voters to judge the incumbent. This dissertation similarly focuses on these dynamics of electoral accountability.

Meanwhile, in a bureaucratic setting, bureaucrats serve as the agents of some form of political principal – often Congress, the president, or the judiciary (Gailmard 2012). Although both actors are political in nature, the delegation of a tasks often takes place because bureaucrats possess specialized knowledge and/or capacity in order to implement broad policies set forth by political principles. Principals wish for implementation to match the preferences contained in their authorizing act, though that may not always be the case (Niskanen 1971*a*; Bendor and Meirowitz 2004). Thus, capturing the “responsiveness" of bureaucratic policy outcomes to the preferences of political principals is a central aim in studies of bureaucratic accountability.

1.2 Information Defined

“Information" in the field of political science can take on a myriad of meanings across different subfields. In formal models, information can refer to how certain or uncertain players are of other players’ payoffs or types (Harsanyi 1967; Gates and Humes 1997). In experimental work, information can refer to the directions or tasks given to subjects across treatment assignments (Druckman et al. 2006). Or even from the researcher’s perspective, information can refer to the data used to produce results that test hypotheses.

Couching “information" within the umbrella of “accountability" provides a useful tool to more narrowly define information. Accountability entails the performance of an agent and the ability of a principal to judge that performance. Each of these components requires information. The agent requires knowledge of the delegated task, the delegating principal, and their own preferences in order to perform; while the principal requires knowledge of the performance, the agent, and their own preferences in order to judge. “Information," then, constitutes the knowledge of any or all of these components.

To illustrate, let us return to both the electoral and bureaucratic contexts. In an electoral setting, information would refer to the voters and candidates, their preferences over outcomes, and any messages regarding these elements. In a bureaucratic setting, information would refer to knowledge of the agencies and their political principals, their preferences, as well as outputs by the agent or tasks delegated by the principal.

1.3 Papers in Summary

Three papers constitute this dissertation, and each studies information and accountability in a different context within American politics. Paper 1 and Paper 2 focus on electoral accountability through different perspectives – Paper 1 examines information and voter behavior in a non-strategic setting, while Paper 2 considers incumbent messaging in a strategic electoral setting. Meanwhile, Paper 3 considers accountability in bureaucratic politics

through the related concept of “responsiveness," and attempts to measure responsiveness across agencies and time. Given this piecemeal approach to examining information and accountability, each paper places their research contribution within the context of their own academic literatures. I provide a more detailed summary of each of these papers below.

1.3.1 Paper 1: Beyond Party ID? Information Environments and Voter Decisionmaking

The recent rise in polarization has raised legitimate questions about partisanship and accountability in American elections. Do Americans consider candidate preferences anymore when voting for candidates, or do they simply exhibit partisan loyalty? To examine this question, I conduct a conjoint and factorial survey experiment in which I present subjects with different information environments that vary whether partisan information is given, as well as the level of policy-based information that is available. The outcomes that I examine include voters' ability and willingness to choose according to their policy preferences, their dependence on heuristics and cues, intent to turnout, and satisfaction with the electoral process.

Research Questions

- Does the average American vote based on policy preferences or partisan identity?
- How does party ID perform as a heuristic for policy preferences?
- What effect does the information environment have on a voter's intent to turnout, and their satisfaction with the electoral process?

Summary of Findings

Through this study, I find that voters are responsive to policy-relevant information, and match their preferences to the stated preferences of candidates. Voters also rely less

on party ID when policy information is present, and express increased satisfaction with and willingness to turn out to elections with more policy-relevant information. However, perhaps paradoxically, in terms of choosing the candidate that matches on policy overall, partisan cues are just as effective as having complete information.

1.3.2 Paper 2: Electoral Accountability and Exaggerated Credit Claiming

Credit claiming is an essential tool for incumbents to signal their effectiveness to voters. However, as voters do not typically possess the political knowledge or levels of engagement to critically evaluate these claims, incumbents can choose to misrepresent their accomplishments to bolster their constituent support. Although successful challenges of these claims have been shown to reverse these effects, challengers do not actively employ this strategy in an election setting. Why is this the case? To examine this question, I develop a modified signalling game that represents an incumbent choosing to credit claim in an electoral setting, and examine the equilibrium results.

Research Questions

- Why is it rare for an Incumbent to be challenged when credit claiming in an election?
- How do media accountability and voter accountability affect the willingness of Incumbents to exaggerate their record?

Summary of Findings

Through this model of credit claiming in an election, I find that the equilibrium likelihood of ineffective incumbents being successfully challenged on exaggerated claims is lower than one might intuitively expect. Media accountability and voter accountability drive the equilibrium results of this model, and both need to be sufficiently high for incum-

bents to be dissuaded from exaggerating their credit claims. Interestingly, high levels of media accountability *diminish* the effects of high voter accountability in keeping incumbents honest in their messaging.

1.3.3 Paper 3: Measuring Bureaucratic Responsiveness Using Text from the Appropriations Process

The degree to which bureaucratic policy choices reflect the preferences of elected officials is a central question in the study of administrative politics in the United States. It has implications not just for understanding the content of policies but also the quality of democracy. However, measuring agency responsiveness in a consistent way across many agencies over time is a difficult endeavor. In this paper, we examine whether the annual appropriations process provides and avenue to measure bureaucratic responsiveness across agencies and time. This paper is co-authored with Alexander Bolton.

Research Questions

- How does information sharing during the appropriations process affect bureaucratic responsiveness?
- Can bureaucratic responsiveness be measured across agencies and time?

Summary of Findings

We use the approving and disapproving text in appropriations reports to characterize the degree to which Congress approves or disapproves of agency actions in any given year. The regularized nature of the appropriations process and the standard ways in which Congress comments on agencies allows for the comparison of responsiveness across agencies and tasks as well as within agencies over time. We validate this measure by comparing measures of non-responsiveness to the issuance rate of limitation riders, and find that this mea-

sure tracks well with this other well-studied measure of bureaucratic accountability. With the measure validated, we then apply this measure to to study how agency politicization impacts responsiveness to Congress.

1.4 Implications

The aim of this dissertation is to examine the relationship information and accountability in specific contexts within American politics. Paper 1 examines this relationship in the context of voters in elections, examining the information they use to elect candidates. Paper 2 flips the script, examining how the anticipation of various forms of electoral accountability affects the messaging of candidates in elections. Meanwhile, Paper 3 considers this topic in the realm of bureaucratic politics, examining how routinized information-sharing creates an avenue to examine accountability across the bureaucracy.

Given this piecemeal approach to examining “Information and Accountability" as a broad theme, the findings of these studies are most specific to their respective contexts. However, there are a couple of broader points that can be gleaned from these findings as a whole. For one, as the act of accountability becomes coarser (i.e. binary versus continuous), the principal is better off given more condensed information on which to act. In Paper 1, party ID performed just as well as complete policy information when a voter was asked to make a vote choice amongst two candidates. By contrast, in Paper 3, appropriations committees benefit from as much detailed information as possible in order to comment on agency performance and set their next fiscal year budget accordingly.

Another broader point is that information does not immediately lead to accountability. One major point of Paper 2 was to illustrate that, in a strategic environment, the expectation of different forms of accountability can affect the information that is provided in the first place. Even in a non-strategic environment, too much information can hinder accountability. One of the results from Paper 1 illustrated that providing too much policy information

results in voters being *less* able to vote according to their policy preferences.

It is important to note that these points are *gleaned* from results across contexts, and are not specifically verified by the results obtained in this dissertation. However, this provides fruitful avenues for future research, as well as a way to bring more interconnectedness to the broad discipline of political science.

Chapter 2

Beyond Party ID? Information Environments and Voter Decisionmaking

When George Washington left office as the first President of the United States, he delivered a farewell address that warned against the dangers of factions whose influence would replace national interests. If Americans were to vote merely by party loyalty, Washington feared it would foster a "spirit of revenge," and enable the rise of "cunning, ambitious, and unprincipled men" who would "usurp for themselves the reins of government; destroying afterwards the very engines, which have lifted them to unjust dominion" (Washington 1796).

Of course, despite Washington's fears, political parties developed, and have become the most prominent institution that governs function and dysfunction in American politics. In fact, as polarization has taken hold throughout the 20th and 21st centuries, party identification has taken precedence over elaborated policy stances when it comes to making political decisions (Poole and Rosenthal 1984; Abramowitz and Saunders 2008; Fiorina, Abrams et al. 2008; Layman, Carsey and Horowitz 2006; Abramowitz 2010; 2012). If Washing-

ton's fears are correct then, elections do not serve as a meaningful way to hold officials accountable, but rather a way to promote ambitious party loyalists into seats of power.

However, scholars still disagree about whether voters rely on policy preferences or party loyalty when making decisions. Some suggest that policy preferences and issue based policy competition explain electoral competition and voter behavior (e.g. Downs et al. 1957; Tomz and Van Houweling 2008). Meanwhile, others argue that issues are unimportant or even irrelevant to vote choice because party identification is the principle structure of competition in American politics (e.g. Kinder 1998; Polsby, Wildavsky and Hopkins 2008; Barber and Pope 2019). From another angle, political psychologists argue that party identification can serve as a stable and potent heuristic, or information shortcut, to guide their vote choice (Campbell et al. 1980; Rahn 1993; Green, Palmquist and Schickler 2004; Kirkland and Coppock 2018). Since Washington's fears were rooted in the effects of factions on American voters, and the subsequent lack of electoral accountability if party loyalty were the key driver, resolving this debate about the relationship between partisanship and voter behavior is important in understanding the state of electoral accountability in modern American politics.

This study seeks to provide more concrete evidence to this debate between issue-based and identity-based voting. More specifically, I set out to answer the following questions: (i) does the average voter decide based on policy preferences or partisan identities; (ii) for voters that use party ID as a heuristic, how does it perform as a shortcut to choosing based on policy preferences; and (iii) what effects does the information environment have on a voter's intent to turnout, and their satisfaction with the electoral process? To examine these questions, I conducted a conjoint, factorial survey experiment that varies the amount of information presented to individuals – both in terms of the number of policy stances given about candidates, as well as whether party identification is given. This design enables me to not only examine the effect of different levels of information on voter decision-making, but also to measure how subjects use other factors to choose candidates. Combined, this

design provides a more thorough and fine-grained examination of how different types of candidate information affect voters' decisions.

To preview the results, I find that voters are responsive to policy-relevant information, and match their preferences to the stated preferences of candidates. Voters also rely less on party ID when policy information is present, and express increased satisfaction with and willingness to turn out to elections with more policy-relevant information. However, perhaps paradoxically, in terms of choosing the candidate that matches on policy overall, partisan cues are just as effective as having complete information.

The rest of this paper will proceed as follows. First, I discuss prior literature on the topics of issue-based and identity-based political behavior, heuristics, and methods of assessing voter choice. Then, I develop simple models of voter behavior that distinguish issue-based voters from identity-based voters and create hypothesis for how exposure to policy-relevant information would affect their decisionmaking. Then I explain the experimental design used to assess these hypotheses. The results follow, and I conclude with answers to the research questions, and how these results contribute to the broader literatures on voter decisionmaking and elections as an accountability mechanism.

2.1 Literature on Information and Voter Decision-making

Before launching into the experiment that I conducted, I first review relevant literature in several areas: (i) policy, partisanship and political behavior; (ii) information processing and heuristics; and (iii) methodological approaches that have been used to study voter choice. Not only will such an examination place this study within the literature, but it will also inform the theoretical arguments that I will use to derive testable hypotheses about voter behavior.

2.1.1 Policy, Partisanship, and Political Behavior

Academics have long debated whether the average voter engages in issue-based politics or identity-based politics. Some suggest that policy preferences and issue based policy competition explain electoral competition and voter behavior (e.g. Downs et al. 1957; Key et al. 1966; Tomz and Van Houweling 2008). Under this conception of voter behavior, voters judge their policy preferences against those of the candidates, and select the candidates whose preferences more closely match theirs. Meanwhile, others argue that issues are unimportant or even irrelevant to vote choice because party identification is the principle structure of competition in American politics (e.g. Kinder 1998; Polsby, Wildavsky and Hopkins 2008; Barber and Pope 2019). This literature has shown that voter issue preferences are unstable over time and across issues, while party identification remains a stable trait that voters draw from when making their decision.

Other scholars note that the reason that this debate continues is due to issues of measurement (Hill 2022), and scholars have sought measurement techniques to further contribute to this debate. For example, when you average across issue preferences of voters rather than consider issues individually, voter issue preferences become more stable and predictive of vote choice (Ansolabehere, Rodden and Snyder 2008). Similarly, considering cases of shared partisanship between candidates and voters, Jessee (2012) finds that issue alignment is a strong predictor of vote choice.

This paper similarly contributes to this debate between issue-based or identity-based voting by incorporating more recent methodological advances in experimental research on voter behavior, which are explained below. But first, I touch on literature related to heuristics and decisionmaking, as it informs the theoretical distinction between issue-based and identity-based voters.

2.1.2 Information and Heuristics

Many democratic theorists, from Mill to Dahl, have considered an informed citizenry as key to the functioning democratic state. Not only do informed citizens participate more in civic society and better internalize democratic norms, but they also make more informed citizens, which would lead to more informed representation and policy Galston (2001). Yet, ever since political scientists began studying voter behavior, one trend among Americans has remained consistent. Generally speaking, citizens possess uneven and, on average, low levels of political knowledge and interest in political processes (Lazarsfeld, Berelson and Gaudet 1968; Campbell et al. 1980; Converse 2006; Carpini and Keeter 1996). However, this alone may not spell disaster for the proper functioning of elections. Although political knowledge remains generally low, heuristics can guide citizens' decision-making in the absence of more specific policy knowledge (Downs et al. 1957; Tversky and Kahneman 1974; Popkin and Popkin 1991; Lupia 1994; Althaus 2003). Heuristics are rules-of-thumb that can be applied to guide decision-making based on a more limited subset of the available information. For example, when deciding whether or not to bring an umbrella with me, I can look up at the sky rather than take the time to research weather forecasts. Because they rely on less information, heuristics are assumed to facilitate faster decision-making than strategies that require more information.

Of all the heuristics available, party identification has been shown to be the most prominent heuristic used by individuals (Rahn 1993; Kirkland and Coppock 2018). Since political parties operate on a platform that is a collection of ideological beliefs, identification with that group carries with it the association with that collection of beliefs (Popkin and Popkin 1991). For most voters, party identification is an important proxy for a candidate's issue stances and ideological leanings. Party identification is not the only heuristic available to individuals. Absent party ID, voters can look to other traits, such as race (Brady and Sniderman 1985; McDermott 1998), age (Lau and Redlawsk 2008), gender (Huddy and Terkildsen 1993; McDermott 1998), or incumbency status (Erikson 1971) in order to

make inferences about the quality and policy views of a candidate.

However, as individuals make inferences from heuristics, several things can go awry. For one, heuristics are subject to various biases (Tversky and Kahneman 1974; Lau and Redlawsk 2001; Huber, Hill and Lenz 2012; Dancey and Sheagley 2013; Petersen 2015), which can lead to worse inferences than would have been made had the individual had complete information. The reliance on heuristics over more specific information can also contribute to polarized views (Nicholson 2012), and even negative affect towards those that do not rely on the same heuristics (Brady and Sniderman 1985). Thus, while heuristics can help low-information individuals to make decisions, those decisions may or may not be the same ones that would have been chosen under complete information.

2.1.3 Methodological Approaches to Assessing Voter Choice

Before jumping to this theoretical discussion of voter decision-making, it would be sensible to briefly outline methodological innovations in studying voter choice, since they would inform the methodological approaches and estimation strategies used in this experimental design.

For most of the voter behavior literature, causal analysis has taken the form of randomized control trials, whereby different treatments are randomly assigned to subjects, and effects are estimated against a baseline control group. Such methods have been used to study effects such as social pressure on voter turnout (Gerber, Green and Larimer 2008), different forms of contact on voter turnout (Bergan et al. 2005; Gerber and Green 2005), and the thought processes that voters use when making a decision (Tomz and Van Houweling 2008). The leverage that a randomized experiment brings to the analysis is the ability to draw causal inferences with the satisfaction of fewer assumptions than are required of traditional econometric models (Angrist and Pischke 2008; Aronow and Miller 2015).

A factorial design adds on an additional layer of analysis by varying the level of treatment or combination of treatments provided to different subject groups. For example,

Panagopoulos, Larimer and Condon (2014) use a factorial design in order to examine the effects on turnout of receiving one of the following six conditions: no mailing, a mailing with individual vote history only, a mailing with individual vote history and a message emphasizing high (or low) community-level turnout from a previous election, and a mailing emphasizing high (or low) community-level turnout only. In setting up this design, the authors are able to assess not only the effects of a mailing with one's vote history, but also (i) the effects of receiving a message about community turnout, and (ii) the effects of receiving both of these messages. Such a method allows for the assessment of multiple hypotheses, but poses its own complications. For one, having so many groups for comparison presents the problem of multiple comparisons, whereby the baseline probability of a statistically significant result is biased upwards simply due to the number of analytical comparisons one is making. Additionally, clear specification of the comparisons is important, as deriving estimates in a factorial design can be more complicated.

Finally, conjoint analysis allows for the examination of multidimensional preferences, whereby a researcher can examine the salience of different dimensions under different treatments (Hainmueller, Hopkins and Yamamoto 2014). This method has been applied to several areas of voter behavior and preferences, including immigration preferences (Hainmueller and Hopkins 2015), complex policy preferences, and candidate preference (Hainmueller, Hopkins and Yamamoto 2014; Franchino and Zucchini 2015; Kirkland and Coppock 2018). For example, in a conjoint study, Kirkland and Coppock (2018) examine the effects of a nonpartisan electoral setting on the weights that voters place on various aspects of a candidate. They find that, in the absence of party labels, partisans of different affiliations value attributes differently – while Republicans value prior business experience, Democrats value prior political experience. The design of their study raises several considerations that must be taken into account when designing a conjoint experiment. For one, the order of profiles presented and the ordering of the information in those profiles matters. To minimize the threat of order effects biasing the results, it is best to randomize

the order with which both the profiles and the information in those profiles is presented (Hainmueller, Hopkins and Yamamoto 2014; Kirkland and Coppock 2018). Additionally, although conjoint designs are meant to study multidimensional preferences, there are limits to the numbers of attributes and dimensions that can be presented before subjects begin resorting to cognitive shortcuts. Kirkland and Coppock (2018) state that the consensus is that around six or seven attributes is the limit before subjects resort to shortcuts in information processing.

In conducting a randomized, factorial and conjoint experiment, it is necessary to take into account all of these considerations in the experimental setup and the data analysis. When appropriate, I will reference these considerations when explaining the experimental setup and the estimation strategies.

2.2 Theoretical Predictions

To explore these questions related to partisanship, information, and accountability, I first develop simple models of voting behavior. These will provide predictions for how subjects will behave when presented with different types and amounts of information about candidates. They will also provide a basis to assess how well party identification performs as a heuristic for policy information.

2.2.1 “Heuristic Voters" Versus “Identity Voters"

To answer the first research question, I distinguish a voter that uses partisanship as a heuristic from a voter that uses it as an identity trait. This can be accomplished with simple models of a voter’s utility calculus in each scenario. For ease of reference, I refer to these types of individuals as “Heuristic Voters" or “Identify Voters."

The Heuristic Voter

To start, consider a voter using party ID as a heuristic for policy preferences. Generally speaking, the voter in this case is attempting to select the candidate whose policy preferences are closer to their own. In simple spatial models (e.g. Downs et al. 1957; Enelow and Hinich 1984; Jessee 2012), the voter's utility function can be represented as follows:

$$U_V(x) = -\sum_{i=1}^{n}(x_i - v_i)^2$$

For n total issues, a voter is choosing a candidate X that minimizes the distance between their ideal point on each issue (v_i) and that of the candidate (x_i). However, making this calculation requires substantial information regarding the candidate, information that may be time-consuming or otherwise costly to obtain. Moreover, the utility calculation becomes more complicated as n grows larger. A voter may instead wish to rely on the candidate's party identification to serve as an informational shortcut for the candidate's policy preferences, particularly since a party's policy stances are more stable and publicly available. This calculation can be represented as follows:

$$U_V(x) = -(P(x) - v)^2$$

Here, $P(x)$ represents the platform of candidate x's party, while v represents an average of the voter's preferences over the platform issues. In this case, the voter chooses the candidate whose party's platform is closer to their preferences. This type of simplification is adopted from Downs (1957), where the platforms of parties and the preferences of voters are summarized by single points on a liberal-conservative scale, simplifying the decision-making procedure.[1]

[1]This simplistic model overlooks sources of error introduced by this substitution, though this is considered below when theorizing about heuristics and accountability.

Of course, a voter of this type may use some convex combination of these two methods. This would involve partitioning the n-dimensional policy space into a set of policies for which the voter uses the first equation, and a set for which they use the second. Indeed, to the extent that a candidate's policy stances are readily available, the voter could obtain a more precise measure of their utility. The important point is that the voter, intending to match along policy preferences, uses the platform of the candidate's party as a stand-in for some or all of the candidate's policy preferences.

The Identity Voter

Modeling the preferences of an individual that votes using party as an identity is a simpler task. The party identification of a candidate (P_x) and a voter (P_v) may be Democrat (D) or Republican (R). That is, $P_x = R$ or $P_x = D$, and similarly, $P_v = R$ or $P_v = D$. Then, the utility calculation for this type of voter is as follows:

$$U_V(x) = \begin{cases} 0 & \text{if } P_x = P_v \\ -1 & \text{if } P_x \neq P_v \end{cases}$$

This type of voter aims to vote for a candidate that shares their same party identification. Importantly, policy preferences are not a factor for consideration for this type of voter.

Distinguishing between Types

Given these models of how individuals can use party ID in their voting behavior, it is clear that the presence or absence of policy-relevant information would have different impacts on an average voter of each type. If candidate's policy stances were made freely available, then they would have an impact on the Heuristic Voter, and not the Identity Voter. More specifically, the behaviors that would be affected include voting along policy, voting by party ID, satisfaction in the voting process, and intent to turn out.

The relationship between voting in line with policy preferences (“voting along policy") and voter type follows directly from the simple models of each type. If the average voter is a Heuristic Voter, then policy-based voting should increase as a candidate’s policy stances are revealed, subject to limitations discussed in the next section. Meanwhile, if the average voter is an Identity Voter, then policy-based voting should remain unchanged, as this information does not factor into their utility calculus.

- *H1a: If the average voter is a Heuristic Voter, then voting along policy should increase when candidate preferences are revealed.*
- *H1b: If the average voter is an Identity Voter, then voting along policy should remain unchanged when candidate preferences are revealed.*

The relationship between voting in line with one’s party affiliation (“voting by PID") would move in the other direction. For the Heuristic Voter, the availability of a candidate’s policy preferences would mean that they would have to rely on party ID less as a heuristic. Thus, if the average voter is a Heuristic Voter, one would expect voting by PID to decrease once policy-relevant information becomes available. Meanwhile, the behavior for an Identity Voter would remain unchanged, as their entire utility is determined by voting with the candidate with the same party identification.

- *H2a: If the average voter is a Heuristic Voter, then voting by PID should decrease when candidate preferences are revealed.*
- *H2b: If the average voter is an Identity Voter, then voting by PID should remain unchanged when candidate preferences are revealed.*

When considering one’s satisfaction with the voting process, it is important to clarify what this behavior means relative to the theory. There are many external factors that might affect a voter’s satisfaction with the voting process, such as the weather that day, their interaction with the poll workers, or even the availability of mail-in or early voting (e.g.

Claassen et al. 2008; Gomez, Hansford and Krause 2007; Lago and Blais 2019). Within the scope of this theory, I focus on solely the effects of the information environment and voter satisfaction. Generally, a voter that has all of the information needed to make their decision will be more satisfied with the process. Thus, a Heuristic Voter's satisfaction would be affected by the presence of policy information, while that of an Identify Voter would not.

- *H3a: If the average voter is a Heuristic Voter, then their satisfaction with the voting process should increase as more policy preferences are revealed.*
- *H3b: If the average voter is an Identity Voter, then their satisfaction with the voting process should remain unchanged as more policy preferences are revealed.*

As with satisfaction, voter turnout has a long list of external factors that have all been studied in the literature (e.g. Gerber and Green 2000; Gerber, Green and Larimer 2008; Merrifield 1993). Again, I am most concerned with effects of the information environment on an individual's stated intent to turn out. Within this scope, a voter that has all of the information that they need, all else equal, are more likely to turn out to vote. Thus, a Heuristic Voter's intent to turnout would be affected by the presence of policy information, while that of an Identify Voter would not.

- *H4a: If the average voter is a Heuristic Voter, then their intent to turnout should increase as more policy preferences are revealed.*
- *H4b: If the average voter is an Identity Voter, then their intent to turnout should remain unchanged as more policy preferences are revealed.*

2.2.2 Partisan Information and Accountability

Recall that the second research question is conditional upon the average voter being of the "Heuristic Voter" type. As a result, this section further develops the constraints

that a “Heuristic Voter" faces, and compares their decision under those constraints to the limitations of relying on party ID as a heuristic.

One of the main reasons that a “Heuristic Voter" might rely on party ID is that the amount of information to process when making a decision. Prior research has acknowledged that the average capacity of working memory is limited (Miller 1956). Thus, a voter would face diminishing returns when presented with a quantity of information that exceeds their working memory capacity. A simplistic way of factoring this into their utility function would be as follows:

$$U_V(x) = \begin{cases} -\sum_{i=1}^{n}(x_i - v_i)^2 - (n-w)(c) & \text{if } n > w \\ -\sum_{i=1}^{n}(x_i - v_i)^2 & \text{otherwise} \end{cases}$$

Where w represents the working memory capacity of the individual, and c represents a fixed cost of incorporating additional information into their utility calculus. This factor elaborates on Hypothesis 1a by stating the relationship between policy voting and *increases* in policy information, rather than just the presence of it.

- *H5: Conditional upon policy information being present, the “Heuristic Voter" will vote along policy less as the amount of policy-relevant information increases.*

On the other hand, relying on heuristics has its own pitfalls. In being an informational shortcut for more nuanced or specialized information, heuristics can be subject to biases or errors (Tversky and Kahneman 1974; Lau and Redlawsk 2001; Huber, Hill and Lenz 2012; Dancey and Sheagley 2013; Petersen 2015). This would be even more true as a voter’s reliance on the heuristic increases. Given that a “Heuristic Voter" intends to vote in line with their policy preferences, an important measure of accountability would be the extent to which they can when they rely on party ID as a heuristic. An adjusted simple utility

function that accounts for these potential biases and errors would be as follows:

$$U_V(x) = -(P(x) - v)^2 - \varepsilon$$

Where ε represents the average distance between the candidate's preferences and that of their party's platform. The voter does not know the value of ε at the time of their decision-making. However, given their decision to rely on party ID as a heuristic, they assumed that it was less than the informational and computational costs of uncovering and factoring in candidate policy preferences. I am agnostic as to whether the heuristic error rate or the computational costs of the "Heuristic Voter" are higher, and thus the next hypothesis is simply an exploratory one.

- *H6: Party ID as a heuristic may or may not be as effective as having increased, or even complete, information on candidate policy preferences.*

2.3 Empirical Strategy

To assess these hypotheses, I conducted a conjoint and factorial survey experiment, designed through Qualtrics software and uploaded to Amazon's Mechanical Turk (mTurk) online task platform. In this section, I will discuss the design of this experiment, as well as the estimation strategies that will be used to assess each of the above hypotheses.

2.3.1 Description of Subjects

Using the mTurk platform, I obtained 1640 respondents, who were each compensated $1.10 for completion of the survey, which falls above the median level of compensation for mTurk's Human Intelligence Tasks (Hara et al. 2018). Institutional Review Board guidelines generally require that respondents' compensation not be conditioned on completion of the task, which I follow in the implementation of this survey.

Survey experiments sometimes face the flaw of having high internal validity but low external validity (Mutz 2011; Kirkland and Coppock 2018). In conducting the analysis, I am aware of and sympathetic to this concern, yet note the benefits of mTurk over other convenience or financially viable samples. The democratic characteristics of mTurk respondents has been studied extensively, and informs our expectations of the demographic characteristics of our subjects. E., Freese and Druckman (2016) find that, on average: 53.9% of mTurk users are male; 71.8% are white; the mean age is 31.6 years; 46.1% identify as Democrat, 14.6% as Republican, 34.3% as Independent, and 4.9% as "Other;" and 42.6% identify strongly with the Democratic or Republican party. Huff and Tingley (2015) conducted a comprehensive demographic analysis of mTurk respondents, finding that they are consistently similar to respondents of other nationally-representative surveys used in political science research, such as the Cooperative Congressional Election Survey. While Berinsky, Huber and Lenz (2012) found some statistically significant differences between the mTurk population and respondent pools of national surveys, they emphasize that the differences are substantively small, especially when compared to other samples available to political science researchers.

The demographic characteristics of the participants in this study largely fall in line with the trends above. Table 2.1 presents a summary of the proportion of respondents that have various demographic traits of interest.

Particularly in terms of race and gender, the demographics of this sample conform to the samples expected on the mTurk platform. The sample of this study seems slightly more representative along the lines of party ID, while being somewhat less representative in terms of age. Overall, the subject pool does not perfectly align to the US electorate as a whole, yet performs much better than convenience samples.

Trait	Number	Proportion
Gender		
Male	882	0.538
Female	756	0.461
Age		
Under 40	1066	0.65
40 to 50	249	0.152
Over 50	325	0.198
Party ID		
Republican	497	0.303
Democrat	885	0.54
Independent	248	0.151
Race		
White	1239	0.755
Black	178	0.109
Hispanic	56	0.034
Asian	132	0.08
Other	29	0.02

Table 2.1: Demographic Characteristics of Subjects

2.3.2 Details of the Experiment and Treatment Assignment

Prior to treatment assignment, subjects were asked a number of demographic questions in order to obtain covariates for balance testing and controls in the data analysis. Then, subjects were asked questions about their levels of political engagement and prior turnout in voting. After that, subjects were asked for their partisan identification, and their stances on five salient political issues: (1) immigration, (2) abortion, (3) gun control, (4) healthcare, and (5) tax policy. These issues were chosen because they are extremely salient issues in American politics that have been asked frequently by the American National Election Studies (ANES) Time Series Pre-election Questionnaire. In the survey, the questions and responses to these prompts are modelled after ANES, or taken directly from the 2016 ANES Questionnaire when possible, as a way to improve external validity. In so modelling the responses on ANES, I was also able to ensure that the response options were not open-ended, yet still allowed enough variation to get at different ideological stances on each issue. Subject responses to these prompts were coded from 1 to 4, where 1 represents the

most conservative response and 4 represents the most liberal response. The essential point of this step in the design was that I was able to obtain a vector of policy preferences for each respondent. Subjects were also asked how important they view each issue, which I use to weight the sum of the policy differences between them and the candidates.

After responding to these pre-treatment prompts, subjects underwent complete randomization into one of nine conditions. All of these conditions are the same in that subjects were then exposed to 15 pairs of candidates, and then asked to select the candidate that they would vote for in an election. The conditions vary in the amount of information that the subjects have about these pairs of candidates. This information varies in two dimensions: (1) whether or not subjects receive the party identification of the candidates (no PID or PID), and (2) the number of policy stances given about each candidate (zero, one, three, or all five). At the lowest level, subjects only received only demographic information about the candidates (age, gender, and race). This group served as the effective “control group" that was the baseline group for analysis. At the highest level, subjects received demographic information about the candidates, their party IDs, as well as the candidates' positions on the same five issue prompts that the subjects had previously answered before being assigned to treatment. Table 2.2 presents the distribution of subjects into eight of the experimental conditions.

	No PID	PID
+0 Policy Info	177	173
+1 Policy Info	177	175
+3 Policy Info	177	174
+5 Policy Info	176	176

Table 2.2: Summary of Treatment Assignment

In addition to these eight experimental conditions, I also created a placebo treatment group, to which 176 subjects were assigned. In addition to the demographic information, these subjects were given one policy stance of the candidates, as well as two random “fun facts" that are not connected to policy preferences in any meaningful way. The purpose of

this placebo group, which will be discussed at length in the data analysis, is to ascertain the extent to which treatment effects are due to the presence of “policy-relevant" information versus simply receiving more information about the candidates in general.

Figure 2.1 presents a depiction of the information that subjects saw across select treatment groups. The upper left image shows the information that the “No PID, No Policy Info" group received about each pair of candidates, while the upper right image shows the information that the “No PID, +1 Policy Info" group received. Meanwhile, the bottom left image show the information that the “No PID, +5 Policy Info" group received, and the bottom right image shows the information that the “PID, +5 Policy Info" group received.

With these different treatments in mind, it is worth discussing how these profiles of the candidates were created and presented in the study. Thirty candidates were randomly assigned demographic characteristics and preferences across the five issues in R. Then, an “average issue score" was computed for each candidate by calculating the mean of the 1-4 scores across the five issue areas. The party ID of each candidate was then determined by a cutoff of these “average issue scores" – candidates with scores above a 2.5 were labelled “Democrat," while candidates with scores below 2.5 were labelled “Republican." This is to ensure that party ID is correlated with, yet not necessarily fully predictive of, their policy positions. Using randomly generated candidate profiles ensured that there were both moderate and extreme candidates in both parties across the thirty candidates.

With these candidate profiles created, they were then paired randomly such that each of the 15 pairs featured one Democrat and one Republican. To determine which policy stances were featured across the conditions, I started with the full information condition (“PID, +5 Policy Info"), and for each pair in each condition, randomly (selected at random in R) deleted one piece of information from the candidate profile so that each pair featured different pieces of policy information. This process was repeated iteratively until the profiles contained only the demographic information (for the “No PID, No Policy Info" condition). This process ensures that across the 15 pairs, subjects that do not receive full

	Candidate A	Candidate B
Gender	Male	Male
Race	Hispanic	Asian
Age	35	55

Which candidate would you vote for?
- Candidate A
- Candidate B

	Candidate A	Candidate B
Position on Tax Policy	"Government should tax only to raise money for its essential functions, such as enforcing contracts, maintaining basic infrastructure and national security, and protecting citizens against criminals."	"Government should tax to raise money for its functions, citizen services, and to provide a universal basic income."
Gender	Male	Male
Race	Hispanic	Asian
Age	35	55

Which candidate would you vote for?
- Candidate A
- Candidate B

	Candidate A	Candidate B
Position on Tax Policy	"Government should tax only to raise money for its essential functions, such as enforcing contracts, maintaining basic infrastructure and national security, and protecting citizens against criminals."	"Government should tax to raise money for its functions, citizen services, and to provide a universal basic income."
Gender	Male	Male
Race	Hispanic	Asian
Position on Abortion	"By law, abortion should never be permitted."	"The law should permit abortion for reasons other than rape, incest, or danger to the woman's life, but only after the need for the abortion has clearly been established."
Position on Healthcare	"Healthcare should be managed by the private marketplace, except for citizens that qualify for Medicaid or Medicare."	"Healthcare should be managed by the private marketplace, except for citizens that qualify for Medicaid or Medicare."
Position on Gun Ownership	"Any adult citizen should be able to purchase a firearm in any state, without restrictions."	"Any adult citizen wishing to purchase a firearm in the US should undergo and pass rigorous background checks, wait periods, and certified gun safety training; and the government should maintain a database to track gun sales and ownership."
Age	35	55
Position on Immigration	"Allow unauthorized immigrants to remain in the United States and eventually qualify for U.S. citizenship, but only if they meet certain requirements (paying back taxes, learning English, and passing background checks)."	"Allow unauthorized immigrants to remain in the United States and eventually qualify for U.S. citizenship, but only if they meet certain requirements (paying back taxes, learning English, and passing background checks)."

Which candidate would you vote for?
- Candidate A
- Candidate B

	Candidate A	Candidate B
Position on Tax Policy	"Government should tax only to raise money for its essential functions, such as enforcing contracts, maintaining basic infrastructure and national security, and protecting citizens against criminals."	"Government should tax to raise money for its functions, citizen services, and to provide a universal basic income."
Gender	Male	Male
Race	Hispanic	Asian
Position on Abortion	"By law, abortion should never be permitted."	"The law should permit abortion for reasons other than rape, incest, or danger to the woman's life, but only after the need for the abortion has clearly been established."
Position on Healthcare	"Healthcare should be managed by the private marketplace, except for citizens that qualify for Medicaid or Medicare."	"Healthcare should be managed by the private marketplace, except for citizens that qualify for Medicaid or Medicare."
Position on Gun Ownership	"Any adult citizen should be able to purchase a firearm in any state, without restrictions."	"Any adult citizen wishing to purchase a firearm in the US should undergo and pass rigorous background checks, wait periods, and certified gun safety training; and the government should maintain a database to track gun sales and ownership."
Age	35	55
Position on Immigration	"Allow unauthorized immigrants to remain in the United States and eventually qualify for U.S. citizenship, but only if they meet certain requirements (paying back taxes, learning English, and passing background checks)."	"Allow unauthorized immigrants to remain in the United States and eventually qualify for U.S. citizenship, but only if they meet certain requirements (paying back taxes, learning English, and passing background checks)."
Party Identification	Republican	Democrat

Which candidate would you vote for?
- Candidate A
- Candidate B

Figure 2.1: Example Experimental Stimuli for Select Treatment Groups

information are exposed to a variety of information about the candidates.

To mitigate the concern of order effects, the 15 pairs of candidates were presented to the subjects in a random order. Additionally, the ordering of the information on the candidate profiles was randomized to ensure that subjects did not over-weight the first few pieces of information presented to them, in a manner similar to Hainmueller, Hopkins and Yamamoto (2014) and Kirkland and Coppock (2018). The code used to generate the candidate profiles, their pairing, and the ordering of pairs and information, is available in the Supplementary Materials.

As mentioned above, for each of the 15 pairs of candidates presented, subjects were asked to select the candidate that they would vote for in an election. Based on their choice, the corresponding candidate number was recorded to streamline analysis. After their responses to the 15 pairs of candidates, all subjects were asked three post-treatment questions. First, they were asked, on a scale of 0-10 (where 0 = not at all confidence, and 10 = extremely confident), how confident they felt in their choices. Next, the subjects were asked, on a scale of 0-10 (where 0 = not at all satisfied, and 10 = extremely satisfied), how satisfied they would be with elections that provided the amount of information that they received in their treatment assignment. Finally, subjects were asked, on a scale of -5 to 5 (where -5 = much less likely, and 5 = much more likely), how much more or less likely they would turn out to vote if elections featured the amount of information that they had received in their treatment assignment.

Following these post-treatment questions, subjects were taken to a page with a brief survey debrief, as well as a randomly generated number that they then input into mTurk to receive compensation. Upon continuing from this point, the subject responses are recorded, as the time taken by the subject to complete the study.

Of the 1640 subjects that opted into the survey, 1581 observations were kept for analysis. Subjects were dropped from analysis for one of two reasons: (1) they did not answer any of the electoral choice questions, and (2) their survey duration was longer than three

hours, suggesting their survey had timed out (since data collection was completed in that amount of time). Appendix A.1 goes into more detail about data missingness, ensuring that missingness was not correlated with treatment assignment.

2.3.3 Estimation of Treatment Effects

Before analyzing the data, I conducted a balance test across pre-treatment covariates in order to assess the success of the randomization procedure. Importantly, a successful balance test provides evidence that the Conditional Independence Assumption is satisfied, allowing the results to have a causal interpretation (Angrist and Pischke 2008; Aronow and Miller 2015). The extent to which the treatment groups are balanced also determines whether adding controls to the model specifications will increase the precision of treatment effect estimates. The results of this balance test are included in Appendix A.2.

In order to assess H1 (Voting Along Policy), I first determined, for each selection that a subject made across a candidate pair, whether they selected the candidate that is more aligned to their policy preferences in the policy areas given. For example, if a subject is presented with stances on immigration and gun control, did they choose the candidate that is closer to them on those two issues? This was achieved by comparing the subject's policy vector to that of the candidate that they did support and did not support. If weighted sum of the distances between the subject's policy vector and their chosen candidate's vector was lower than that of their vector and the candidate not chosen, the entry was coded as a "1" in the dataset; "0" otherwise.[2] Based on this scoring, I was able to then calculate the proportion of times that the subject had voted in line with their stated policy preferences. Note that for the "No PID + 0" and "PID + 0" groups, it would be impossible for a subject

[2]This process was automated using R, and the code for this data cleaning is included in the supplementary materials. Additionally, If the distance between the subject was identical for both candidates in a pair, they were given a "1" in either case. If the subject did not select a candidate, they were coded as a "0" for that selection. Finally, if the subject did not list a preference on an issue, that issue was excluded when calculating distances.

to vote along the given policy information because no such information was given.

From here, in order to estimate the effects of treatment assignment on the likelihood of voting in line with policy preferences, I ran the following OLS model:[3]

$$prop.policy = \beta_0 + \beta_1 * Treat + \beta * \mathbf{X} + \varepsilon$$

X in this case is a matrix of pre-treatment covariates (education, age, gender, race, party ID). This model was run six times in total, with the "No PID + 0" group serving as the comparison group.[4]

To assess H2 (Voting by PID), I estimate the effect of treatment assignments on the proportion of times that a subject votes in line with their party ID. The model specification in this case is as follows:

$$prop.PID = \beta_0 + \beta_1 * Treat + \beta * \mathbf{X} + \varepsilon$$

This model was run against the other treatments that featured party ID, three times in total. The "PID + 0" group served as the comparison group for this analysis.

To assess H3 (Satisfaction) and H4 (Turnout), I similarly run a linear model that assesses the effect of treatment assignment on subjects' reported change in satisfaction and intent to turn-out, respectively. These outcome variables are on different scales – 0 to 10 and -5 to 5 – so I ensure to report the baselines so as to speak to the substantive impact of any significant effects.

Assessing Hypothesis 5 (Vote Along Policy as Information Increases) can be accomplished using the same models for Hypothesis 1. However, instead of examining the overall

---

[3]In Appendix A.3, I show that the results are robust to a logistic regression model specification.

[4]The "PID + 0" group could have just as easily served as the comparison group here, as they also received no policy information.

proportion of times that a subject votes in line with their policy preferences for the given profiles, I examine the change of that proportion as the amount of information increases. Specifically, if the proportion of "Vote by Policy" has a statistically significant decrease as the number of policy stances across profiles increases, I would conclude that the data supports the hypothesis.

Finally, Hypothesis 6 (Heuristic Performance) is assessed using the same set-up as the models for Hypothesis 1, but just for subjects that received the candidates' party identification. However, instead of determining whether a subject voted in line with the policy preferences given, I measure whether the subject voted in line with all of the policy preferences that the candidate had, whether they were shown or not. This adequately assessed the performance of party ID as a heuristic, since the candidate's party ID was determined using a cutoff rule for the average of their policy preferences.[5]

Given all of the models specified above, as well as the number treatments to compare against the control groups, the number of coefficient estimates and standard errors that this analysis will render will be quite large. For clarity of presentation, I present the results as coefficient plots, and include the associated regression tables in Appendix A.4.

2.4 Results

Given the estimation strategies outlined above, I now present the results of the experiment in the sections that follow. I also include the result of a placebo analysis to ensure that it is the policy-relevance of the information that is driving the results, and not just the amount of information overall.

[5]The cutoff rule is the most conservative one possible (assigning Democrat as just left of center and Republican as just right), so the estimates from the data analysis will be conservative in nature. It would be reasonable to suspect that heuristic performance will increase as the cutoff rule is more aggressive, which would signify more polarized parties. This would be an interesting avenue for further research.

2.4.1 Hypothesis 1 - Voting Along Policy

Recall that H1 (Voting Along Policy) makes predictions for the behavior of a Heuristic Voter versus an Identity Voter. If the average Voter is a Heuristic Voter, then voting along policy should increase when candidate preferences are revealed. But if the average voter is an Identity Voter, then voting along policy should remain unchanged.

The experimental results related to this hypothesis are presented in Figure 2.2, and the exact coefficient estimates and standard errors for the models are included in Appendix A.4.

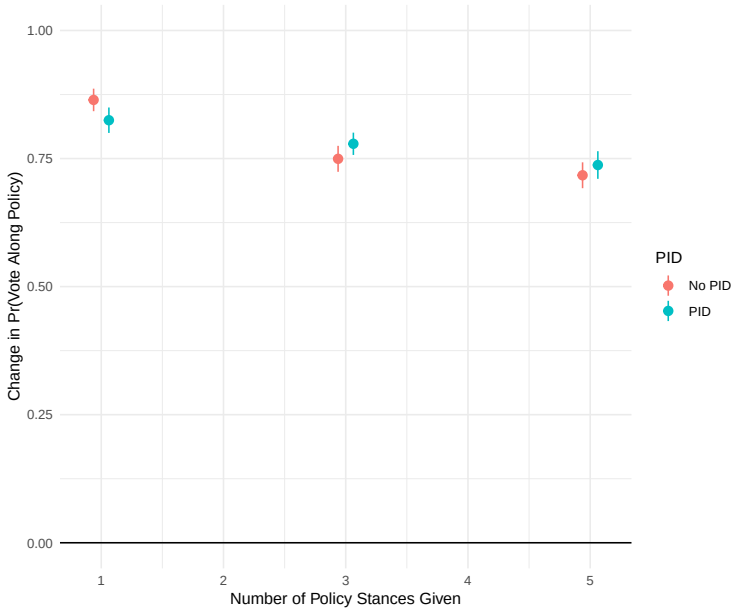


Figure 2.2: Change in Policy Voting When Policy-Relevant Information is Present

The results presented here overwhelmingly support the idea that the average voter is a Heuristic Voter. When voters are presented with policy-relevant information, they are anywhere from 71.7% to 86.4% likely to choose the candidate that was closer to them amongst the policies given.[6] The data here suggests that the average voter is responsive to knowing the policy preferences of candidates, and votes so as to match their policy

[6]Note that the control group likelihood is zero, since they were given no policy-relevant information to select on.

preferences to that of the candidate's.

2.4.2 Hypothesis 2 - Voting Along Party ID

The second hypothesis concerns the extent to which party ID becomes more or less salient in candidate choice as subjects receive more policy-relevant information. If the average voter is a Heuristic Voter, then voting along party ID should decrease once policy preferences are revealed. If the average voter is an Identity Voter, then this value should remain unchanged. The experimental results for partisan voting are presented in Figure 2.3, and the exact coefficient estimates and standard errors for the models are included in Appendix A.4.

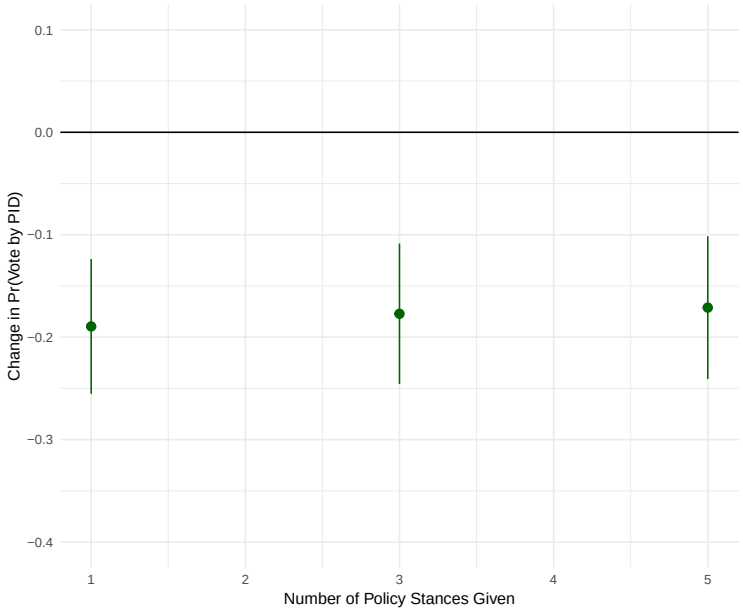


Figure 2.3: Change in Voting by PID When Policy-Relevant Information is Present

When compared to the "PID + 0" condition, we see that subjects in conditions with party ID and policy-relevant information prioritize party ID less when making their decision between candidates. Specifically, we see a reduction in anywhere from 17.1% and 18.9% in the frequency of voting along party ID, which supports the average voter being a Heuristic Voter. Given the intensity of partisanship in contemporary American politics,

these levels are particularly striking. This effect is statistically consistent amongst the treatment groups that received any level of policy-relevant information.[7]

2.4.3 Hypothesis 3 - Satisfaction

The next two hypotheses are concerned with the effects of treatment assignment on responses given at the end of the survey. Hypothesis 3 assesses the levels of satisfaction that subjects would have with electoral processes that provided the same amount of information that they had in the study. More precisely the wording of the post-treatment question was as follows: “On a scale of 0-10, how satisfied would you be with elections that provided the level of information that you received in this study?" Subjects were then given a scale from 0 to 10 to score their satisfaction, where 0 meant “Not at all Satisfied" and 10 meant "Extremely Satisfied." As mentioned above, H4 (Satisfaction) predicts that as subjects receive more information about candidates in elections, they would be more satisfied with the electoral process.

The experimental results for satisfaction are presented in Figure 2.4, and the exact coefficient estimates and standard errors for the models are included in Appendix A.4.

In this case, the data strongly supports the idea that the average voter is a Heuristic Voter. Statistically significant effects appeared with the presence of just party ID, but increased and became even more statistically distinguishable as subjects received more policy-relevant information. Starting with just being given party ID, stated likelihood of turnout increased by 1.09 points on a 0 to 10 scale. These effects increased to 2.83 points and 3.12 points for the “No PID, +5 Policy Info" and “PID, +5 Policy Info" groups respectively. Given a baseline score of 4.023 for the “No PID + 0" control group, it is clear that

[7]The change in reduction from the +1 to the +5 group, combined with the results in the H1 analysis, might suggest a switch to using party ID in the +5 group. This conjecture is inconclusive, though, due to the inability to statistically distinguish the groups in these H2 results.

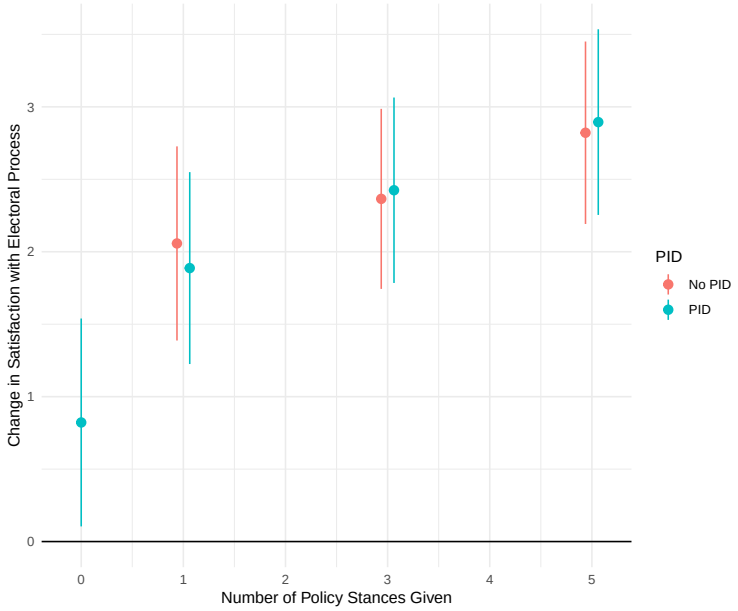


Figure 2.4: Change in Satisfaction by Treatment Group

just having party ID increased satisfaction to neutral, yet having policy-relevant information increased satisfaction to a more positive score on this scale. Thus, a voter's satisfaction with the voting process depends on having a candidate's policy preferences, which is information that a Heuristic Voter needs to make their decision.

2.4.4 Hypothesis 4 - Turnout

Hypothesis 4 considers the effects of the information environment on stated turnout. At the end of the survey, subjects were asked, "If elections featured the information that you received in this study, how much more or less likely would you be to turn out to vote?" Subjects were then given a scale from -5 to 5 to score their responses, where -5 meant "Much Less Likely" and 5 meant "Much More Likely." As mentioned above, H4 (Turnout) predicts that if the average Voter is a Heuristic Voter (Identity Voter), then intent to turnout should increase (remain unchanged) as more policy preferences are revealed.

The experimental results for turnout are presented in Figure 2.5, and the exact coefficient estimates and standard errors for the models are included in Appendix A.4.

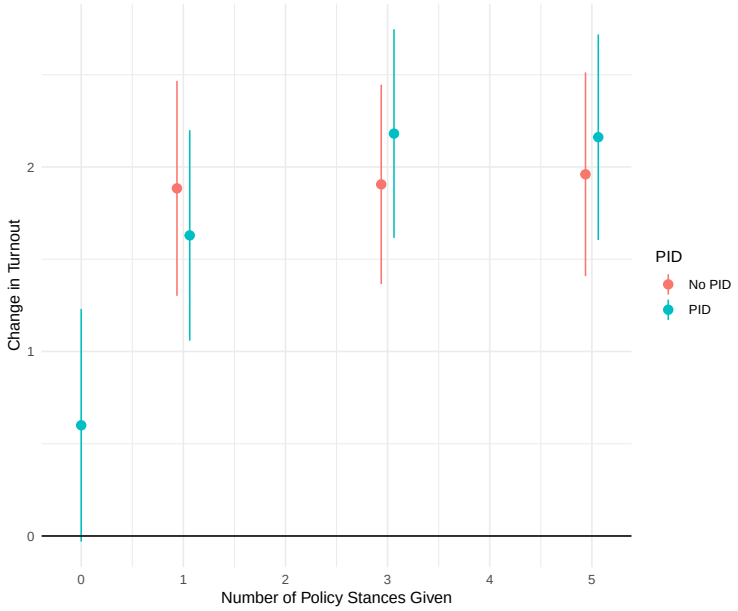


Figure 2.5: Change in Stated Turnout by Treatment Group

Overall, the data support the average voter being a Heuristic Voter. While just being given party ID has no statistically significant effect on the stated likelihood of turnout, being given policy-relevant information has a positive and statistically significant effect on stated turnout. While estimates do increase as more policy information is given, the increases are not enough to be statistically distinguishable from one another. The average score amongst the control group was -0.099, thus an increase of approximately 1.8 to 2.3 points on this scale is substantively large as well. These effects are also robust to including pre-treatment turnout as a covariate.

One may wonder whether social desirability bias might play a role in the higher turnout estimates here. This concern is mitigated in two ways. First, subjects were asked about their rates of turnout prior to treatment assignment, in a way that acknowledges the reality of constrained time impacting turnout. More importantly, assuming that randomization worked and that the excludability assumption holds, which are fundamental in experimental designs, means that the differences reported here are purely a result of treatment assignment. Thus, overall, it appears that increases in policy-relevant information cause increases in stated likelihood of turnout out to vote.

2.4.5 Hypothesis 5 - Vote Along Policy as Information Increases

The final two hypotheses are conditional upon the average voter being a Heuristic Voter, which has seen overwhelming support from the results thus far. Hypothesis 5 assesses whether voting along policy decreases as the amount of information increases. Put more plainly, do voters face difficulty processing larger amounts of information to vote along their own policy preferences?

Referring back to Figure 2.2, we see that policy voting decreases as the amount of information to process increases. The +1 groups are statistically distinguishable from the +3 groups at the $p < 0.1$ level,[8] and are distinguishable from the +5 groups at the $p < 0.05$ level. Although not resounding support, I nonetheless find decent support that, in line with predictions, subjects are less able to vote in line with policy preferences when there is too much information to process.

2.4.6 Hypothesis 6 - Heuristic Performance

Finally, Hypothesis posits an exploration of the performance of party ID as a heuristic for policy preferences. How does choosing based on just party ID compare to choosing based on incomplete, or even complete, policy information? Recall that to assess this, subjects were coded as having “voted along policy" if they voted for the candidate that more closely matched along all possible policy preferences, not just the ones displayed for the subject. Figure 2.6 presents the results of this comparison, with the estimates and standard errors being reported in Appendix A.4.

In this figure, the performance of each treatment group is being compared to that of the “PID + 0" group. We see that, for low levels of policy-relevant information, subjects perform worse than just having party ID to use while voting. With higher levels of policy-

[8]These groups are *almost* distinguishable at the $p < 0.05$ level, but the confidence intervals overlap by 0.000586 units.

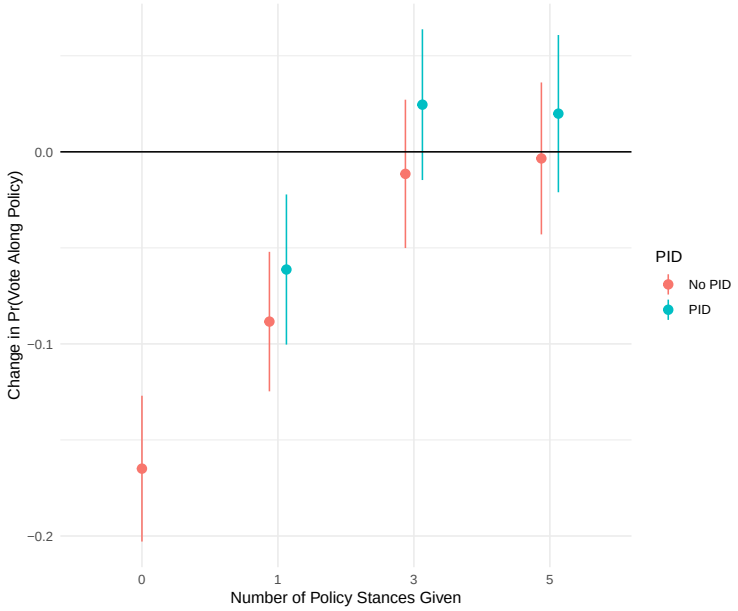


Figure 2.6: Performance of Party ID vs. Policy Stances

relevant information, and even complete information, subjects perform no better than just having partisan information. This is particularly surprising in the case of the +5 groups, as they had complete policy information about the candidates. Overall, these results suggest that party ID is indeed a useful and stable heuristic for policy preferences.

2.4.7 Placebo Analysis

Throughout this analysis, one may wonder whether the policy-relevance of the information is important, particularly when it comes to policy-based voting. Do voters filter through policy-irrelevant information to locate policy-relevant information when making their vote choice?

To test this theory, I included a placebo treatment group in the study. For each candidate pair, subjects in this group received the following information: (1) demographic information (age, gender, race), (2) one random policy stance, and (3) two randomized “fun facts" about the candidate. The “fun fact" statements were designed to suggest individual hobbies and experiences in a way that would not correlate to policy preferences in a meaningful

way.[9] Figure 2.7 below presents an example of what one such candidate pair looks like.

	Candidate A	Candidate B
Position on Tax Policy	"Government should tax only to raise money for its essential functions, such as enforcing contracts, maintaining basic infrastructure and national security, and protecting citizens against criminals."	"Government should tax to raise money for its functions, citizen services, and to provide a universal basic income."
Gender	Male	Male
Race	Hispanic	Asian
Interesting Fact	Owns a golden retriever	Has travelled on the Orient Express
Age	35	55
Interesting Fact	Has visited twenty-five countries	Played baseball as a child

Which candidate would you vote for?

○ Candidate A

○ Candidate B

Figure 2.7: Example Experimental Stimuli for Placebo Group

As with the other candidate profiles and pairings, the information was randomly determined for each candidate, and the ordering of candidate pairs and information for each candidate was randomized to minimize the threat of order effects. Thus, the question becomes whether this group is comparable to the "No PID + 1" group, the "No PID + 3" group, or neither. If the answer is the first group, then that would suggest that it is the policy relevance of the information that was driving the above results. If the answer is the second group, then it might be the case that it is the amount of information that matters. In cases that the answer is neither, it might be the case that nature of the "fun facts" are affecting voter choices.[10]

[9]The Supplementary Materials include all of the "fun fact" statements that were possible for selection, as well as the code used to randomly assign these statements to the candidates.

[10]One may well argue that these "fun facts" prime other reasons that voters might choose certain candidates, such as their charisma, likeability, or just as a matter of expression.

Figure 2.8 presents a comparison of the placebo group to the “No PID + 1" and “No PID +3" groups in terms of voting along the policy preferences given (H1).

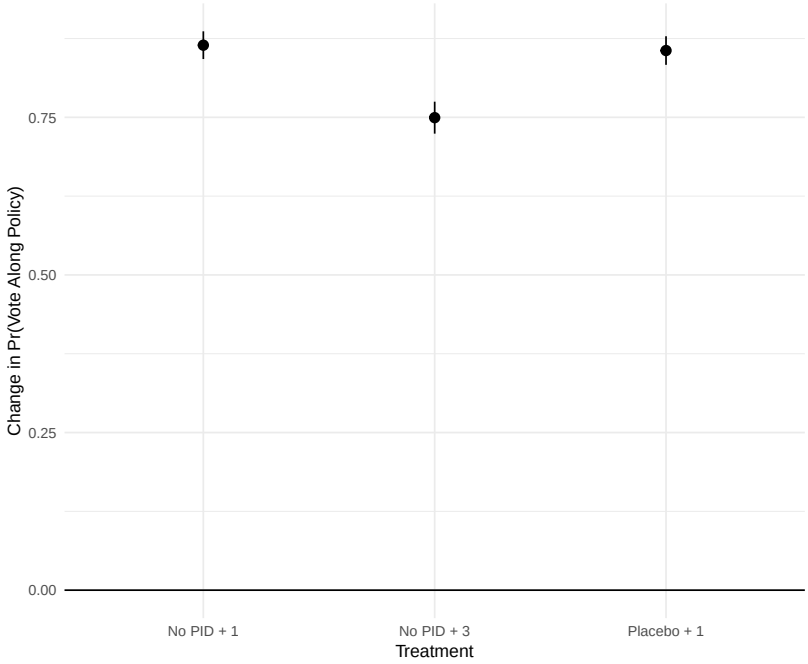


Figure 2.8: Placebo Group Analysis for H1

In terms of the point estimates, the placebo group is closer to that of the “No PID +1" and statistically distinguishable from the “No PID + 3" group. This seems to suggest that voters do filter through information and choose according to policy-relevant information. Additionally, the fact that the placebo group is statistically indistinguishable from the “No PID + 1" group suggests that the irrelevant information did not hinder subjects’ ability to perform a policy proximity estimation.

In terms of matching along policy preferences overall, Figure 2.9 provides a comparison between the three groups’ performance in that area.

In this case, the placebo group performed similar to that of the “No PID + 1" group, and is statistically distinguishable from the performance of the “No PID + 3" group. This also

---

These factors lie outside the scope of analysis. However, as will be discussed in the conclusion, the general methods employed by this paper provide an avenue to study these forms of voting as well.

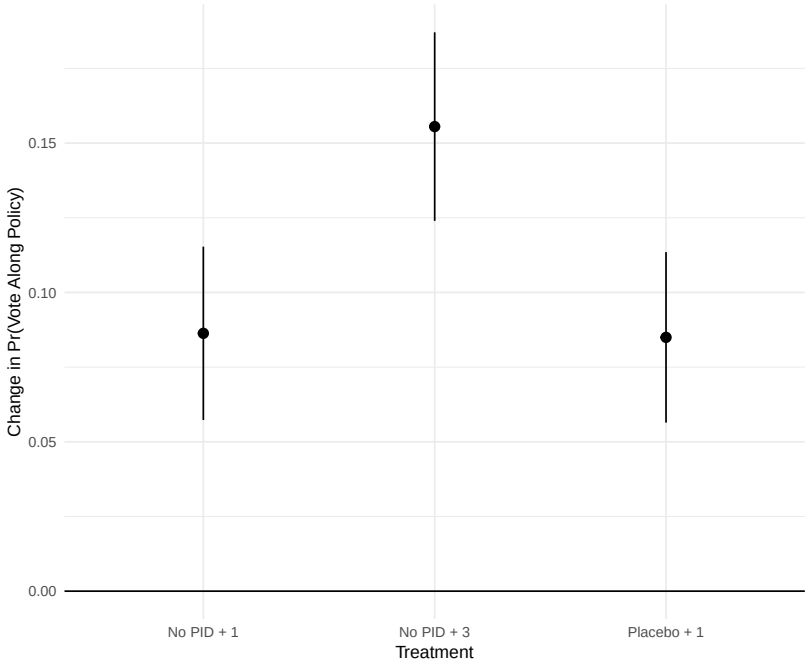


Figure 2.9: Placebo Group Analysis for Overall Vote Performance

confirms that fact that the policy-relevance of the information is driving the vote choice.

2.5 Discussion & Conclusion

From this conjoint and factorial survey experiment, we have learned a lot regarding how voters make decisions in different information environments. For one, this study suggests that George Washington's concern about the threat of factions is not so dire. On average, voters respond to policy-relevant information when it is given to them, voting for the candidates that more closely match those preferences. Voters are also prioritize party identification less when given policy-relevant information. Yet paradoxically, this might be to their detriment, since in line with past studies on party ID as a useful heuristic (e.g. Rahn 1993; Green, Palmquist and Schickler 2004), I find that party ID as a heuristic performs just as well for voters as having the policy stances given to them.

This study also made contributions to our understandings of voter turnout and satisfaction with our electoral institutions. Namely, as voters are provided with more information

about their candidates, they are more satisfied with the elections and are more likely to turn out. The question remains here as to who is responsible for acquiring or providing this information – is this an institutional issue, or an area in which individuals should take responsibility?

Taken together, these findings have many implications for policymakers and scholars alike. For policymakers, the results of this study would inform campaign strategies that could increase turnout, electoral satisfaction and institutional legitimacy, and policy-based voting. It also raises the question of whether institutions that provide candidate information might improve our elections. Indeed, several organizations, such as Vote411, HeadCount, Vote Smart, and Vote-USA, are attempting to fill that role. However, the effectiveness of these organizations will likely depend on the extent to which candidates genuinely reveal their policy preferences. It could be the case that candidates strategically adapt to these new institutional norms, causing voters to be no better off than before.

The results of this study also speak to various literatures within the discipline of political science. Relative to the polarization literature, the results of this study imply that focusing on policy preferences, rather than partisanship, might decrease the level of affective polarization in American society. Whether or not information will reduce ideological polarization depends on which theory of polarization one considers more valid. If polarization occurs only among the elites (Fiorina et al. 2008), then it would appear that increasing the flows of policy-relevant information might assuage the polarization problem. However, if one supports the idea that ideological polarization is truly prevalent amongst the mass public (Abramowitz 2012), then more information might only serve to reinforce these polarized beliefs.

Relative to the voter turnout literature, the results of this study contributes meaningfully to our understanding of the effects of information on voter turnout without raising ethical concerns.[11] It would be desirable, however, to consider ways to bolster the external validity

[11]We are of course reminded of the “Rock the Vote" experiment that caused controversy

of an experiment such as this one. Repeating the experiment on a more nationally representative platform, such as YouGov, would be one step forward. Another possible alternative might involve constructing candidate profiles that directly align with real candidates. Precautions would have to be taken to ensure that candidate selection effects do not introduce bias into the results.

The results of this study also speak to the broader literature on democratic theory and institutional legitimacy. Since it has been shown the more information increases the satisfaction with electoral institutions, then it may be worth considering how to shape electoral institutions so that they prioritize informing voters. As suggested in the policy implications above, though, care would need to be taken to anticipate strategic adaptation and manipulation by the institutional reformers.

The use of a conjoint and factorial design also provides a methodological contribution to the discipline. While these techniques have been used separately, their combination is a more novel concept. This combination could be used to study other aspects of voter behavior, including voters' relative weighing of experience versus policy congruence, or how the prioritization of partisanship might change depending on the electoral setting, just to name a few.

---

in Montana during the 2014 election.

Chapter 3

Electoral Accountability and Exaggerated Credit Claiming

Political representation and accountability rest on the ability of constituents to hold elected officials accountable for their actions while in office. However, voters often possess low levels of knowledge on political activities (Lazarsfeld, Berelson and Gaudet 1968; Campbell et al. 1980; Converse 2006; Carpini and Keeter 1996; Bickers and Stein 1996), and instead rely on messages and signals about the incumbent's effectiveness and accomplishments while in office when deciding whether or not to re-elect them (e.g. Stein and Bickers 1994; Evans 1994; Grimmer, Messing and Westwood 2012).

One of the most common ways that an incumbent communicates their accomplishments to constituents is through credit claiming, whereby they act "so as to generate a belief" that they were effective while in office (Mayhew 1974). In fact, in an electoral setting, the ability of an incumbent to message their past record of accomplishments through credit claiming is a strategy that is uniquely available to them.[1] A large literature has also shown that incumbents often utilize this strategy to bolster constituent support (e.g. Cain, Ferejohn

[1]Of course, challengers from different offices might very well point to their past records, but the alignment of this record would still not match that of an incumbent.

and Fiorina 2013; Ferejohn 1974; Fenno 1978; Lazarus and Reilly 2010; Mayhew 1974). Whether this strategy is successful depends on how constituents allocate credit for credit claiming messages.

Several experimental studies have sought to document the credit that constituents allocate for different types of messages from elected officials. Grimmer, Messing and Westwood (2012) show that, due to the low levels of political knowledge and engagement, voters allocate credit for the frequency of messaging, rather than its substance. Such a result could incentivize strategic incumbents to claim credit for indirect accomplishments (Stein and Bickers 1994; Evans 1994; Grimmer, Messing and Westwood 2012), or even strategically exaggerate their involvement with government actions (Grimmer, Westwood and Messing 2014).[2] Left unchecked, this would be concerning from the standpoint of democratic accountability. Yet experimental evidence has shown that, if successfully challenged, these types of exaggerated credit claims actually damage the electoral prospects of the incumbent (Grimmer, Westwood and Messing 2014). Thus, these concerns would be mitigated if challengers sought to expose these types of exaggerated credit claims.

However, conditional upon entering an election, challengers do not generally employ such a strategy against incumbents that campaign on their record (Grimmer, Messing and Westwood 2012). Given these prior findings, why would this be the case? To answer this question, I develop a modified signalling game that formally models these types of strategies by incumbents and challengers in an electoral setting. In equilibrium, I find that the likelihood for ineffective incumbents being successfully challenged on exaggerated

[2]An important caveat to this characterization of credit claiming is that I am not universally suggesting that legislators deserve zero credit for these types of accomplishments. Indeed, a legislator may have been actively involved with the applications, follow-ups, and meetings that led to successful grants being awarded to their district by a federal agency. However, empirical and experimental evidence has shown that the subsequent constituent credit allocation does not accurately reflect how instrumental or decisive lawmakers were in their accomplishments (Stein and Bickers 1994; Evans 1994; Grimmer, Messing and Westwood 2012; Grimmer, Westwood and Messing 2014). This creates an incentive to exaggerate one's claim to the credit that is being sought.

credit claims is necessarily lower than one might intuitively expect. Comparative statics from the model show that this effect is largely driven by two factors: (i) the likelihood of successful challenges of ineffective incumbents, (ii) the level of voter accountability towards exaggerated credit claims.

The use of a formal model in this setting makes several theoretical contributions to the literature on credit claiming and campaigns. For one, it seeks to formalize the incentives and strategies around credit claiming, which has been missing from the literature (Grimmer, Messing and Westwood 2012). Second, and related, it contributes to a growing literature on challengers that actively challenge the signal of an incumbent (Dziuda and Howell 2021; Ogden and Medina 2020; Shotts and Ashworth 2011). In this case, the model allows for a receiver that can challenge the sender's signal, with some probabilistic outcome that is distributed asymmetrically across sender types. This type of formal structure can be extended to model other settings in which the receiver would challenge the sender.

These results also suggest future empirical work that could bolster our understanding of credit claiming, campaign messaging, and constituent demand for accountability and honesty in elections. This work emphasizes the need to establish more precise empirical knowledge of different types of credit claiming behavior, as well as the subsequent responses from constituents. My findings also provide the foundations to suggest that constituent demand for accountability and honesty may have taken a back seat to polarized ideological preferences in the American context.

The rest of this paper will proceed as follows. First, I will review the relevant literature on credit claiming and challengers in elections, primarily to emphasize the motivating puzzle that this model seeks to answer and its theoretical contributions. Then, I outline an informal theory for why it might be the case that challengers do not challenge credit claiming activity in elections. To test this argument, I then present the modified signaling game. Solving for the equilibrium and analyzing the comparative statics provides a different answer than the informal theory. This also suggests opportunities for theoretical and

empirical work in this area, which I discuss in the conclusion.

3.1 Literature Review

Before theorizing about the dynamics of credit claiming and challengers in electoral settings, it is important to highlight previous work that has been done in these areas. Doing so will not only underscore the puzzle that this model seeks to address, but also will work to emphasize the theoretical contributions of this type of model.

3.1.1 Credit Claiming and Constituent Credit Allocation

Mayhew (1974) first coined the term credit claiming in a political context, defining it as "acting so as to generate a belief in a relevant political actor (or actors) that one is personally responsible for causing the government, or some unit thereof, to do something that the actor (or actors) considers desirable" (52-53). Seeing credit claiming as an avenue towards reelection, he proclaimed it as "highly important to congressmen, with the consequence that much of congressional life is a relentless search for opportunities to engage in it" (53).

Since credit claiming is intended to "generate a belief," one may wonder how constituents respond to these types of messages. Traditional models of constituent credit allocation assume that voters possess the capacity and will to continuously monitor and track these types of messages in order to accurately allocate credit (e.g. Levitt and Snyder Jr 1997; Stein and Bickers 1994). However, as has been documented thoroughly in the literature, voters possess low and uneven levels of political knowledge and engagement (Lazarsfeld, Berelson and Gaudet 1968; Campbell et al. 1980; Converse 2006; Carpini and Keeter 1996; Bickers and Stein 1996). Thus, if voters encounter these types of messages, it would follow that credit claiming affects constituents' *impressions* of legislator effectiveness, which may or may not reflect reality (Lodge, Steenbergen and Brau 1995).

Empirical and experimental work has shown this to be the case. Empirically, it has been

shown that legislators claim credit for activities that they are, at best, *indirectly* involved with (Stein and Bickers 1994; Evans 1994; Grimmer, Messing and Westwood 2012; Grimmer, Westwood and Messing 2014).[3] Nonetheless, upon encountering these types of messages, experimental work has shown that constituents award credit to the sender as if they were instrumental or pivotal in the completion of the task. As a result, the authors argue that constituent credit allocation can be detached from levels of involvement with a task – constituents need only to encounter the credit claiming message to award credit, even if the actual involvement was small (Grimmer, Messing and Westwood 2012; Grimmer, Westwood and Messing 2014).

This possibility of credit claiming messages to be detached from actual involvement creates a problem of accountability, as it could incentivize incumbents to claim credit for indirect accomplishments (Stein and Bickers 1994; Evans 1994; Grimmer, Messing and Westwood 2012), strategically exaggerate their involvement with government actions (Grimmer, Westwood and Messing 2014), or even lie about their involvement in order to pander to voters. If voters do not actively monitor the *content* of these messages, legislators could easily get away with misrepresenting their accomplishments.

Further experimental work into constituent credit allocation has sought to mitigate this concern. Grimmer, Westwood and Messing (2014) show that if a subject is introduced to such an exaggerated credit claiming message, and then provided a corrective lesson on the actual process by which the accomplishment had occurred, then the subject evaluates the incumbent as *less* effective at their job.[4] In the context of an election, then, it would

[3]While these studies focus on claiming credit for grants awarded by bureaucratic agencies, there are other examples of this type of behavior. Announcing your role as a co-sponsor on a bill can create an inflated impression of involvement, given the low cost of signing on as a co-sponsor. Claiming credit for the policy outcomes reached by other lawmakers and administrations has also been a prevalent strategy. In all of these cases, the intent is to create an exaggerated impression of one's effectiveness in accomplishing tasks.

[4]Grimmer, Westwood and Messing (2014) accomplish this in the context of credit

stand to reason that a challenger that targets these types of messages would gain support from voters. This type of strategy would also alleviate the accountability concern that exaggerated credit claiming has the potential to create.

However, empirically, challengers have not been known to adopt this strategy (Grimmer, Messing and Westwood 2012). Why would this be the case, given these potential benefits? The problem is not the supply of exaggerated credit claims, as empirical work has shown that there is both quantity and variation in legislators that engage in this type of activity (Stein and Bickers 1994; Evans 1994; Grimmer, Messing and Westwood 2012). This suggests that an answer to this question must involve the willingness of, or demand for, challengers that confront these exaggerated credit claims. Taken together, these observations inform the dynamics that a model must feature in order to capture the strategic tensions that determine outcomes.

3.1.2 Challengers' Roles in Elections

With this crucial role of the challenger in deterring exaggerated credit claims, it is worth considering how challengers have been featured as strategic actors in the theoretical literature. Traditionally, challengers are conceived of as passive players in models of electoral politics. Most originally, they were featured in models of challenger entry, where they simply choose to enter an election or not (e.g. Epstein and Zemsky 1995; Goodliffe 2005). Further developments conceive of the decision to enter as a costly signal of their quality (e.g Gordon, Huber and Landa 2007). While some models allow the challenger to adopt party platforms (e.g. Kramer 1977), for the most part, challengers merely exist in these games, and their existence drives the incumbent's incentives and strategies.

More recent theoretical literature, however, has expanded the role of challengers in elections. In these models, challengers can: gather hard, verifiable information on incumbent

---

claiming on bureaucratic grants, but it would stand to reason that the same corrective approach would work on other types of exaggerated credit claims.

performance (Shotts and Ashworth 2011); trigger scandals of varying informational value (Dziuda and Howell 2021); or expend effort on oppositional research (Ogden and Medina 2020). These efforts differ from previous models because in this case, the incumbent is not merely worried about whether a challenger will enter, but is also concerned with how the challenger will respond to their election strategy.

I argue that such a challenger role is necessary for a model of credit claiming, since without some potential risk of being “found out," an incumbent would be incentivized to claim credit all the time, even on things with which they had no involvement. Thus, the ability of a challenger to challenge the incumbent’s signal will be a key part of the model in this paper.

3.2 Theory

To examine the dynamics of credit claiming and challenges in an election, I model a dynamic game of incomplete information in which an incumbent selects a campaign strategy and a challenger responds with their own campaign strategy. But first, and in order to derive baseline expectations for credit claiming and challenges in elections, I present an intuitive “soft theory" of these dynamics.

3.2.1 Soft Theory

As the literature has shown, constituents allocate credit for the frequency of credit claiming messages, and not necessarily the content (Grimmer, Messing and Westwood 2012). This implies that if a credit claiming message goes unchallenged, then constituents would more evaluate the incumbent more favorably. In a two-candidate election, this necessarily means that the challenger would be evaluated less favorably. Furthermore, experimental work on credit allocation has shown that, when successfully challenged, exaggerated credit claims *damage* the reputation of the incumbent (Grimmer, Westwood and

Messing 2014). Thus, the effect of exaggerated credit claims on constituent credit allocation necessarily depends on whether challenges to them: (i) take place, and (ii) are successful.

From this reasoning, we can derive some expectations about the frequency of exaggerated credit claims and challenges as the probability of successful challenges changes. Namely, as the probability of a successful challenge increases, the net benefit to an incumbent that exaggerates their claims would decrease, resulting in less credit claiming. For the challenger, this would necessarily mean that the net benefit would increase, and we should expect more challenges to credit claims. The converse would hold in the event that the probability of successful challenges decreases. Table 3.1 below summarizes these expected relationships.

Candidate	Pr(Successful Challenge)	Net Effect	Expectation
Incumbent	*increases*	–	Less claiming
	decreases	+	More claiming
Challenger	*increases*	+	More challenging
	decreases	–	Less challenging

Table 3.1: “Soft Theory" Expectations

Applied to the central research question, these intuitions would suggest that the reason that we do not observe challenges to credit claiming activity in electoral settings is that the probability of successful challenges is too low. But does such a simple explanation fully characterize the strategies and incentives of candidates in an electoral setting? A formal model would more fully assess these intuitions, as well as the research question.

3.2.2 The Model

In this model, we have a campaign with two players, an Incumbent and a Challenger. The Incumbent is of two possible types – $\theta_I \in \{Lo, Hi\}$ – which indicates their effectiveness in office. This type is private information to the Incumbent, but all players know the probability of the Incumbent being a high type, with $\Pr(\theta_I = Hi) = \mu \in (0,1)$.

The Incumbent and the Challenger engage in a short game that represents a campaign. First, the Incumbent moves, with a decision to either make campaign promises, or to claim achievements from their previous term in office. Thus, $a_I = \{\text{Promise, Claim}\}$. After the Incumbent selects a campaign strategy, the Challenger moves. Since the Challenger does not have a previous record off of which to make claims, their decision space is different. The Challenger can make promises of what they will accomplish if elected, or the Challenger can challenge the campaign messaging of the Incumbent. Thus, $a_C = \{\text{Promise, Challenge}\}$. While a simplified form of the nuances of language in credit claiming, it is not a stretch to assume that low-type Incumbents are the ones that would engage in the exaggerated credit claiming activity that is susceptible to challenges.[5]

Both players care about winning the campaign and getting elected. Thus, payoffs will represent their expected probabilities of winning. Since the model excludes a representative voter,[6] it is necessary to make some assumptions about how different action combinations would affect payoffs.

In the simplest case, in which the candidates choose (Promise, Promise), one could imagine that the electorate is not learning much in the campaign. Thus, one could assume that the Incumbent would be elected with some baseline probability q, where incumbency advantage would suggest that $q > \frac{1}{2}$. This implies that the baseline probability for the Challenger winning would be $(1-q)$, such that $(1-q) < \frac{1}{2}$.

---

[5]Indeed, as variations of the model that are explained below show, the model is robust to alternative explanations that allow for high-types to engage in some form of exaggerated credit claiming.

[6]The primary reason for excluding a representative voter from this model is that this actor would need to update over multiple signals. A more elaborate version of the model could expand to a continuous type space, and have a representative voter that updates on a Beta distribution (e.g. Gailmard and Patty 2019). Yet such a model would likely need to make some cutpoint assumptions in order to simplify the set of strategy profiles. Alternatively, this model could be altered to introduce a voter as a non-strategic player, yet the results would carry through.

In the next case, in which the candidates choose (Claim, Promise), one can imagine that the electorate would view the Incumbent much more favorably, since they gave a clearer signal of their effectiveness that went unchallenged in the campaign. Thus, we can assume that the Incumbent would be elected with a higher probability, $q + \delta_{min}$, subject to the constraint that $(q + \delta_{min}) < 1$. Accordingly, the Challenger would be elected with a probability $(1 - q - \delta_{min})$, subject to the constraint that $(1 - q - \delta_{min}) \geq 0$.

In the case that the candidates choose (Promise, Challenge), the probability of each candidate winning depends on the outcome of the challenge.[7] A challenge is successful against low types with probability $p \sim \mathscr{U}(0,1)$. Both players know the distribution of p, but not the particular draw for the game. Meanwhile, if the Incumbent is of a high type, then a challenge is never successful. One can think of this as there not being any verifiable "dirt" on the Incumbent because they are in fact a high type.[8] If the challenge fails, the Challenger receives a knock to their probability of winning, from $(1 - q)$ to $(1 - q - \delta_{min})$, where δ_{min} represents the cost of a failed challenge. Accordingly, the probability of the Incumbent winning is now $(q + \delta_{min})$. These changes to the probabilities of winning can be thought of as perceived incompetence from a lost challenge, or as a boost to the Incumbent's campaign promises in light of a failed challenge to them. Meanwhile, if the challenge is successful, then the reverse happens – the probability of the Incumbent winning becomes $(q - \delta_{min})$, while the probability of the Challenger winning becomes $(1 - q + \delta_{min})$.

In the case that the candidates choose (Claim, Challenge), a similar dynamic happens. The challenge is successful against low types with probability $p \sim \mathscr{U}(0,1)$, and never suc-

---

[7] In this type of election, the Challenger would be asserting that the Incumbent's promises are not realistic or otherwise feasible. This type of information is less verifiable than challenges to credit claims, hence a lower risk/reward for doing so.

[8] As will be discussed below, and is proven in Appendix B.1.3, the results of this model are robust to specifications where the probability for successful challenges of high types is not zero, but still lower than the likelihood of successful challenges against low types.

cessful against high type.[9] As with the previous case, whether the challenge is successful determines who is penalized and rewarded as a result. However, this time, the penalty or reward to their probabilities of winning is δ_{max}, where $\delta_{max} \geq \delta_{min}$. This represents the risk involved in credit claiming in a campaign when the Challenger is focused on challenging the Incumbent's effectiveness. If the challenge is successful, not only does the Incumbent look less effective, but their campaign messages and credit claims throughout are called into question, since they were effectively called out on a lie. Meanwhile, if the challenge fails, the Incumbent not only looks more effective, but the perceived lack of verifiable "dirt" makes the credit claims from the Incumbent seem even more credible.

To summarize the game, the order of play and observable information is as follows:

- Nature draws θ_I and p
- Incumbent privately observes θ_I, and chooses $a_I \in \{\text{Promise, Claim}\}$
- Challenger observes a_I, and then chooses $a_C \in \{\text{Promise, Challenge}\}$
- Payoffs are realized

Figure 3.1 depicts this game visually. Such a depiction emphasizes the theoretical contribution of this model to the discipline – it represents a modified signalling game in which the receiver can challenge the signal, with some probabilistic outcome that is distributed asymmetrically across sender types.

3.2.3 Equilibria

In this section, I outline three classes of equilibria that will serve as the basis for analysis: (1) a separating equilibrium in which high-types Claim and low-types Promise, (2)

[9] One may reasonably wonder if the probability of success for challenges of promises should be the same as challenges of claims. As will be discussed below, and is proven in Appendix B.1.4, the results of this model are robust to a specification in which promises are less challengeable than claims.

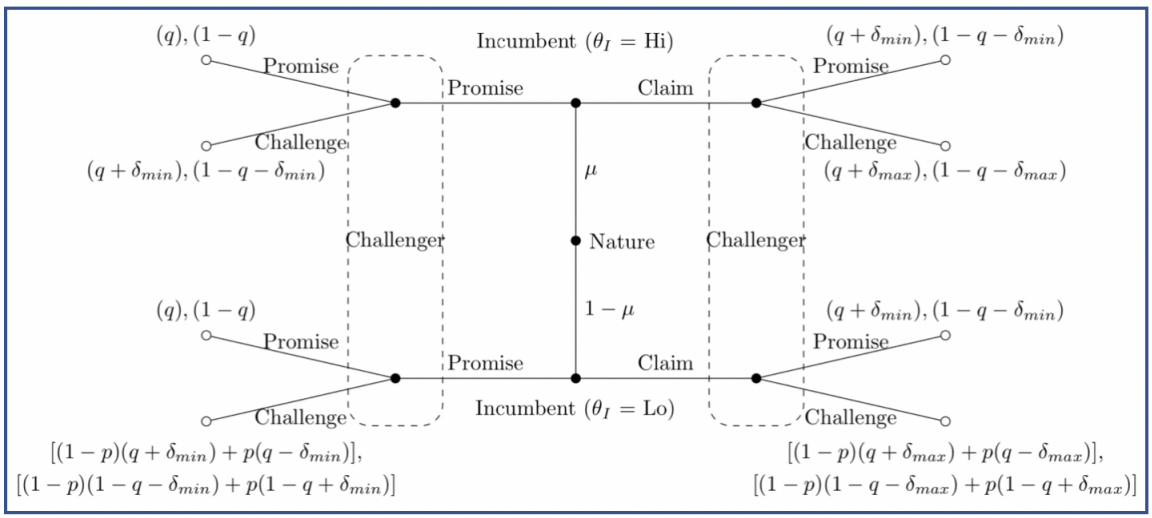


Figure 3.1: Formal Model Game Tree

a pooling equilibrium in which both Incumbent types Claim, and (3) a semi-separating equilibrium in which high-types Claim and low-types mix between the two actions.

Separating Equilibrium

From the set-up of the game, several observations are immediately apparent. First, it is the case that high-type Incumbents are always incentivized to select $a_I = Claim$. If the Challenger were to select $a_C = Promise$, then they would receive a payoff of $(q+\delta_{min})$, as opposed to a payoff of q if they chose $a_I = Promise$. Meanwhile, if the Challenger were to select $a_C = Challenge$, then they would receive a payoff of $(q+\delta_{max})$, as opposed to a payoff of $(q+\delta_{min})$ if they chose $a_I = Promise$. Another observation comes from the perspective of the Challenger – if the Challenger were to know that θ_I = Hi, then they would prefer to choose $a_C = Promise$. This is due to the fact that if a Challenger faces a high-type Incumbent, then they would earn $(1-q-\delta_{max})$ by choosing $a_C = Challenge$, as opposed to the higher payoff of $(1-q-\delta_{min})$ by choosing $a_C = Promise$.

Deducing what the Challenger would expect of low-type Incumbents lays the foundation for arriving at separating and pooling equilibria. First, let us assume the Challenger believes that high-types choose Claim and low-types choose Promise. With these beliefs,

they would challenge a Promise if $p > \frac{1}{2}$, and only be willing to challenge a Claim if $\delta_{max} = \delta_{min}$. Under these conditions, high and low Incumbent types would stick to this assigned strategy, leading to a separating equilibrium.

Proposition 1.

There exists a separating equilibrium in which high-type Incumbents Claim and low-type Incumbents Promise, while Challengers challenge the Incumbent's messaging. This occurs when $p > \frac{1}{2}$, and if $\delta_{max} = \delta_{min}$.

The occurrence of this equilibrium under very limited conditions necessitates the examination of other types of equilibria. I turn to pooling equilibria and semi-separating equilibria next.

Pooling Equilibrium

Under what conditions do all Incumbent types claim? This is the pooling scenario that is of the most interest, as it represents the trends in the credit claiming literature. It also represents the case in which there is the least amount of accountability, as low types are incentivized exaggerate their record in a campaign setting.

Starting with the initial observations above, high Incumbent types are always incentivized to Claim. Meanwhile, low Incumbent types are incentivized to Claim when the threat of a challenge is not sufficiently credible. Given the symmetry of payoffs in the outcome of a challenge, this would occur when $p < \frac{1}{2}$. In this case, a challenge is favorable to all types of Incumbents. Specifying what is incentive-compatible for the Challenger when $p < \frac{1}{2}$, as well as their beliefs for off-path play, leads to two classes of pooling equilibria.

Proposition 2.

There exists a pooling equilibrium in which both Incumbent types Claim, and the Chal-

lenger, with beliefs $\beta_C(\theta_I = Hi | a_I = Promise) \leq \frac{2p\delta_{min} - \delta_{min}}{2p\delta_{min}}$, chooses Promise. This holds when $p < \frac{1}{2}$ and $p < \frac{\delta_{max} - \delta_{min}}{2p\delta_{max}(1-\mu)}$.

Further, there exists a pooling equilibrium in which both Incumbent types Claim, and the Challenger, with with beliefs $\beta_C(\theta_I = Hi | a_I = Promise) \leq \frac{2p\delta_{min} - \delta_{min}}{2p\delta_{min}}$, chooses Challenge. This holds when $\frac{\delta_{max} - \delta_{min}}{2p\delta_{max}(1-\mu)} < p < \frac{1}{2}$.

These equilibria start to confirm the intuitions laid out at the onset of this paper. When p is low, challengers are not incentivized to challenge the messaging of the Incumbent. However, as p increases, Challengers are more willing to challenge, despite anticipating that both types of Incumbents will Claim.

Semi-Separating Equilibrium

Given the overall goal of determining when low-type Incumbents are successfully challenged on their exaggerated credit claiming messages, I restrict my attention to a semi-separating equilibrium that defines: (1) the probability that low-type Incumbents choose *Claim*, and (2) the probability that the Challenger selects *Challenge* upon observing $a_I = Claim$. In such an equilibrium: (1) high-type Incumbents choose $a_I = Claim$, while low-type Incumbents mix between $a_I = Claim$ and $a_I = Promise$; (2) Challengers pick $a_C = Challenge$ upon observing $a_I = Promise$, and mix between $a_C = Challenge$ and $a_C = Promise$ upon observing $a_I = Claim$. Proposition 3 states such an equilibrium, the conditions for its existence, as well as its robustness to alternate reasonable specifications of the parameter p. Proofs of each part of Proposition 3 are given in Appendix B.1.

Proposition 3.

i. $\exists$ a semi-separating equilibrium in which:

   - High Incumbent types choose *Claim*

- Low Incumbent types choose *Claim* with probability $x^* = \frac{1-q-\delta_{min}-\mu(1-q-\delta_{max})}{(1-\mu)(1-q-\delta_{max}+2p\delta_{max})}$, and *Promise* with probability $(1-x^*)$
- Challengers that observe *Promise* choose *Challenge*
- Challengers that observe *Claim* choose *Challenge* with probability $y^* = \frac{2p\delta_{min}}{2p\delta_{max}-\delta_{max}+\delta_{min}}$, and *Promise* with probability $(1-y^*)$

ii. This semi-separating equilibrium holds when:

- $p > \frac{\delta_{max}-\delta_{min}}{2\delta_{max}}$
- $\mu < \frac{2p\delta_{max}-\delta_{max}+\delta_{min}}{2p\delta_{max}}$

* This equilibrium is robust to alternate specifications of p in which:

iii. Successful challenges to high types are less likely than that for low types

iv. Successful challenges to claims are less than that for promises

It is worth noting the robustness of these results to different specifications of successful challenges (p), as indicated by parts *iii.* and *iv.* of Proposition 1. The first alternate specification addresses the point that it might be reasonable to suspect that both low-type and high-type incumbents engage in exaggerated credit claiming. This is supported by the literature, which acknowledges both variation in credit claiming activity, as well as "exaggerated" claims even amongst the most secure incumbents (Stein and Bickers 1994; Evans 1994; Grimmer, Messing and Westwood 2012). However, it is also reasonable to assume that effective incumbents would have a richer record from which to draw than ineffective ones, and thus would exaggerate less often. In expectation, then, it would be reasonable to assume an alternative specification in which p_H = Pr[Challenge Successful | a_I = Claim, θ_I = Hi], p_L = Pr[Challenge Successful | a_I = Claim, θ_I = Lo], and $p_H < p_L$. Appendix B.1.3 proves that the results of this model are robust to this specification.

Another alternate and reasonable specification would suggest that a campaign promise is fundamentally different from a credit claim. In being prospective, a promise is extremely

difficult to falsify, since a candidate that made promises in the election would only be held to account once they have reaped the benefits of office. Meanwhile, credit claiming activity, in being based on a record of achievements, is much more easily falsifiable. Thus, it would stand to reason that the probability of successful challenges to promises would be lower than that for claims. Appendix B.1.4 proves that, if we specify that p_P = Pr[Challenge Successful | a_I = Promise] and p_C = Pr[Challenge Successful | a_I = Claim], then the equilibrium only holds when $p_P < p_C$, which is exactly in line with these expectations. Thus, the model is clearly robust to this type of specification.

3.3 Semi-Separating Equilibrium Analysis

This project seeks to answer why challenging exaggerated credit claims is not a viable strategy for challengers, given experimental results that show such a strategy to benefit challengers (Grimmer, Westwood and Messing 2014). Through this model, I have derived the essential semi-separating equilibrium strategies necessary for probing this puzzle, and have summarized them below in Table 3.2. For brevity, I will use the corresponding symbols in the analysis.

Player	Strategy	Symbol	Value
Incumbent	Pr[a_I = Claim | θ_I = Lo]	x^*	$\frac{1-q-\delta_{min}-\mu(1-q-\delta_{max})}{(1-\mu)(1-q-\delta_{max}+2p\delta_{max})}$
Challenger	Pr[a_C = Challenge | a_I = Claim]	y^*	$\frac{2p\delta_{min}}{2p\delta_{max}-\delta_{max}+\delta_{min}}$

Table 3.2: Equilibrium Results

Given these results, it is necessary to assess the following: (1) how x^* and y^* each vary with respect to p; (2) how these relationships affect the likelihood of certain types of elections, when compared to the intuitive expectations derived earlier; and (3) the magnitude of these effects on different types of elections. I present each of these in turn.

3.3.1 Effects of p on Equilibrium Strategies

Without a doubt, the most vital comparative static to examine would be p, Pr[Challenge Successful | θ_I = Lo]. Proposition 4 below outlines the relationships between x^* and p, as well as y^* and p.

Proposition 4.

i. As the likelihood of a successful challenge increases, $Pr(a_I = Claim|\theta_I = Lo)$ decreases (that is, $\frac{\partial x^*}{\partial p} < 0$)

ii. As the likelihood of a successful challenge increases, $Pr(a_C = Challenge|a_I = Claim)$ decreases (that is, $\frac{\partial y^*}{\partial p} < 0$)

Appendix B.2 provides the proofs for these relationships. In line with the intuitive expectations earlier, x^* decreases as p increases. However, I find that y^* *also decreases* as p increases. This is a finding that is not in line with the intuitive expectations earlier, but is necessary for the equilibrium to hold. To see this, recall that, in order to keep a low-type Incumbent mixing between a_I = Promise and a_I = Claim, it must the be case that they are indifferent between the two options. This necessarily holds when EU_I(Promise | θ_I = Lo, y) = EU_I(Claim | θ_I = Lo, y), or when:

$$p(q-\delta_{min})+(1-p)(q+\delta_{min}) = y^*[p(q-\delta_{max})+(1-p)(q+\delta_{max})]+(1-y^*)(q+\delta_{min})$$

From this equation, if p were to increase, y^* would necessarily need to decrease in order to maintain the indifference, thereby maintaining the equilibrium strategy. If y^* did not decrease, it would be the case that EU_I(Promise | θ_I = Lo, y) > EU_I(Claim | θ_I = Lo, y), which would tend towards a separating equilibrium, which is outside the scope of this

paper.[10] Thus, and perhaps paradoxically when compared to earlier intuitions, it is the case that a higher probability of success induces *fewer* challenges of claims.

3.3.2 Probabilities of Different Types of Elections

Having established the relationships between x^* and p, as well as y^* and p, it stands to show that, when compared to the intuitions laid out earlier, the likelihood of low-types being called out on exaggerated claiming is lower than one might reasonably expect. To analyze these effects, I first present a comparison of the expected relationships versus the ones I derived below.

Method	Candidate	Pr(Successful Challenge)	Expectation
Intuitions	Incumbent Challenger	*increases*	Less claiming **More challenging**
Model	Incumbent Challenger	*increases*	Less claiming **Less challenging**

Table 3.3: Comparison: Expectations vs. Results

Note that this relationship applies when the semi-separating conditions apply, or when $p > \frac{\delta_{max}-\delta_{min}}{2\delta_{max}}$. For lower values, the pooling equilibrium occurs. Thus, the intuitions laid out at the start are correct up to the p values that induce semi-separation.

To determine the impact of these relationships on different types of elections, I first define these probabilities in Table 3.4 below. These likelihoods logically follow from parameters defined in the model.[11] For brevity, I refer to these probabilities by their respective

[10]For institutional designers, the conditions for separation might be interesting in their own right, as different institutions that influence the magnitude of p might induce more honest campaigning.

[11]One could reasonably argue that the second probability, that of low-type Incumbents not being called out in an election, would be $(1 - Pr[*])$, since from the perspective of a voter, $(1 - Pr[*])$ and $Pr[\neg *]$ would look the same. I proceed with $Pr[\neg *]$ as a conservative estimate for this probability, which I believe is necessary since the model does not feature a representative voter as a strategic player.

symbols in the analyses that follow.

Election Type	Symbol	Value
Low-types successfully called out on claiming	$Pr[*]$	$(1-\mu)(p)(x^*)(y^*)$
Low-types not called out on claiming	$Pr[\neg *]$	$(1-\mu)(x^*)(1-y^*)$

Table 3.4: Probabilities of Different Types of Elections

By comparing the relationships in Table 3.3 with the values in Table 3.4, one can clearly see an answer to the central research question. From the reasonable intuitions at the start of this paper, an increase in p would cause an increase in y^*, which would necessarily increase the value of $Pr[*]$. However, since this relationship has been shown to be the inverse, this necessarily means that $Pr[*]$ is lower than one might reasonably expect.

Although the relationships above provide an answer to the question of why challengers do not challenge exaggerated credit claims, one might reasonably question the *magnitude* of these effects. After all, if the relationship is decreasing at an extremely slow rate, then we should still empirically observe challengers using this strategy in elections. Figure 3.2 illustrates the values of $Pr[*]$ and $Pr[\neg *]$ for set values of other parameters.

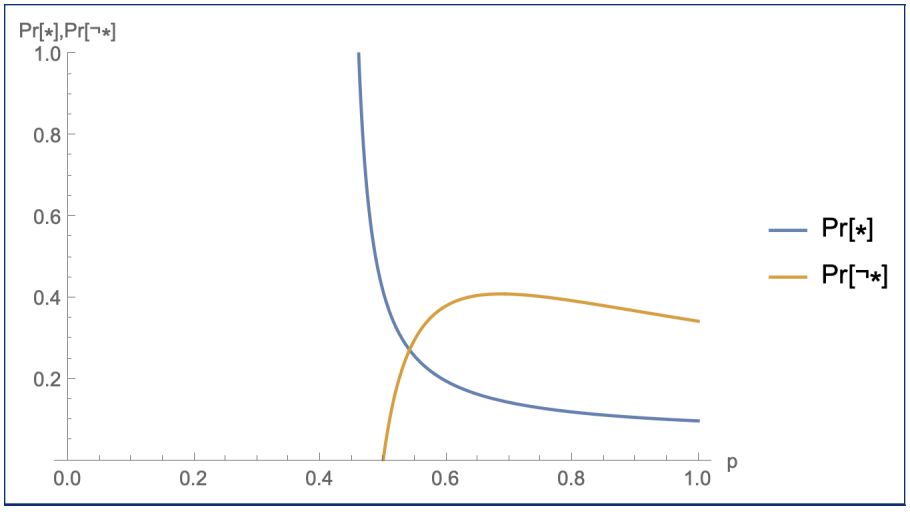


Figure 3.2: Likelihood of Different Election Types

From this graph, it is clear that, for an overwhelming majority of values for which p

exists in this equilibrium,[12] it is the case that $Pr[\neg *]$ is roughly double that of $Pr[*]$.[13] It is only when the value of p approaches the lower bound of equilibrium existence, $p > \frac{\delta_{max}-\delta_{min}}{2\delta_{max}}$, that we would expect $Pr[*]$ to be greater than $Pr[\neg *]$. Overall, for a majority of values for which p exists, it is less likely that we see low-type incumbents successfully called out on their exaggerated claims.

3.4 Broader Implications: Accountability and Equilibrium Outcomes

A broader way to consider the implications of this model is to examine the effects of different parameters on equilibrium outcomes. Given changes in comparative statics, when are Incumbent types likely to pool, separate, or semi-separate? To conduct this analysis, I first conceptualize two different parameters, p and $\delta_{max} - \delta_{min}$, as *media accountability* and *voter accountability*, respectively. Then I examine how changes to these parameters shift the equilibrium outcomes.

Conceptualizing "Media Accountability"

In the analysis above, I derived p as a parameter that affects equilibrium outcomes. In the model, p represents the probability that a challenge to an Incumbent's credit claiming activity is successful. The institution that would be most equipped to facilitate this process would be the media. Given the contentious nature of a challenge, media outlets would be able to investigate whether the Incumbent's credit claim matches their record, and report

[12] Specifically, for this equilibrium, as represented by Figure 3.2, $\delta_{min} = 0.05$, $\delta_{max} = 0.4$, $\mu = 0.5$, and $q = 0.55$.

[13] When using $(1 - Pr[*])$ as the alternative probability (see footnote 10), it is roughly four times as large on average.

the results of their investigation.[14] In this context, then, p could be conceived of as measure of media's willingness or ability to investigate Incumbent's claims in light of a challenge. Thus, for the following analysis, I consider p to be a measure of *media accountability*.

Conceptualizing "Voter Accountability"

Recall that in the model, δ_{max} and δ_{min} represented the changes in the baseline probability of winning that the candidates would receive, depending on their actions and the outcomes of challenges. These values were assumptions made about how a representative voter would view the candidates, given their messaging strategies. Moreover, the fact that $\delta_{max} > \delta_{min}$ represented the idea that credit claiming in the face of a challenge is riskier – if successfully challenged, an incumbent not only looks less effective, but they were caught exaggerating their record to voters.

With this in mind, the value $\delta_{max} - \delta_{min}$ could be conceived of as the extent to which a representative voter cares about this exaggeration of the record, which I call *voter accountability*. Larger values of $\delta_{max} - \delta_{min}$ would represent a voter that views these exaggerations as particularly scandalous, while smaller values would represent a voter that does not really care. The extent to which this value drives equilibrium results and the types of elections, then, could inform how factors such as polarization might affect voter responses to exaggerated credit claims.

The Effects of Accountability on Equilibrium Outcomes

Having conceptualized these forms of accountability and established these effects, let us examine how *media accountability* and *voter accountability* affect the equilibrium spaces

[14] A series of recent examples of what this would look like occurred when Republicans claimed credit on expenditures in 2021 Infrastructure Investment and Jobs Act, despite voting against the bill. Many of these members of Congress were called out for doing so on both cable news and social media.

for this model. Figure 3.3 illustrates the equilibrium space for low levels of voter accountability, while Figure 3.4 illustrates the equilibrium space for high levels of voter accountability. Note that both parameters are defined on open intervals from 0 to 1, and μ represents the prior beliefs of the Incumbent's type.[15]

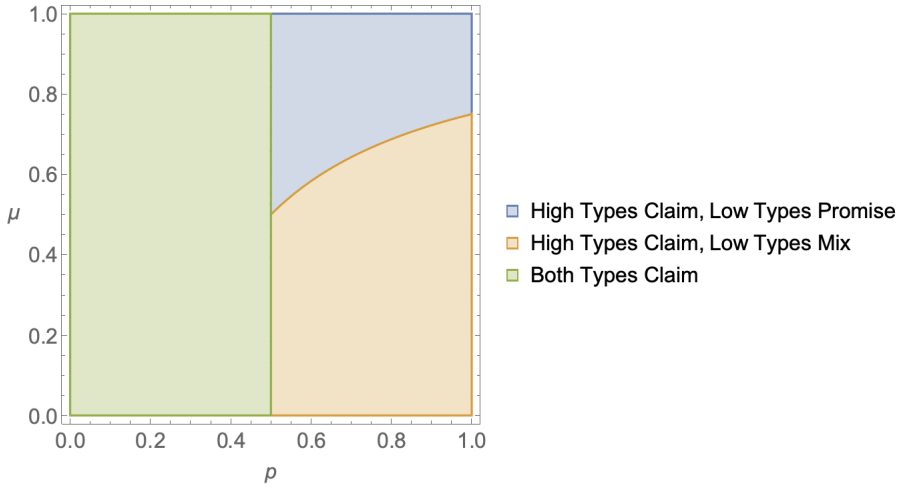


Figure 3.3: Equilibrium Space with Low Levels of Voter Accountability

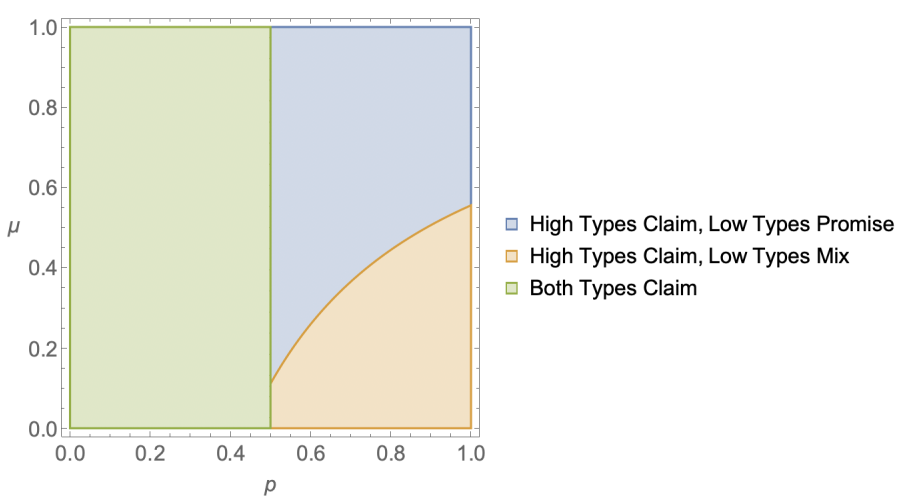


Figure 3.4: Equilibrium Space with High Levels of Voter Accountability

The legends on these figures gives and indication of which equilibrium spaces are better

[15]For Figure 3.3, $\delta_{max} = 0.10$ and $\delta_{min} = 0.05$. For Figure 3.4, $\delta_{max} = 0.45$ and $\delta_{min} = 0.05$. These results carry through for any change to the quantity $\delta_{max} - \delta_{min}$ outlined in the previous section.

from the standpoint of accountability.[16]. The blue region represents the separating equilibrium in which high Incumbent types claim while low Incumbent types promise. This is the ideal scenario in terms of accountability, as there is no exaggerated credit claiming that is occurring. The semi-separating equilibrium, which has been the subject of most of this analysis, is represented by the orange region. This is the second-best scenario, wherein a fraction of the pool of low Incumbent types are incentivized to claim credit.[17] Finally, the pooling equilibrium in green represents the worst scenario, wherein all Incumbent types are incentivized to claim credit.

Comparing the relative areas of the regions between Figure 3.3 and Figure 3.4 makes the effects of accountability clear. Media accountability (p) needs to be sufficiently high for low Incumbent types to be dissuaded from claiming credit at all. And when voter accountability ($\delta_{max} - \delta_{min}$) is high, the separating equilibrium space grows larger relative to the semi-separating equilibrium space, indicating that fewer low Incumbent types are incentivized to claim. Thus, both types of accountability are necessary to prevent exaggerated credit claiming.

However, an interesting result emerges once one considers the interactive effects of media accountability and voter accountability. Once media accountability is sufficiently high ($p > \frac{1}{2}$), increasing media accountability further *actually diminishes the effects of voter accountability*. This is shown in both figures by the slope of the line between the blue and orange regions. For larger values of media accountability, the size of the separating equilibrium region shrinks relative to the size of the semi-separating equilibrium region. This indicates that, from the broader standpoint of electoral accountability, increas-

[16]For ease of interpretation, they are listed from highest accountability on top to lowest on the bottom.

[17]Note that, in line with comparative statics above, that $x*$ and $y*$ both decrease in this area as p increases. This means that the semi-separating equilibrium will appear more like the separating equilibrium when moving rightwards cross the orange region.

ing voter accountability is unambiguously good. However, media accountability must be only moderately high to maximize the proportion of low Incumbent types that are honest in their messaging.

3.5 Discussion & Conclusion

Credit claiming is an essential tool for incumbents to signal their effectiveness to voters (e.g. Cain, Ferejohn and Fiorina 2013; Ferejohn 1974; Fenno 1978; Lazarus and Reilly 2010; Mayhew 1974). However, as voters do not typically possess the political knowledge or levels of engagement to critically evaluate these claims, incumbents can choose to misrepresent their accomplishments to bolster their constituent support (Stein and Bickers 1994; Evans 1994; Grimmer, Messing and Westwood 2012; Grimmer, Westwood and Messing 2014). Although successful challenges of these claims have been shown to reverse these effects, challengers do not actively employ this strategy in an election setting (Grimmer, Messing and Westwood 2012; Grimmer, Westwood and Messing 2014).

Through the use of a formal model, I have shown that the likelihood for ineffective incumbents being successfully challenged on exaggerated credit claims is necessarily lower than one might intuitively expect. Comparative statics from the model show that this effect is largely driven by two factors: (i) the likelihood of successful challenges of low-types, (ii) the level of voter accountability towards exaggerated credit claims.

Not only does this model contribute to our understanding of credit claiming in elections, but it also contributes to a growing theoretical literature on challengers that can actually challenge the incumbent in elections (Shotts and Ashworth 2011; Dziuda and Howell 2021; Ogden and Medina 2020). In this case, the model provides a modified signalling game in which the receiver can challenge the signal, with some probabilistic outcome that is distributed asymmetrically across sender types. While I have used this version of a signalling game to examine credit claiming activity and electoral politics, there are undoubtedly many

other possible applications for this type of model.

Relevant to this specific domain, however, there are potential paths forward in further refining the model. The interactive effects of media accountability and voter accountability suggest that further work should be done to de-couple these respective actors, and make their roles more strategic. Currently, the media and the voter are included as adjustable parameters to the model.[18] Further work could include them as strategic players, and the model would then capture the incentives of the media and the voter to report and believe the Incumbent's record of credit claiming.

Given the importance of "voter accountability" in driving these results, it would also be interesting to consider ideology as another dimension that a representative voter would care about. This would necessarily entail conceiving the voter as an actor in the game, which would require thinking more about how this voter would update over multiple signals. Once ideology is considered, a further step in a formal model could add multiple stages to account for the differences between a primary and a general election. Such a layer would add the desire for consistent messaging to the incentives that would drive candidate behavior.

More broadly, this project also motivates further work on the extent to which voter accountability affects the demand for non-ideologically valenced effectiveness and honesty on the part of lawmakers. Such a model would have candidates signaling along multiple dimensions – ideological concerns and valence – and examining equilibrium results when voters prioritize these dimensions differently. This type of work could add depth to our understanding of polarization in American politics, and its effects on voter demands from elected officials.

---

[18]They could be included as non-strategic actors whose actions are driven by these parameter values, and the current results would carry through.

Chapter 4

Measuring Bureaucratic Responsiveness Using Text from the Appropriations Process

Whether or not bureaucratic policy outcomes are responsive to the preferences of political principals, such as Congress and the president, is a perennial question in the study of the bureaucracy. Who "controls" the bureaucracy and its policy activities has implications not just for understanding the broad contours, dynamics, and content of public policy in the United States but also for assessing the quality of democracy. Given that bureaucratic officials do not face the same election incentives as legislators or chief executives, it is important to ascertain whether or not their policy choices reflect the preferences of elected officials at any point in time (and thus, ostensibly, those of the public). The question of whether the bureaucracy is responsive to the preferences of elected officials goes directly to the question of the legitimacy of the administrative state and the government broadly. Given these stakes, it unsurprising that a robust literature has developed that has focused on assessing whether or not agency activities are responsive to the preferences of various principals and under what conditions they may be more or less likely to be responsive.

However, researchers face a number of challenges in characterizing responsiveness. First, agencies are generally engaged in non-comparable missions. Different agencies carry out wildly different tasks at any point in time, making it difficult to assess differences across them. Even within the same agency, different units often produce different types of outputs. Some organizations conduct inspections. Others write regulations. Some do both and others do neither. How can we quantitatively make comparisons across these agencies? Second, it is often difficult to characterize principals' preferences over any of these given activities in a fine-grained way. How many inspections would Congress like to see? How many regulations? What should those regulations say? Often researchers must rely on relatively high-level measures of principals' preferences because the discipline lacks more specific measures of preferences. Third, while it may sometimes be easy to characterize the quantitative nature of outputs, it is more difficult for researchers to ascertain the qualitative nature of the same outputs and their concordance with principals' preferences. For instance, it is easy to count how many workplace inspections are carried out by the Occupational Safety and Health Administration. However, it is more difficult to simultaneously quantify the rigor, severity, or quality of the inspections, which may also be just as important (if not more) from the perspective of principals. Similarly, it easy to observe when agencies issue regulations, but the content of those regulations and whether Congress prefers them to the status quo are more complex questions to systematically answer on a broad scale.

We propose a measure of responsiveness that aims to overcome some of these hurdles and characterize the degree to which an agency is generally acting in accord with the desires of congressional committees in a given year. To do this, we turn to the annual appropriations process and focus specifically on the reports that appropriations subcommittees issue to accompany the legislation that they write. Much of the text of these reports discusses the operations and policy decisions that agencies have made or are in the process of making, along with the committee's judgments about the prudence of those activities. We use supervised learning techniques to code each sentence in the reports as to whether or not they

contain content indicating that the agency has engaged in behavior of which the appropriations subcommittee disapproves. This yields a responsiveness score for each sentence in the reports, which we then aggregate to the agency level to create characterization of the responsiveness of each agency to the House and Senate appropriations committees in every year from 1997–2016.

Importantly, a measure is only credible insofar as it meaningfully captures the ideas related to its corresponding concept (Adcock and Collier 2001). To assess the content validity of this measure, we examine the correlation between our measure of responsiveness and another plausible indicator of responsiveness, the volume of limitation riders included in appropriations legislation. We find that agency non-responsiveness, as measured by our method, is substantively and significantly correlated with the number of limitation riders contained in appropriations bills pertaining to that agency.

This measure allows researchers to overcome many of the challenges described above. First, it is agnostic as to the types of tasks that an agency performs. All that it picks up on is the degree to which the committee approves of the agency's current activities. This allows for comparability across agencies and, perhaps even more enticingly, within agencies over time. This allows researchers to move beyond single-agency case studies to more systematically study responsiveness. Second, the measure does not require researchers to specify congressional preferences about specific agency outputs ex ante. Rather than being forced to make indirect inferences about whether agencies are responsive to Congress based on the degree of correlation between general measures of preferences and outputs, this measure gauges responsiveness directly based on what Congress says about the agency's performance. Finally, the measure moves beyond previous work that tends to focus on a single or handful of specific agency activities and instead yields a more general characterization of the agency's responsiveness in the judgment of the appropriations committees.

At the same time, there are some drawbacks to the measure, which we also review in the paper. First, its focus on the appropriations process limits the set of actors for whom we can

make inferences about responsiveness. Perhaps most notably, we are unable to characterize agency responsiveness to the president using this measure. Second, the focus on appropriations means that issues related to management of spending are likely over-represented in the measure, though as we show, committees comment on far more than funding levels when discussing agency behaviors. Third, given the jurisdiction of the appropriations committees, agency activities in non-discretionary programs will not be captured through this measure.

Still, we believe that this measure has much to recommend it and will be of utility to political scientists interested in studying questions of bureaucratic responsiveness. To demonstrate its versatility, we examine a long-standing question in political science: to what degree do the tools of the administrative presidency bias policy outcomes away from congressional preferences? To do this, we leverage within-agency variation in the degree of politicization. Overall, we find that agency politicization does reduce responsiveness, but only under conditions where the majority and administration are not of the same party. While interesting in its own right, this analysis demonstrates the flexibility that this new measure will give political scientists in answering long-standing questions in American politics about the degree to which administrative organizations are responsive to the wishes of political principals.

4.1 The Concept of Responsiveness

Before turning to measurement, we first discuss our conceptualization of responsiveness. Doing so is necessary for both guiding our measurement strategy and delineating it from others that may be based on different (and similarly valid) ideas about what responsiveness entails. Saltzstein (1992) challenges researchers studying bureaucratic responsiveness to clearly answer three questions when defining the concept for their purposes: "[R]esponsiveness to whom? To what? And in what form?"

With respect to the first two questions, we are primarily concerned in this paper with the responsiveness of agencies to their congressional principals. In particular, we focus on the link between agency activities and the preferences of the congressional appropriations committees. This orientation reflects a normative framework that emphasizes elected politicians as the ultimate source of direction for the bureaucracy and its policy activities and discounts others, such the professional expertise or norms of employees in agencies (see, e.g., Durant 1995). In this conception, elected politicians are held to be the legitimate representatives of the people, and as such democracy demands no less than administrative responsiveness to these preferences. This is the dominant (though not exclusive) orientation of theories and empirical studies of bureaucratic responsiveness in the extant political science literature.

This idea is also consonant with (and, at least in part, grew out of) the principal-agent paradigm for analyzing the relationships between political principals and bureaucracies that gained prominence in the bureaucratic politics literature after the introduction of the new economics of organization (e.g., Moe 1984). This understanding of bureaucratic politics takes elected officials, such as Congress or the president, as the principals of agencies, which may have different preferences over policy outcomes but possess superior information about how policy choices map into outcomes. This latter characteristic leads elected officials to delegate to agencies, but in doing so, agencies may be able to skew policies away from the preferred outcomes of principals (e.g. Bendor and Meirowitz 2004; Epstein and O'Halloran 1999; Niskanen 1971*b*). As a result, political principals have developed myriad techniques to "control" the bureaucracy and compel the responsiveness of agencies to their preferences. This bureaucratic control literature has a long tradition and is centrally focused on responsiveness (e.g., Weingast and Moran 1983; Moe 1985*b*).

Finally, to answer the, "in what form" piece of Saltzstein's formulation, we focus on policy decisions made by agencies. Of course, there may be plausible conceptualizations of responsiveness that would point to other outcomes. However, we believe that this ul-

timately is the most important outcome for the way that we conceptualize responsiveness and has the clearest consequences for the empirical and normative questions that motivate it.

Of course, we should note that this is not the only way in which one could imagine conceiving responsiveness, and the measure we propose may not be able to fully capture the dynamics of other concepts. For instance, some work in political science emphasizes the reciprocal nature of the contract between politicians and bureaucrats, with each influencing the other (e.g., Carpenter 2001; Krause 1999; Workman 2008; Wood and Waterman 1994). Carpenter and Krause (2014) refer to this notion of bureaucratic-politician relationships as being based on "transactional authority." It focuses less on principals imposing contracts on agents and instead on a give-and-take relationship between politicians and bureaucrats that is far flatter than conceived here. Other concepts of responsiveness might be focused on bureaucratic loyalty to the broader state or adherence to professional norms and expertise rather than the desires of elected officials (e.g. Ingraham and Ban 1988; Mashaw 1983). The measure that we propose is less able to capture these dynamics but still may be of use for thinking about some aspects of responsiveness conceived in these ways.

4.2 Previous Studies of Responsiveness and Control

We are far from the first to take up the question of responsiveness in the bureaucracy. As such, there is a long line of studies in which scholars have attempted to determine the degree to which agencies adjust their behavior to bring it in line with preferences of elected principals. Some scholars have argued that bureaucracies have substantial (if not total) ability to mold policy outcomes to their liking, potentially in opposition to what their political overseers would prefer. In this, they are aided by large delegations of authority, constituencies that they cultivate to support them, the relative indifference of political principals, their superior policymaking expertise, and an institutional environment that makes accountabil-

ity difficult (e.g. Carpenter 2001; Dodd and Schott 1979; Eisner and Meier 1990; Hammond and Knott 1996; Lowi 1969; Niskanen 1971*b*; Potter 2019). Other work, however, emphasizes the important ways in which Congress and the president are generally able to compel responsiveness from the bureaucracy (e.g. Moe 1985*a*; Wood and Waterman 1991). Congress is able to use tools like passing laws and oversight to bring administrative policy outcomes in line with their preferences (Arnold 1979; Kriner and Schickler 2016; Weingast and Moran 1983). Presidents too draw on a number of strategies to bring about responsiveness, such as politicization and centralization (Lewis 2008; Moe 1985*b*; Nathan 1983; Randall 1979; Rudalevige 2002).

Here, we discuss two prominent approaches to measuring and characterizing responsiveness, with an emphasis on responsiveness to Congress since that is the focus of our measure. The first examines the correlation between some summary measure of the preferences of Congress and a policy output from an agency. The second is focused on the correlation between particular actions that Congress or another political principal takes and the resulting shift in agency policy activity.

Of these two approaches, the first method is the most common. As discussed above, the importation of agency theory into the study of the bureaucracy led to predictions that if the tools of elected principals to direct agency actions (e.g., oversight, legislation, etc.) were effective, then the policy choices of agencies should be reflective of the preferences of principals. A corollary to this prediction is that changes in principals' preferences should also be correlated with changes in the policy choices that agencies make. That is, if an agency's congressional overseers became more liberal, the policy choices that the agency makes should likewise become more liberal. In this view, the correlation between preferences and agency outputs is the indicator of responsiveness. When that correlation is high, scholars typically take it to mean that the agency is responsive to the principals' preferences.

Weingast and Moran (1983) present on archetypal example of this approach. In their study of the Federal Trade Commission (FTC), they ascribe different ideological valence

to cases the agency brought under different statutes. For instance, they take cases related to textiles or small businesses to be inherently a more conservative case choice than ones that are related to credit and truth in lending. They then regress the FTC's choice of case mix onto the Americans for Democratic Action (ADA) liberalism scores of key members of the House and the Senate, expecting that as ADA scores rise (indicating that the member of Congress is relatively more liberal) the more "liberal" types of cases should become more prevalent. Thus, the correlation between congressional preferences and its sign and statistical significance are the key tests of FTC responsiveness in their study.

The second approach to ascertaining responsiveness is exemplified by Wood and Waterman (1991; 1994). Instead of examining the responsiveness of agency activities to measures of preferences, they instead evaluate whether the application of particular tools of political control (e.g., the passage of new legislation or the appointment of a new administrator) is correlated with a change in particular types of activities. For example, in their study of the FTC, they examine whether the chairmanship of James Miller impacted the number of enforcement actions that the agency brought during the 1980s, finding that they decreased by nearly 50% during the time he presided over the commission. They take this as evidence of presidential control over the agency, given that Reagan had appointed him as chair.

These studies of responsiveness have much to recommend them, as they are clearly intended to tap into the same concept of responsiveness that we outlined in the previous section, with a particular focus on dynamic responsiveness and the substantively important policy choices that agencies make. However, there are some drawbacks to these approaches.

First, researchers are left to determine the preferences of Congress over policy outcomes and the ideological valence of agency actions. Typically, scholars have turned to aggregated measures of congressional preferences such as DW-NOMINATE or ADA scores to characterize the preferences of members of Congress and label certain agency actions as

conservative or liberal. In some cases, this may be easy to do, however, in others it may be more difficult. Aggregated measures of congressional preferences may elide important or idiosyncratic preferences that key members of Congress have on certain issues. This is problematic when examining one type of policy output, as most of the studies described above do, and may induce measurement error.

Perhaps more problematically, scholars must also ascribe ideological orientations to different types of actions and their volumes. This may be difficult to do, even in straightforward cases because of qualitative differences between different instances of the same action. For instance, in the case of inspections, agency officials could engage in more searching inspections or more surface-level ones. These qualities of agency activities may be of just as much (or even more) interest to principals than the overall levels of the activity. Some studies attempt to account for these qualitative differences, however, they can understandably do so only in blunt ways. For instance, Weingast and Moran (1983) are forced to ascribe ideological motivations to different types of cases and Moe (1985*a*) attempts to characterize NLRB decisions as either pro-business or pro-union with little space between.

Second, because scholars are focused on particular types of agency outputs, they are consigned to studying responsiveness in just one agency and often for just one subset of its activities at a time. While this has the advantage of internal validity (presuming, of course, that the chosen agency activity of study is representative of the larger behavior of the agency and does not interact in important ways with the other work of the agency), it also naturally leads to questions about the generalizability and case selection in any given study. Agencies are engaged in wildly incomparable tasks. The General Services Agency procures pencils and leases workspaces while the Federal Communications Commission auctions off spectrum and adjudicates claims about the decency of broadcasting. Comparing results across studies examining all of these activities necessarily requires caveats.

Altogether, this discussion of previous studies highlights a few ways in which the empirical study of responsiveness might be improved. First, measures that more fully capture

the scope of agency activities would provide greater confidence in the generalizability of findings. Second, the focus on a single type of policy activity has also lead studies to be primarily centered on just one agency with few comparisons across them. A measure that is comparable across agencies and within agencies over time would allow for expanded studies of responsiveness in which results across agencies could be meaningfully compared. Finally, previous measures have required researchers to make assumptions about the preferences of politicians over certain types of agency actions and the ideological meaning of volumes of particular types of activities. Relaxing these assumptions would provide greater credibility to empirical findings. We believe that our measure based on outputs from the appropriations process makes substantial progress on these fronts.

4.3 Our Measure of Responsiveness

We propose a measure of responsiveness based on text outputs from the congressional appropriations process. The appropriations process serves as an important venue in which oversight of the bureaucracy occurs on a regular basis. In addition to producing appropriations legislation, appropriations subcommittees also produce reports that accompany this legislation. Unlike the typical committee report, however, appropriations reports play an important and substantive role in the policymaking process and include both evaluations of an agency's current activities and instructions about how funds are to be used in the future (Bolton and Thrower 2019; Schick 2008).

The measure that we propose is based directly on the text of these reports and the committees' statements about agencies and their activities. For each agency, we calculate the number of the sentences in the report that are indicative of the agency pursuing actions inconsistent with the preferences of the committee. When this is higher, it indicates that there was more agency non-responsiveness in a given year. When it is lower, it indicates that agency activities were more in line with the preferences of the appropriations com-

mittee. We divide the number of negative statements by the overall number to account for differences in the amount of activities agencies are engaged in.[1] One negative sentence means much more for the Marine Mammal Commission than it does for the Environmental Protection Agency in our measure, given that the former has an average of 1.69 sentences per report while the latter has 288.32. This is appropriate in our view because the latter's mission is far more expansive than the former's. However, we note in the content validity check below that this choice does not appear to unduly drive variation in our measure.

We argue that this measure has a few important advantages over others that we hope will facilitate further study of bureaucratic responsiveness. Because oversight through the appropriations process is roughly the same for all of the agencies that rely on discretionary funding, we can be fairly confident that the data generating process is equivalent across the sample of agencies that we examine. The yearly annual nature of the appropriations process helps to avoid selection effects that are associated with looking at the application of specific types of control mechanisms by Congress, such as oversight hearings, to ascertain responsiveness. Furthermore, because we shift the focus from specific types of agency outputs to the direct expressions of congressional committees, the quantity of interest is also comparable across agencies, even when they are engaged in very different types of missions. Finally, the measure does not require users to make assumptions about either congressional preferences over certain types of activities or the salience of those activities to members of Congress. Instead, committees will choose to write about what they care about (either positively or negatively) and will ignore things that they do not.

We now turn to describing the appropriations process and reports, the way in which we

[1]It is important to stress that, since our measure focuses negative statements, it captures "non-responsiveness." One may very well wonder whether its complement, residual category fully captures "responsiveness," given that statements in this category might be more neutral with respect to policy. We argue that it is more appropriate to associate statements with no clear statement of satisfaction or dissatisfaction with agency activities with responsiveness since satisfaction with the status quo would indicate "responsiveness" more than it would "non-responsiveness."

coded the texts, a validation of the measure using limitation riders in appropriations bills, and a more in-depth discussion of its strengths and weaknesses.

4.3.1 Bureaucratic Responsiveness and the Appropriations Process

These reports are published by the House and Senate appropriations committees as a regular part of the annual appropriations process. Traditionally, the reports, which accompany the committee-passed legislation, contain more detailed guidance to agencies than is provided in related appropriations bills or conference reports, generally including detailed spending instructions, directives, expectations, and, sometimes, spending restrictions. Importantly for our purposes, the reports also contain judgments from the committee about the prudence of an agency's current policy activities, with accompanying reprobation or praise. Although these reports do not have the legal force that the bill does, all evidence suggests that agencies are quite attentive to the legislative pronouncements included in them due to the repeated nature of their interactions with the appropriations subcommittees. They will be back in a year's time asking for more funds, and if they have disregarded the committee's instructions they may face negative consequences (see Schick 2008).

We focus on these reports for several reasons: (i) the oversight occurs during the annual appropriations process, providing a relatively low-cost opportunity for legislative oversight; (ii) the reports come from the perspective of the principal actor; and (iii) the reports contain language intended to express satisfaction or dissatisfaction with agency performance. The existing literature on these reports emphasizes the important role that they play in the oversight of agencies on a regular basis and Congress's use of them as an important forum in which to reflect on the agency's activities in the previous fiscal year and instruct them on their expectations for the coming one (Bolton and Thrower 2019; Fenno 1966; Kirst 1969; Schick 2008).

An example of this language in the reports, which we use as a running case in this study, is shown below in Figure 4.1. The text comes directly from the 2000 Senate Interior and

Environment subcommittee on appropriations. Here, the Senate subcommittee is clearly dissatisfied with the fact that the Department of Energy has not supported drilling research and development in Alaska. We argue that this negative language indicates the agency has acted against the wishes of the subcommittee. These types of judgments and pronouncements are not isolated to this case. Its occurrence broadly across agencies and programs serves as the core of our measure.

> The State of Alaska produces nearly a quarter of the nation's oil, has roughly half the nation's remaining oil reserves, and contains over half the country's coal. Yet the Department of Energy has provided almost no support to develop drilling techniques that would reduce surface disturbance, reduce greenhouse gas emissions from North Slope drilling, develop heavy oil reserves, or develop Alaska's extremely low-sulfur coal reserves.
>
> The Committee is deeply disturbed by this failure, and directs the Secretary to consider the development of a Federal arctic technology center in Alaska that would focus on production and use of energy in the Arctic environment. Such a center could also conduct research on the most cost effective manner to generate electric power, which cost Alaskans as much as 500 percent more than the national average. The Secretary shall consult with the State of Alaska, the Alaska Oil and Gas Association, energy producers and generators, and other interested parties to determine the need for such a center, and shall submit a report to the Committee no later than March 1, 2000. A total of $1,000,000 has been provided in the oil and natural gas research programs to accelerate Arctic energy research efforts, including carbon sequestration, heavy oil recovery and directional drilling technology.

Figure 4.1: 2000 Senate Interior and Environment Subcommittee Report Text

Overall, these reports represent a key opportunity to understand the legislature's response to actions undertaken by executive agencies. Our task, which we turn to next, is to identify these types of passages in reports and their frequency for agencies in a given year. To do so, we first isolate the relevant policy language from the reports and use a combination of human and machine coding to identify sentences indicative of non-responsiveness. We describe these methods in the sections that follow.

4.3.2 Coding the Content of Appropriations Reports

To begin, we collected pertinent language in the reports from both the House and Senate Committees on Appropriations for fiscal years 1998 to 2017. For our purposes "pertinent" language refers to the committee's discussion of an agency's activities and appropriations. We omit language consisting of boilerplate introductory remarks about the agency and descriptions of an agency's general mission that tend to recur from year to year. In practice, this is a very small proportion of the overall report.[2]

Altogether, we collected 370,810 sentences from the House and Senate appropriations reports published from 1997-2016 (FY 1998-2017), which make up the heart of our measure of responsiveness. This text is linked to 6,587 agency-year-chamber observations. Hand-coding all of these sentences as expressing "non-responsive" language or not would be prohibitively costly. Fortunately, supervised machine learning techniques are one option for coding large amounts of text relatively cheaply.

Supervised learning entails human labeling of a set of documents (referred to as a training set), which is then used to train machine learning algorithms about which features of those documents (and combinations of features) are associated with different labels. In our case, the documents that we examine are sentences from the appropriations reports and the features of those documents are the words. We use a binary classification scheme to code each sentence as either indicative of agency responsiveness to the preferences of the appropriations subcommittee or non-responsiveness.[3]

---

[2]Research assistants were required to achieve a correlation of at least 0.95 on three reports in terms of the amount of words collected for each agency relative to the authors before they could continue collecting data. Appendix C.1 fully details the data collection protocol provided to each coder during the data collection process. It also includes an example from the 2016 House Transportation, Housing, and Urban Development subcommittee on appropriations, that illustrates what types of sentences were collected for interested readers.

[3]Note that one could imagine using other subsets of the appropriations report as "documents" to code, such as specific words, paragraphs, or the entire text devoted to discussing

Our hand-coded training set consisted of a stratified random sample of 11,577 sentences from the appropriations reports. This sample set of sentences was stratified in that we used key words inspired by Kirst (1969) and our own reading of the reports when sampling to ensure that non-responsive sentences would be well-represented in the training set. Words that were used to initially identify likely "non-responsive" sentences included "frustrated," "concerned," "troubled," "aggrieved," "disturbed," "disagrees," "discouraged," and "shocked." The training set also included randomly selected sentences that did not feature any of these words. Importantly, the words used to form this stratified sample were not deterministic in the coding procedure for the training set. This is because appropriations committees might very well express negative affect toward broader issues (i.e. the opioid epidemic, in the case of the Department of Health and Human Services) that have no direct bearing on their evaluation of the agency's performance. Thus, when hand-coding the training set, coders were told to only code sentences as "non-responsive" if the sentence was an evaluation of agency decisions. Table 4.1 uses examples from the reports to make this distinction clearer.

In all, about one-quarter of the training set consisted of sentences indicating non-responsiveness. As before, we ensured that there was sufficiently high intercoder reliability on a sample of sentences before dividing them among human coders. We then used this set of labeled documents to train machine learning algorithms to code the other 359,233 sentences in our dataset.

Scholars have developed many supervised learning algorithms to move from training sets to coding new datasets. Rather than choose just one algorithm to code our dataset, we

---

a given agency. We opted for sentences because (unlike words) they offer an entire coherent thought. Moving to larger chunks of text runs the risk of introducing multiple ideas and topics that make human coding of the overall responsiveness or non-responsiveness ambiguous (which will lead to poor algorithm performance overall). However, the trade-off associated with using smaller amounts of text is that there may be missing context for a single sentence from words or sentences near it that induce measurement error.

Table 4.1: Example Hand-coding for Training Set

Sentence	Source	Coding
The Committee is concerned about the increasing number of unintentional overdose deaths attributable to prescription and nonprescription opioids.	2017, House, LHHSE	1
The Committee is frustrated with the disappointing effort displayed over the past year by the Departments of Veterans Affairs and Defense to continue development of an integrated Electronic Health Record [iEHR].	2014, Senate, MV	0
The Committee remains aggrieved by the IRS' past attempt to propose new regulations for determining the tax-exempt status of 501(c)(4) organizations, which offended organizations across the political spectrum.	2017, House, FS	0

Note: 0 indicates "non-responsive," while 1 indicates "responsive" or "no valence"

instead rely on ensemble learning. The key idea behind ensemble methods is to use multiple algorithms to score documents and then aggregate those ratings to come up with a single code for each document. While each of these algorithms may be lower in accuracy than would be desirable, aggregation serves to enhance the overall accuracy of coding (Goodfellow, Bengio and Courville 2016). The ensemble that we rely on is made up of results from seven different algorithms implemented using the RTextTools software (Jurka et al. 2013): (1) support vector machine; (2) maximum entropy; (3) scaled linear discriminant analysis; (4) bagging; (5) boosting; (6) random forest; and (7) regression tree.[4]

After each of the algorithms classifies a document, the different ratings must be aggregated in some way. This can take a number of forms, including simple "majority rule" or some other weighted average of the algorithm classifications. In this paper, we employ logistic regression using the test set to estimate weights for each of the algorithm scores. In particular, we regress the manual code on each of the algorithm-determined codes in the test set. The estimated coefficients from the regression are then used to weight the algo-

[4]Note that the text was pre-processed in standard ways before these algorithms were applied, including the stemming of words and the removal of common linking words and words that appeared in more than 99.5% of sentences of less than 0.5% of sentences.

rithm scores for the documents that are not part of the coded training set. The results of this model are reported in Table C.1 in Appendix C.2.

Doing this yields a predicted probability that each document falls into the responsive or non-responsive category. If the predicted probability of being coded as responsive is greater than 0.94, we code the document as responsive. Otherwise, it is coded as non-responsive.[5] Two common diagnostic statistics for machine learning methods are precision (proportion of true positives out of total positive predicted) and recall (proportion of true positive predictions relative to total positively-coded documents). Using this threshold, we estimate a precision of 0.98 and a recall of 0.89. The overall accuracy of classification (i.e., the percentage of labels that were correctly predicted) was 90.1%.

Another common method of assessing model performance in the machine learning is examining the area under the receiver operating characteristic (ROC) curve. This metric (often referred to as the AUC) is useful because it combines information about both the true and false positive rates for each category (Burkov 2019). The maximum possible AUC is one. An area greater than 0.5 implies that the model is better than randomly assigning labels with equal probability. We plot the ROC curve for our logistic regression combining all of the models in the ensemble in Figure 4.2. The AUC for our ensemble model is 0.951, which suggests that overall the ensemble is performing extremely well.

Given this, we then applied the algorithms to the broader set of uncoded documents, using the logistic regression weights from the test set to calculate the predicted probability that a document fell into either category to code them.

---

[5]We chose 0.94 because this is the threshold that most closely equalizes the false positive and false negative rates in the sample. In practice, other thresholds that we examined, such as 0.5 and 0.76 (i.e., the proportion of ”responsive” documents in the training set) yield similar results.

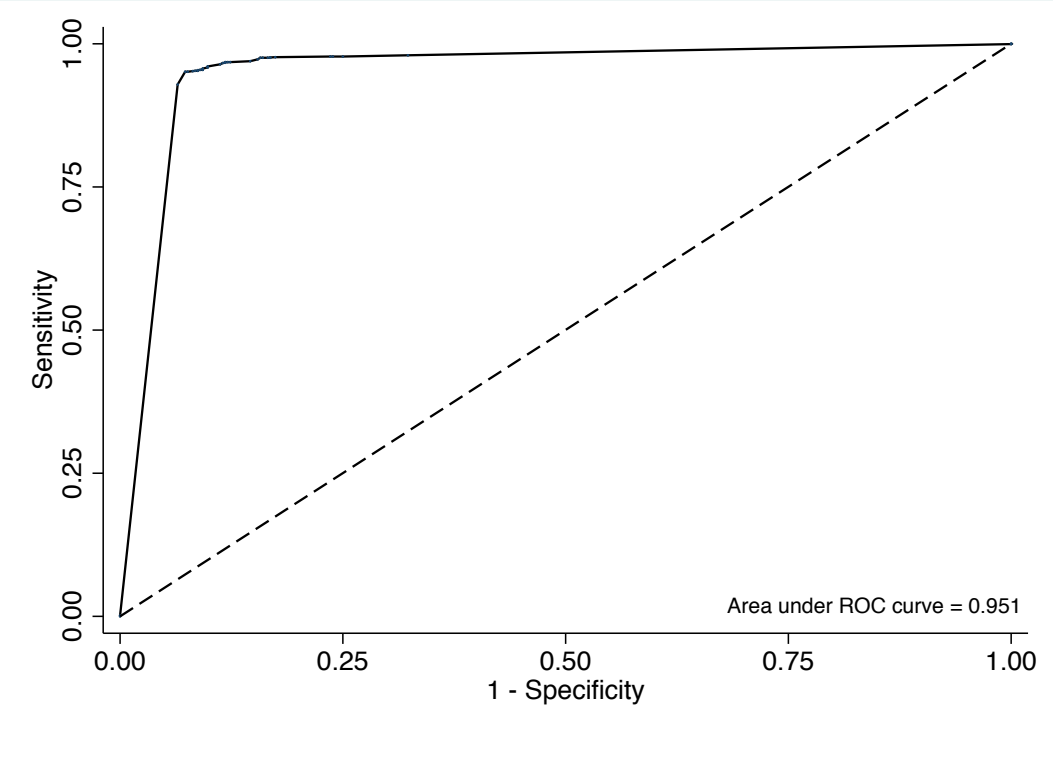


Figure 4.2: ROC Curve – Ensemble Coding Weighted by Logistic Regression

4.3.3 Summary Statistics

In all, the machine learning algorithms were used to score 359,233 sentences as having "non-responsive" or "responsive/non-valenced" language. 41,719 or approximately 11.6% of the sentences overall were coded as having "non-responsive" content. Table 4.2 illustrates the number and proportion of sentences coded as "non-responsive" by each appropriations subcommittee throughout the time period (aggregating the House and the Senate together).

The measure can also used to characterize non-responsiveness at the agency level as well. Figure 4.3 displays box-and-whisker plots of non-responsiveness for each of the agencies included in the Trump administration's Cabinet.[6] At a glance, one can see that the

[6]Note that most of the Cabinet agencies consist of multiple sub-agencies with different estimated responsiveness levels in a given year. For example, within the Department of Homeland Security, there are many bureaus, including the Transportation Security Admin-

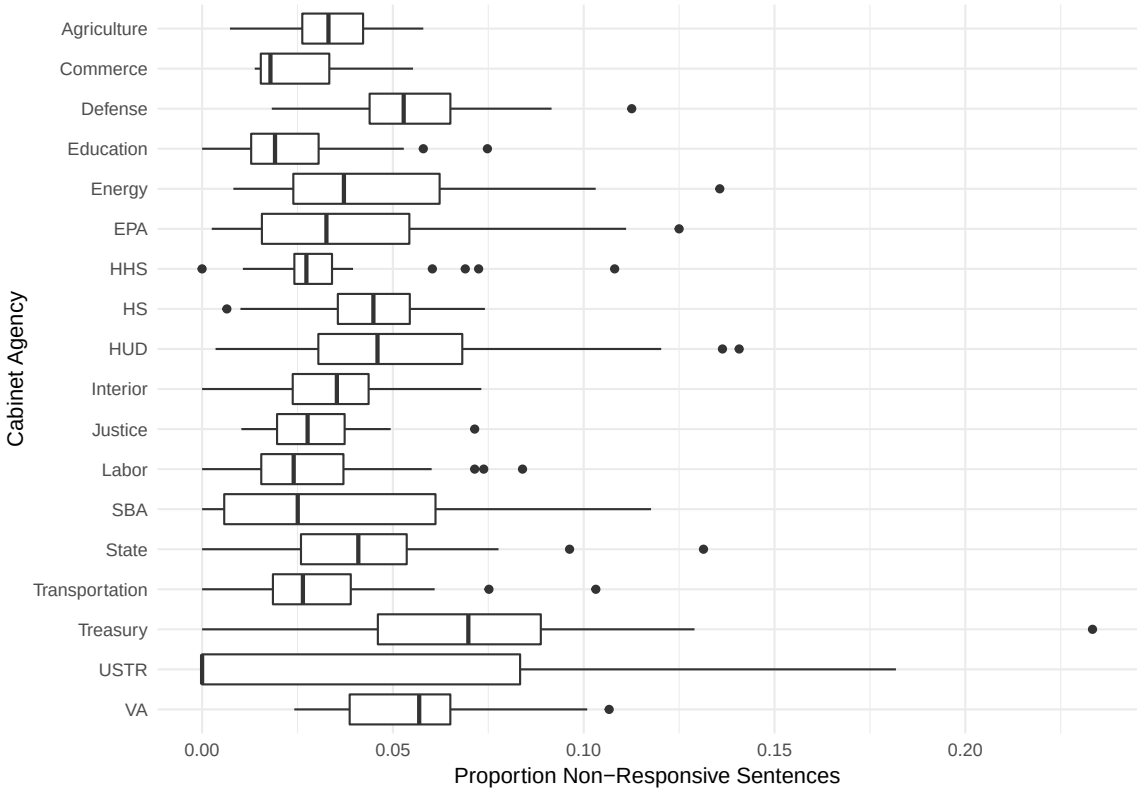
Table 4.2: Descriptive Statistics by Appropriations Subcommittee

Subcommittee	Years	No. Sentences	No. NR	Prop. NR
Agriculture	1998-2017	14743	1820	0.1234
Commerce/Justice/Science	2008-2017	18425	1329	0.0721
Commerce/Justice/State	1998-2007	12182	2049	0.1682
Defense	1998-2017	26433	3551	0.1343
Energy, Water	1998-2017	28128	3373	0.1199
Foreign Operations	1998-2007	13402	1506	0.1124
Financial Services	2008-2017	9831	1794	0.1825
Homeland Security	2004-2017	29808	3538	0.1187
Interior, Environment	1998-2017	26019	3602	0.1384
Labor, HHS, Education	1998-2017	84081	7472	0.0889
Military Construction	1998-2005	2487	288	0.1158
Military, Veterans	2006-2017	10755	1561	0.1451
State, Foreign Operations	2008-2017	14829	1402	0.0945
Transportation, HUD	2004-2017	25283	3112	0.1231
Treasury, Postal	1998-2003	5314	957	0.1801
Transportation	1998-2003	19283	2031	0.1053
Transportation, Treasury	2004-2005	2259	325	0.1439
Veterans, HUD	1998-2005	15971	2009	0.1258
Total		359233	41719	. 0.1161

Note: NR = Non-Responsive. This table displays the number of sentences collected by appropriations subcommittee, the number of sentences coded as negative, and the proportion of negative sentences. As suggested by the "Years Collected" column, some shuffling of committee jurisdictions occurred in the years 2003-2004 and 2007-2008.

measure captures across-agency variation in non-responsiveness. It also varies over time and sub-agency, as evidenced by the spread of proportions within each agency across the years measured.

Figure 4.3: Proportion Non-Responsiveness by Cabinet Agency



Abbreviations are as follows: EPA = Environmental Protection Agency; HHS = Health and Human Services; HS = Homeland Security; HUD = Housing and Urban Development; SBA = Small Business Administration; USTR = Office of the United States Trade Representative; VA = Veterans Affairs

To give more clarity to the coding that took place using the machine learning algorithms, we return to the Department of Energy excerpt presented in Figure 4.1. Table 4.3 breaks down the excerpt, sentence by sentence, displaying the codes assigned by the

istration and Immigration and Customs Enforcement. We estimate separate responsiveness scores for each of these bureaus. The box-and-whiskers plots in the graph reflect the distributions from all of the bureaus in a given Cabinet department.

weighted ensemble. Although a single case, this descriptive example provides some indication that the algorithms were picking up on the correct types of language that indicate "non-responsiveness." Additional examples of the machine coding are included in Appendix C.3.

Table 4.3: Algorithm Scores for Dept. of Energy Excerpt

Sentence	Score
The State of Alaska produces nearly a quarter of the nation's oil, has roughly half the nation's remaining oil reserves, and contains over half the country's coal.	1
Yet the Department of Energy has provided almost no support to develop drilling techniques that would reduce surface disturbance, reduce greenhouse gas emissions from North Slope drilling, develop heavy oil reserves, or develop Alaska's extremely low-sulfur coal reserves.	0
The Committee is deeply disturbed by this failure, and directs the Secretary to consider the development of a Federal arctic technology center in Alaska that would focus on production and use of energy in the Arctic environment.	0
Such a center could also conduct research on the most cost effective manner to generate electric power, which cost Alaskans as much as 500 percent more than the national average.	1
The Secretary shall consult with the State of Alaska, the Alaska Oil and Gas Association, energy producers and generators, and other interested parties to determine the need for such a center, and shall submit a report to the Committee no later than March 1, 2000.	1
A total of $1,000,000 has been provided in the oil and natural gas research programs to accelerate Arctic energy research efforts, including carbon sequestration, heavy oil recovery and directional drilling technology.	1

Note: 0 indicates "non-responsive," while 1 indicates "responsive" or "no valence"

4.3.4 Content Validity: Limitation Riders

To statistically probe the content validity of our proposed measure of responsiveness, we turn to another plausible indicator of non-responsiveness in the appropriations process: limitation riders. These legislative provisions are included directly in appropriations bills

and act to prohibit executive branch officials from using budget authority for specified purposes (Copeland 2008; MacDonald 2010). A typical rider may begin, "None of the funds made available by this Act may be used to…" These provisions can target specific actions that agencies take with regard to policymaking (e.g., prohibiting the agency from moving forward on a specific regulation it is seeking to promulgate) or more administrative-focused activities in the agency (e.g., restricting the number of vehicles in an agency's fleet or the number of employees it can hire).

Previous work suggests that these legislative provisions are likely indicators of non-responsiveness. For example, MacDonald (2010) finds that they are significantly more likely to occur during periods of divided government, when the president is most likely to direct agencies to take actions in opposition to congressional preferences. The riders target specific types of agency activities and are plausibly reactive to the actions that agencies take or are contemplating. As such, their presence or absence in legislation may serve as a useful indicator of non-responsiveness that we will leverage to further assess the content validity of our measure.

To do this, we collected data on the volume of limitation riders included in appropriations legislation reported to the House and the Senate during the fiscal years of our study. Note that while our measure exists at the agency-year level, we choose to conduct this analysis at the subcommittee-level (i.e., the unit of analysis is the bill/subcommittee report-year). Limitation riders are sometimes easily linked to specific agencies within appropriations legislation. However, many times, the riders will appear in the General Provisions titles of the legislation. In those cases, they will sometimes target specific agencies directly, but sometimes they are more subtle or target all of the agencies in the legislation. Given the necessarily inexact nature of assigning riders to specific agencies, we instead aggregate our measure up to the bill/report-level. This decreases the power of the overall test of content validity, however, it also limits the degree of measurement error introduced by inexactly linking specific riders to agencies.

In the analysis we model the number of limitation riders included in the legislation reported to the House or Senate as function of the percentage of sentences in the accompanying report that are coded as non-responsive. Because the dependent variable in the analysis is a count, we use negative binomial regressions to estimate this relationship. We account for within-subcommittee correlation in errors by clustering the standard errors at the subcommittee level in all of the analyses. Models 1–4 in Table 4.4 display the results of this first analysis.

In Model 1, we run a bivariate model regressing the number of riders onto the percentage of sentences in a report that are negative. The estimated relationship is positive, which is in line with our expectations, however it fails to reach conventional levels of statistical significance. In Models 2–4, we add additional controls that may account for confounding variables in the analysis. In Model 2, we include fixed effects for the subcommittee writing the legislation and report.[7] There is a fair amount of persistence of specific riders over time (i.e., once it is included in the bill it sometimes continues for the rest of the years in the dataset), so including these fixed effects allows us to estimate the correlations within subcommittees and bills. In Model 3, we add fixed effects for years. This accounts for factors such as divided government that may impact both outcomes. Finally, in Model 4, we include an indicator for the overall size of the bill, as measured by the logged budget authority in the legislation.[8] Larger bills may naturally lead to longer texts (thus depressing the percentage of negative statements) while also increasing the number of riders given that they deal with more programs.

In these models, we continue to see a positive, and now statistically significant and larger magnitude, relationship between the percentage of negative statements and the num-

[7]Note that we use different indicators for the House and Senate subcommittees. For example, the House Agriculture and Senate Agriculture appropriations subcommittees each have a different fixed effect.

[8]All monetary values were converted to 2017 dollars using the GDP deflator.

Table 4.4: Content Validity – Responsiveness Measure and Volume of Limitation Riders

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Percent Non-Responsive	0.798	2.432***	2.033***	1.993***				
	(1.711)	(0.789)	(0.604)	(0.630)				
Ln(Non-Responsive Sentences)					0.337***	0.143*	0.163**	0.162**
					(0.106)	(0.0825)	(0.0738)	(0.0711)
Ln(Total Sentences)					-0.0434	0.00200	0.0414	0.0322
					(0.122)	(0.0795)	(0.0704)	(0.0683)
Ln(Budget Authority)				-0.0610				-0.0359
				(0.0552)				(0.0303)
N	326	326	326	326	326	326	326	326
Subcommittee FE		✓	✓	✓		✓	✓	✓
Year FE			✓	✓			✓	✓

Negative binomial coefficients and standard errors clustered by appropriations subcommittee in parentheses. Significance codes: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

ber of limitation riders. Using the estimates from Model 4, moving from one standard deviation below the mean to above is associated with an increase in the predicted number of riders from 46.6 to 53.7 (i.e. about a 15% increase). These substantive effects sizes are in line with those of other politically-relevant variables. For comparison, using the same data we estimate that a shift from the party of the president constituting a majority on the subcommittee to not is associated with an 11% increase in the number of limitation riders included in the legislation.

One concern with using a percentage as the measure of non-responsiveness, is that it is not clear if changes are being driven by the numerator (i.e., the number of non-responsiveness sentences) or the denominator (i.e., the overall length of the report). The design of the measure is based on the idea that changes are fundamentally being driven by the variation in indications of non-responsiveness by the subcommittee. To assess this, in Models 5–8 of Table 4.4, we separate out these two components of the measure and model the number of limitation riders in the bill as a function of the logged number of non-responsive statements and the logged number of total sentences. The addition of the other controls follows the same pattern as in Models 1–4: Model 5 includes only these two variables in the regression; Model 6 adds subcommittee fixed effects; Model 7 adds year fixed effects; and Model 8 includes the amount of budgetary authority in the legislation.

Across all of these models, we see a positive and statistically significant estimated coefficient for the logged number of negative sentences in the bill, while the coefficient for the overall bill length is small in magnitude and statistically insignificant in all four models. These results add to our confidence in the content validity of the measure, and specifically, that the results in Model 4 are being driven primarily by the degree of indicated non-responsiveness in the reports. Overall, the fairly strong statistical relationship between our measure of non-responsiveness and the number of limitation riders in the associated legislation adds credence to the content validity of the measure.

4.3.5 Strengths and Weaknesses of the Measure

While we believe that this measure offers an improvement over existing ones and hope that it will facilitate research in this area, we also acknowledge that it has important limitations. In this section we review both its positive and negative features.

On the positive side, the measure is legible across agencies. Focusing on specific types of agency activities to characterize responsiveness limits the degree to which we can pursue empirical strategies beyond a single agency case study. Because congressional oversight of the bureaucracy through the annual appropriations process is relatively routinized, the data generating process for the measure is similar across agencies. By moving away from specific agency outputs and focusing on outputs from Congress, we are able to characterize responsiveness in a way that allows for comparisons across agencies and within agencies over time. This will allow researchers to generate more generalizable findings about agency activities, while still conducting analyses within agencies (by using fixed effects) if they would like.

Re-focusing attention on direct congressional outputs also allows us to avoid making assumptions about congressional preferences over very specific types of agency outputs and the salience of those outputs relative to others. When focusing on specific activities, generalized measures of preferences may or may not do a good job characterizing congressional attitudes about a particular issue. With our measure, researchers are not required to make these judgments because it is based wholly on the expressed opinions of the committee via the report.

Because the appropriations process touches on many facets of agency activities, the measure we propose also yields a characterization of responsiveness that is more general and not potentially dependent on the type of activity that one chooses to examine. The general nature of the measure also has the benefit of circumventing issues surrounding the qualitative nature of agency outputs that are lost when examining aggregate agency outputs. Thus, whether or not the Food Safety and Inspection Service is conducting inspections at

the level of rigor that Congress desires would be potentially difficult to infer from looking at patterns in terms of the number of inspections but could be articulated in the reports that we base our measure on.

While these are important benefits, the measure is not without downsides. The focus on the appropriations process has some important limitations. First, an increasing amount of agency outlays are mandatory and thus outside the scope of the appropriations process. The reports are thus overwhelmingly focused on discretionary spending activity, which is making up a smaller and smaller (albeit still important) portion of overall federal spending. Second, a disproportionate amount of the policy judgments will be related to funding issues and agency decisions in this regard. Of course, decisions about how agencies allocate funds have consequential policy effects, however, potential users of these measures should be aware of this. Third, we can only confidently use the reports to characterize responsiveness to the appropriations committees. To the degree that these bodies are not representative of the larger chamber, then we may be observing only one slice of the overall responsiveness to Congress.

Finally, some researchers may be concerned that the committees are strategic in the types of language that they are using in the report and that their words cannot be taken as sincere expressions of the committee's dismay or approval. We do not think that this is likely a major issue. There is not a clear incentive for the committee to misrepresent its judgments of agency activities in the report given that they are not highly salient to the public. The main audience for them is the agency itself. If there are strategic incentives, they may be borne out of intercameral disputes about policy. However, it is not clear that those would lead to outright directional misrepresentations about satisfaction or dissatisfaction. Similarly, one might be concerned that committees will pursue less formal means of indicating displeasure if they generally support the administration. However, Lowande (2018*b*) demonstrates that there are no real ideological patterns in the use of these sorts of informal communications, at least from the perspective of individual members of Congress.

While we held up the general nature of the measure as a benefit earlier, some scholars might also view this as a drawback if they are interested in a particular type of activity. A related concern is that the measure may not reflect the fact that some agencies are engaged in activities that are easy to adapt in response to congressional committees while others are not and that Congress may not care and still choose to reprimand agencies about these problems. For example, Gordon (2011) demonstrates that the General Services Agency has more flexibility in some types of activities relative to others and responsiveness is enhanced for those activities. There are two potential responses to this concern. First, whether or not the agency's activity is easy to change or flexible is not directly relevant – it still is unresponsive in a way that Congress cares about whether this is due to willful disregard for congressional preferences or capacity problems. Indeed, the source of non-responsiveness may actually be an object of interest. Our measure seeks to capture this non-responsiveness, not its causes. Second, we would encourage users of the measure to employ agency fixed effects (or demeaned scores) to account for any time-invariant agency differences, such as the policy activities or mission of the agency, that they may be concerned about.

4.4 Putting the Measure to Use: Politicization, Interbranch Conflict, and Bureaucratic Responsiveness

We now turn to an application that to illustrate the utility of this measure and uses it to shed light on an important question in the study of administrative politics: how does presidential politicization of agencies impact their responsiveness to Congress? Presidents enter office with lofty goals and only a short time to achieve them. As such, they must quickly harness the apparatus of the administrative state to put their policy programs into action. However, presidents often view existing administrative arrangements as unsuited to their purposes. Looking out onto the bureaucracy, they largely see career officials committed to previous administrations and the status quo, as well as organizations and procedures

unsuited to quick, energetic action (e.g., Moe 1985*b*; Wilson 1989).

As such, presidents have developed an assortment of tools to facilitate their control over bureaucratic organizations. One such tool is politicization, which entails the installation of loyal appointees in key leadership and decision-making roles in agencies (Lewis 2008; Moe 1985*a*; Weko 1995). In doing so, presidents seek to mitigate agency problems in the development and implementation of their policy programs. If the strategy works as intended, the policy outputs of the bureaucracy will more closely reflect the president's preferences than if the political appointees were not present in the agency (Randall 1979; Wood and Waterman 1991). However, this may be at the expense of responsiveness to congressional preferences, particularly if the president and congressional majorities have divergent preferences. Thus, if politicization is indeed serving the interests of presidents, we should expect to see a positive correlation between politicization levels and non-responsiveness as the preferences of Congress and the president diverge.

Of course, there may be reasons to doubt that politicization would impact responsiveness to Congress in this way. After all, Congress possesses tools to sanction agencies (and appointees, in particular) that pursue policies divergent with its preferences. Oversight, for instance, is often cited as one means through which Congress can impose costs on bureaucratic actors. Requiring agency leaders to publicly testify before Congress during charged hearings can impose both opportunity costs (by requiring them to take time away from other activities to prepare testimony and actually deliver it) as well as reputational costs. In this way, congressional retribution, or at least the threat of it, may deter agency officials from straying too far from congressional desires in the policymaking process.

Furthermore, political officials may be even better suited than career ones to act in accordance with congressional preferences. These individuals may have a greater sense of the political importance and consequences of their decisions. They may be more highly attuned to congressional and presidential preferences throughout the policy development and implementation process and may do more to consult and appease congressional over-

seers because of this. Indeed, previous work demonstrates that, at least when political appointees and careerists have productive relationships, politicization can serve to enhance overall agency performance (Krause, Lewis and Douglas 2006; Resh 2015). If this is the case, then politicization might actually facilitate responsiveness to Congress.

Lowande (2018*a*) examines as similar question using a different measure of responsiveness. There, responsiveness is operationalized as the amount of time that it takes agencies to respond to correspondence from members of Congress. Lowande finds that politicization increases response times for all members on average, and especially so for members of Congress from the party opposed to the president on case work requests. On policy-related requests, the study finds evidence that the effects of politicization are largely confined to opposite-party legislators.

This response time measure of responsiveness differs in important ways from the measure we present. In particular, it does not have the same clear policy relevance that measures focused on direct mission-related activities do (though, to be sure, it does clearly gesture toward the agency's priorities, a quantity clearly worthy of study and scholarly interest). Further, the qualitative content of agency responses is not captured by the measure and would seemingly have at least as much if not more bearing on the responsiveness of the agency in terms of policy as the length of time that it takes to respond (though, clearly, members of Congress likely care about both dimensions to some degree). Thus, there are reasons to expect our measure would potentially yield different results with respect to these arguments.

To examine this question, we explore the relationship between our measure of responsiveness, the levels of political appointees in agencies, and the partisan relationship between appropriations subcommittee majorities and the administration. In these analyses, our unit of analysis is the agency-year and the dependent variable is the percentage of sentences in the agency's section of the appropriations report that is indicative of non-responsiveness. The results of this analysis are presented in Table **??**. Because there is some skew in this

variable, we also examine a model in which we use the natural logarithm of this value (plus one) in Model 4. Notably, the results are substantively unchanged by this transformation.

The first key independent variable is an indicator of partisan divergence between the president and the appropriations subcommittee majorities. This variable (*Divided*) is coded as "1" if the subcommittee majority and the president are from different parties and as "0" otherwise. Note that within a given year, this can be different for the House and the Senate because this variable is determined at the chamber level.

Measuring politicization is somewhat less straightforward. We ultimately operationalize it in four different ways. First, following Lowande (2018*a*), we use the ratio of the number of non-career Senior Executive Service (SES) and Schedule C appointees to the number of career SES employees as one measure of politicization (denoted *Politicization Ratio* below). As this ratio increases, it indicates a larger contingent of political appointees in important managerial positions in the agency and thus higher politicization.[9] This ratio ranges from zero (indicating no political appointees) to 2.6. Because the distribution is somewhat skewed, we also run models using the natural logarithm of the ratio (plus one), which is denoted *Ln(Politicization Ratio)*, in Model 5. Finally, because the ratio is undefined in agencies where there are no career SES members (and thus those agency-years are dropped from the analysis in Models 1–5), we also include a measure that is the number of non-career SES and Schedule C appointees as percentage of the total managerial corps in the agency, *Ln(Politicization Percent)*, in Model 6. Note that when using this measure, we no longer omit agency-years in which there are no non-career SES employeess.

Finally, we also estimate models that interact these measures of politicization with the indicator for divided partisanship. These interactions will allow us to determine the degree to which politicization moderates the impacts of divided government and vice-versa.

---

[9]This measure performs poorly in small agencies because there might only be a few or no members of the career SES leading to highly skewed (in some cases exceeding eleven) or undefined values for the ratio. Because of this, focus on agencies in top three quartiles of agency employment size in our dataset (i.e., more than 223 employees).

We also include controls for the overall size and resources of the agency by including controls for the logged number of employees and budget authority in each year (*Ln(Total Employees)* and *Ln(Budget)*, respectively) in Models 2–6. In addition, all of the models include three sets of fixed effects. First, we include agency fixed effects, which account for any time-invariant features of agencies that might confound the results. The models also include year fixed effects. Note that *Divided* varies within year by chamber. Finally, we include chamber fixed effects, to account for any time-invariant differences between the House and the Senate that might bias the results.

This analysis is closely related to that in Lowande (2018*a*), which models the length of time agencies take to respond to correspondence from members of Congress as a function of member presidential co-partisanship, agency politicization, and their interaction. This analysis will allow us to determine if these variables impact our measure, which is based on more mission-centered policy activities, in similar ways or if there are important distinctions.

Model 1 estimates the uninteracted effects of the *Politicization Ratio* and *Divided* variables. The estimated coefficients for each are positive, indicating that they are each positively correlated with non-responsiveness. However, only the politicization variable is significant at conventional levels. This suggests that, on average, politicization increases the non-responsiveness of agencies to Congress. Adding controls for agency size and resources in Model 2 does not appreciably impact the results.

In Models 3–6, we investigate the potential interactive effects between politicization and partisan divergence. Across all four models, using different operationalizations of both the dependent and independent variables, we estimate a positive and significant interactive effect between the variables, while the estimated coefficients for the politicization and divided government coefficients alone (i.e., when the other is zero) are insignificant. These results suggest that the effects of divided government and politicization on agency responsiveness to Congress depend in a substantial way on the other factor.

Table 4.5: Politicization, Partisanship, and Non-Responsiveness

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Politicization Ratio	2.095*	1.940*	1.303	0.111		
	(1.049)	(1.064)	(0.997)	(0.190)		
Ln(Politicization Ratio)					1.949	
					(1.843)	
Ln(Politicization Percent)						-0.0106
						(0.112)
Divided	0.556	0.595	0.268	0.0199	0.170	0.00935
	(0.412)	(0.396)	(0.453)	(0.0652)	(0.476)	(0.415)
Divided x Politicization Ratio			0.886*	0.171**		
			(0.505)	(0.0750)		
Divided x Ln(Politicization Ratio)					1.553*	
					(0.842)	
Divided x Ln(Politicization Percent)						0.0579**
						(0.0279)
Ln(Total Employees)		2.012	2.030	0.272	2.067	1.597
		(1.538)	(1.538)	(0.245)	(1.546)	(1.447)
Ln(Budget)		0.186**	0.188**	0.109***	0.188**	0.208***
		(0.0781)	(0.0782)	(0.0217)	(0.0783)	(0.0732)
N	1,818	1,807	1,807	1,807	1,807	2,194
Agency FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Chamber FE	✓	✓	✓	✓	✓	✓

OLS coefficients and standard errors clustered by appropriations subcommittee in parentheses. Significance codes: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Note that there are substantially more observations in Model 6 because it includes agency-years in which there are no career SES employees.

When politicization is low, the effect of partisan divergence between Congress and the administration appears to have little significant impact on the degree of responsiveness of the agency to Congress. However, as politicization increases, the non-responsiveness of the agency during periods of partisan divergence also increases. In Figure 4.4, we graph the estimated marginal effect of partisan division at different levels of politicization based on the results reported in Model 3. At the lowest levels, the effect of partisan divergence is small and statistically indistinguishable from zero. However, as the ratio of appointees to career

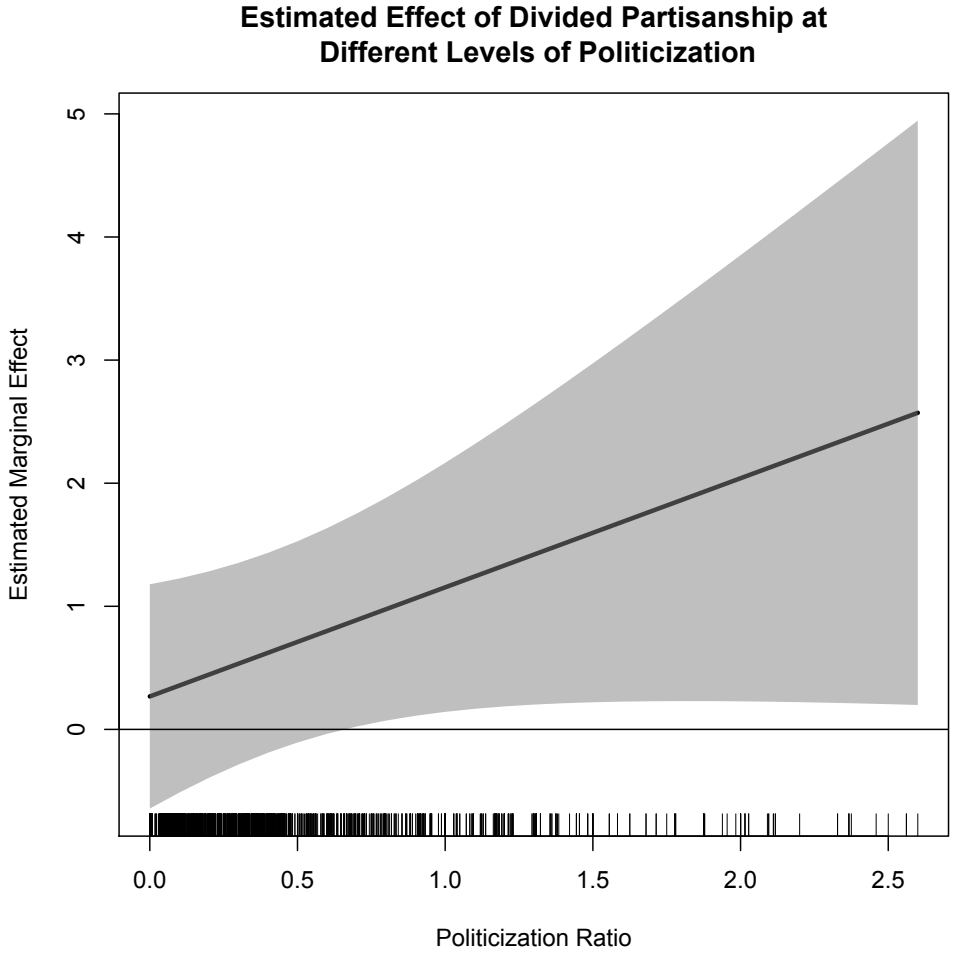


Figure 4.4: Marginal effect of divided government at different levels of politicization with 95 percent confidence intervals. Estimates are based on Model 3 in Table **??**.

SES employees in agencies increases, the effect of partisan division also increases. These results suggest that presidential efforts to politicize agencies do appear to pull policies away from congressional preferences, at least during periods of divided government.

A similar story emerges in examining the relationship between politicization and responsiveness. During periods of partisan harmony between the subcommittee and the agency, there appears to be no significant effect of politicization on agency non-responsiveness (though the estimated coefficient is positive). This suggests that when the president and Congress have relatively shared policy and political goals, politicization makes no substantial difference for responsiveness from the perspective of Congress. However, during

periods of partisan division, the estimated effect of politicization is statistically significant. In particular, the results of Model 3 suggest that a one standard deviation increase in politicization is associated with a statistically significant 0.16 standard deviation increase in non-responsiveness to Congress. Notably, this result differs from Lowande (2018*a*), who finds that politicization significantly increases agency response times to members of Congress whether or not they are a co-partisan of the president (particularly for constituent casework requests).

Overall, the results of this analysis suggest the important policy impacts of politicization. They are most consistent with a portrait of political appointees moving policy in the direction of the president. When congressional majorities are of the same party as the president, this has little impact on the responsiveness of the agency to Congress. However, when majorities and the president have divergent preferences, politicization serves to diminish the responsiveness of agencies to Congress. There is little evidence here for the "congressional dominance" school or the idea that political appointees enhance the responsiveness of agencies (at least with respect to Congress).

This analysis also highlights the flexibility of this measure of agency responsiveness to different types of questions. Here, we were able to evaluate the implications of competing arguments about the role of politicization using data across many different agencies. The comparability of the measure across agencies facilitates this kind of analysis. However, the measure also allows us to identify within-agency temporal changes in responsiveness. Indeed, we leverage this variation to identify the effects of politicization and partisan differences.

4.5 Discussion and Conclusion

Bureaucracies are at the center of policymaking and policy implementation. They have immense power over the details of public policies and their implementation and as such

have profound effects on the lived experiences of citizens. Thus, understanding which actors to whom – if any – these organizations are responsive in their decision-making is a central question in political science with enormous normative stakes. Scholars have made great strides in the last several decades examining these questions from both a theoretical and empirical standpoint. In this paper, we propose a new measure of responsiveness that we believe will help to facilitate new research in this area. The measure is based on congressional text outputs from the appropriations process. This has the benefit of yielding comparable measurements across agencies and within agencies over time and moving from measures faced on specific agency activities to allow for more generalized statements about the (non-)responsiveness of organizations to the appropriations committees in Congress. Using the correlation between our measure and limitation riders included directly in appropriations legislation, we are able to provide support to the content validity of this measure. We believe that researchers will be able to use this measure to explore important questions related to when agencies respond to Congress, the types of institutional structures that facilitate or inhibit responsiveness, how actions by presidents to control their administrations impact responsiveness to Congress, the types of leadership traits and experiences that are correlated with higher levels of responsiveness, and more.

The paper also contributes to a growing literature using text as data in the study of administrative politics (e.g. Haeder and Yackee 2015; Hollibaugh 2019; Marvel and McGrath 2016). Here we demonstrate how supervised learning techniques can be applied to congressional outputs to characterize the behavior of agencies vis-à-vis particular congressional actors. By looking directly to Congress to understand whether they approve of or disapprove of agency activities, we are able to sidestep issues related to measuring congressional preferences over specific agency activities and allow Congress to speak for itself as to whether agencies are doing what it wants or not.

Of course, as with any measure, there are limitations that future users ought to be aware of. Our focus on appropriations necessarily limits the types of policy activities at issue

(notably those funded by discretionary funding). Furthermore, differences that exist across agencies in terms of their missions and activities may make users wary of cross-agency inferences based on the measure. For those users, we recommend agency fixed effects, which would allow for meaningful within-agency inferences with greater power than the single-agency case studies that dominate the literature.

Finally, we use the measure to provide new substantive insights into the role that the administrative presidency plays in influencing the responsiveness of agencies to Congress. In particular, we show agencies with fewer political appointees and greater levels of careerists in their managerial corps tend to be responsive to Congress no matter the ideological orientation of the administration relative to Congress. In agency-years where politicization is higher, however, policy appears skewed away from congressional preferences when the administration is from the opposing party. Similarly, politicization only appears to impact non-responsiveness during periods of partisan division. We hope that this analysis is just the first of many with this measure that will help scholars to better understand the dynamics of agency responsiveness to Congress.

Chapter 5

Conclusion

The goal of this dissertation was to make contributions to our understanding of information and accountability in three specific areas of American politics. Paper 1 considered whether voters demand the type of information necessary for electoral accountability, whether voters have the capacity to process the amount of required information, and how information environments impact voters' evaluations of electoral institutions. Paper 2 examined electoral accountability from the other direction, studying the impacts of different types of accountability on candidate credit claiming strategies. Meanwhile, Paper 3 shifted focus to bureaucratic accountability, and considered whether the routinized information-sharing in the appropriations process provided an avenue to measure "bureaucratic responsiveness" across agencies and time. Each of these papers makes unique contributions to the relationship between information and accountability in the American context.

5.1 Summary of Major Findings

Through the survey experiment conducted in Paper 1, I find that voters are responsive to policy-relevant information, and match their preferences to the stated preferences of candidates. Voters also rely less on party ID when policy information is present, and express increased satisfaction with and willingness to turn out to elections with more policy-

relevant information. However, perhaps paradoxically, in terms of choosing the candidate that matches on policy overall, partisan cues are just as effective as having complete information.

From the results of the formal model in Paper 2, I find that the equilibrium likelihood of ineffective incumbents being successfully challenged on exaggerated claims is lower than one might intuitively expect. Media accountability and voter accountability drive the equilibrium results of this model, and both need to be sufficiently high for incumbents to be dissuaded from exaggerating their credit claims. Meanwhile, high levels of media accountability diminish the effects of high voter accountability in keeping incumbents honest in their messaging.

And finally, through the machine learning and statistical analyses conducted in Paper 3, my co-author and I find that the information-sharing that occurs in the appropriations process does provide the ability to measure bureaucratic responsiveness across agencies and over time. We successfully validate this measure by comparing it to another well-studied measure of accountability, limitation riders. The measure is then applied to a study of how agency politicization affects responsiveness to Congress in order to demonstrate its usefulness.

5.2 Paths for Future Research

One notable limitation to this dissertation involves its piecemeal approach to examining information and accountability in American politics. Given that it studies this topic in three different contexts, this dissertation's findings can only speak to the broader picture in suggestive terms. That is, while very generalized implications can be suggested, the dissertation itself does not test them.

For example, one potential broader implication concerns the courseness of information and its relation to accountability outcomes. The findings of Paper 1 suggest that party ID

as a heuristic is sufficient information, so long as the vote choice remains a binary one between two candidates from two distinct political parties. Meanwhile, the appropriations reports in Paper 3 take in detailed information reported by agencies over the fiscal year, and make a wide variety of judgements on those activities that factor into their final budgetary decision. The measure in Paper 3 would not work if either the information in or the detailed judgements were lacking in granularity. While these studies may suggest this relationship, the dissertation does not explicitly test it. This could be an interesting path for future research for someone interested in contributing to the broader literature.

One broader implication that the dissertation does more thoroughly establish is that more information does not equate to more accountability. Paper 1 found that voters do a worse job matching on policy preferences (despite the intent to) when presented with too much information to process. Meanwhile the findings of Paper 2 speak to the reverse causal direction, that strategic anticipation of different levels of accountability can affect the signalling information provided in the first place. The model outcomes also showed that more media accountability actually diminishes the impact of voter accountability, showing that more is not always better. Even in the case of Paper 3, we found that the length of the report – a naive measure of the amount of judgements made on agency activities – did not drive the validation results of our responsiveness measure.

This implication does point to another fruitful path for research, which involves considering each of these contexts as strategic. Both Paper 1 and Paper 3 were analyzed in non-strategic contexts. For Paper 1, it was a scope restriction on the experimental design, necessary due to budget and time constraints. While for Paper 3, we address the concern and assert that the appropriations reports are less strategic than the actual budgetary decision. In both of these cases, it would be interesting to consider information flows and accountability in a more strategic context. For Paper 1, would knowing how voters prioritize different types of information affect the policies that candidates prioritize? And for Paper 3, would anticipation of negative feedback impact the information that the agency

provides, or the budget request that they make? These would be interesting questions for future research.

On a more methodological note, this dissertation has shown that information and accountability can be studied using a wide variety of methods. Each of the papers in this dissertation utilizes a unique method, suited for the specific set of research questions and context. While not intentional, I hope that this illustrates that any political scientist can contribute to our collective knowledge on this topic.

Appendix A

Beyond Party ID? Information Environments and Voter Decisionmaking

A.1 Addressing Data Missingness

As mentioned in the body of the paper, 1640 subjects enrolled in the experimental survey, yet the results of 1581 subjects were included in the analysis. Subjects were excluded for one of two reasons: (1) they submitted no choice for any of the 15 elections; or (2) they took over three hours to submit a response (the period of data collection), indicating that their session had submitted their current progress.

Missing data, and the exclusion thereof, can bias results if missingness is correlated with treatment assignment. To assess the relationship between these, I subset the data to those observations where no election choices were made, 48 observations in total. Figure A.1 visualizes where missingness occurs in this dataset. Points with dark coloration indicate the prompts for which subject did not submit a response. If a dark line begins and then proceeds to the end of the figure, that would indicate that the subject dropped out of the

study.

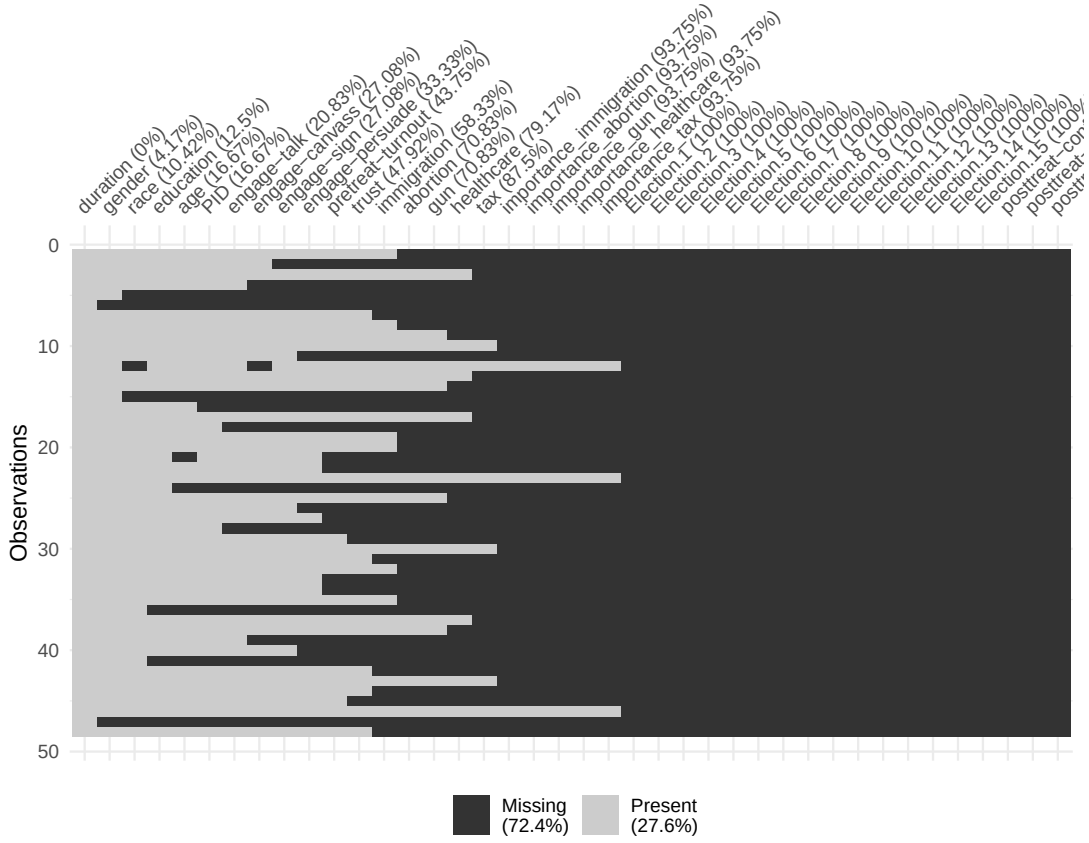


Figure A.1: Missingness in Experiment Data

As this visualization illustrates, much of the missingness is due to attrition that occurred prior to treatment assignment, indicated as column “Election.1" about midway through the figure. In total, only three subjects dropped out at the point that treatment was assigned. The remaining subjects with missing data dropped out earlier, meaning that their data missingness cannot be correlated with treatment assignment. Thus, we can conclude overall that exclusion of subjects that did not make a selection in any election would not significantly bias the results.

In terms of the subjects that took over three hours (13 in total), 3 also belong in the group above. Of the ten remaining, three fell into the “PID + 0" group, while the rest were spread amongst other treatment groups. Thus, time-out duration is not correlated with

treatment in a concerning manner.

A.2 Balance Test

Below I include the results of a balance test to ensure that randomization along important covariates was successful.

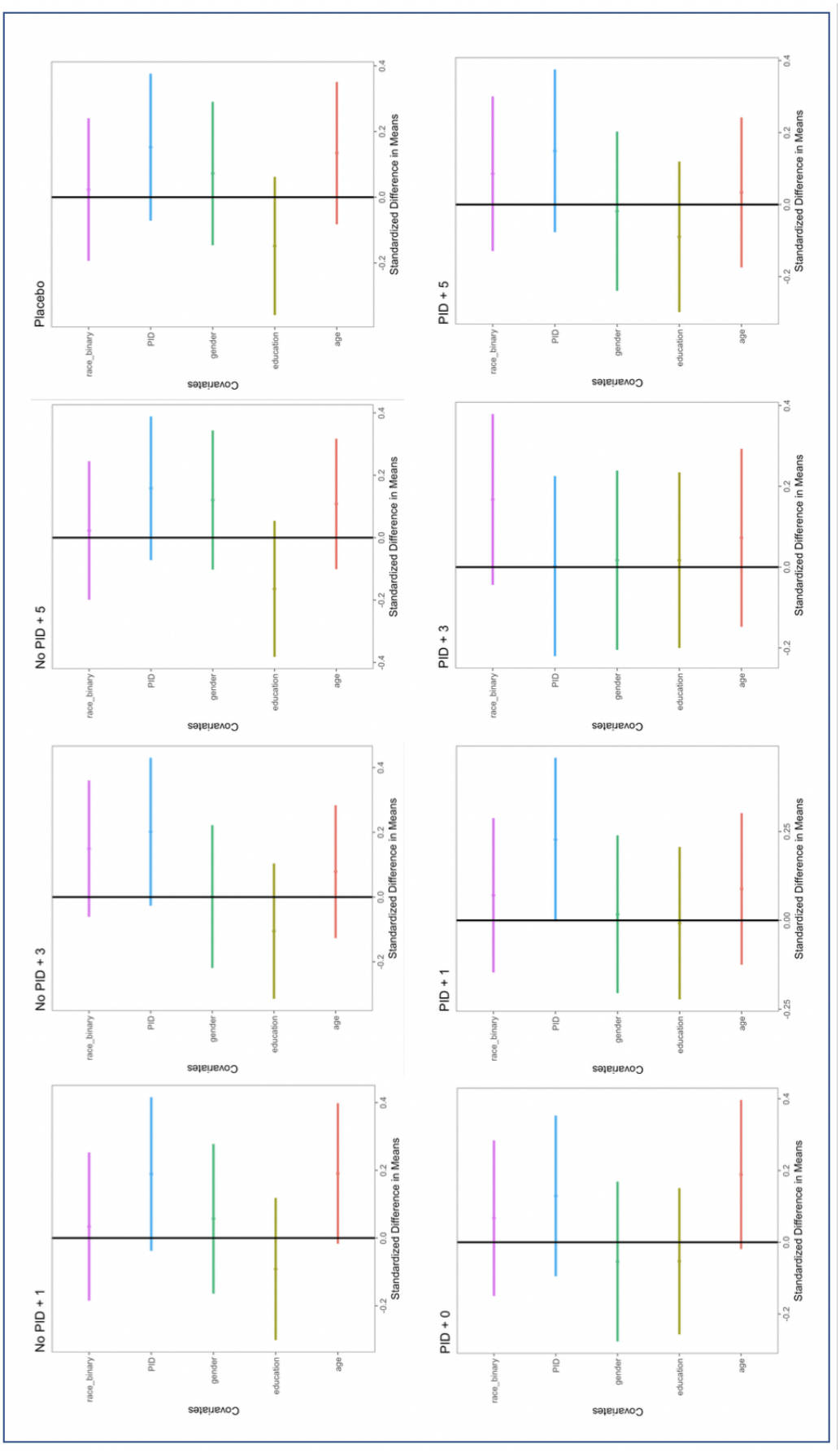


Figure A.2: Balance Test Results

A.3 Analysis Using Logit

Since the dependent variables of H1 and H2 are proportions (or rather, averages of binary outcomes), one may wonder about the degree to which the results depend on the model specification.

For robustness, I check the results of H2 (as well as the heuristic performance analysis) using logistic regression. The coefficient plots below report the results converted out of log-odds and into probabilities.

Figure A.3 presents the results of H2 using a logit model.

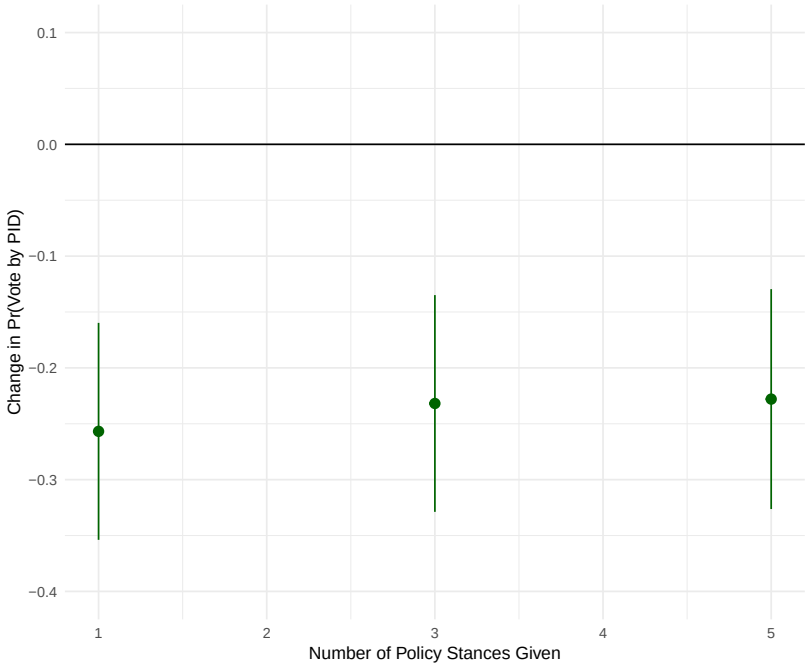


Figure A.3: H2 Results Using Logit Model

The estimates here suggest a further decrease in the use of party ID, though none of the results are statistically distinct from the results derived using OLS.

Figure A.4 presents the results of the heuristic performance assessment using a logit model.

The main results hold here. For low levels of information, subjects do worse than just having party ID. For high levels of information, subjects perform no better than just having

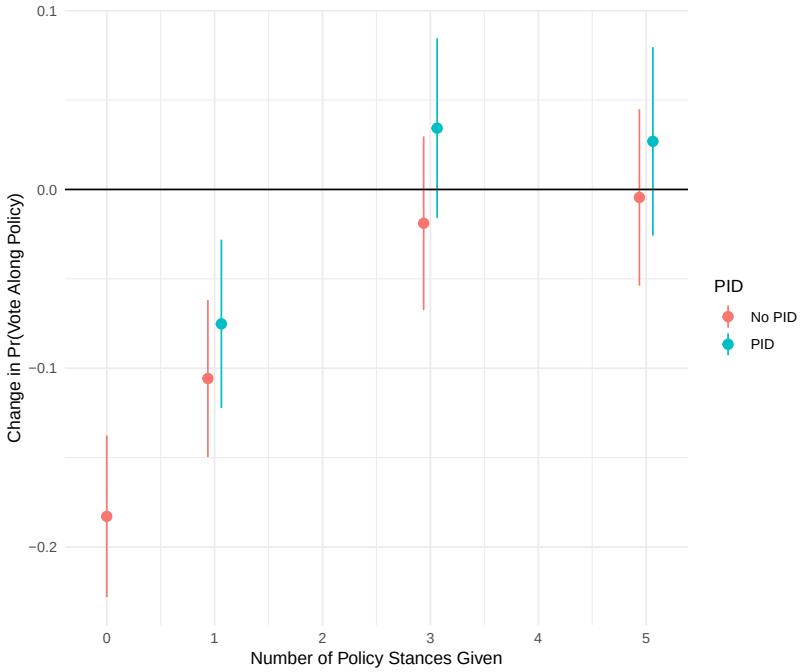


Figure A.4: PID Performance Results Using Logit Model

party ID, suggesting that party ID serves as an effective heuristic for policy preferences.

A.4 Regression Tables for Figures

Table A.1: Hypothesis 1 - Estimates and SEs

	Dependent variable:					
	prop_correct					
	(1)	(2)	(3)	(4)	(5)	(6)
treat_noPID_one	0.864***					
	(0.011)					
treat_noPID_three		0.749***				
		(0.013)				
treat_noPID_five			0.717***			
			(0.013)			
treat_PID_one				0.825***		
				(0.013)		
treat_PID_three					0.779***	
					(0.011)	
treat_PID_five						0.737***
						(0.014)
gender	0.016	−0.004	−0.008	0.009	0.025**	0.005
	(0.011)	(0.013)	(0.013)	(0.013)	(0.011)	(0.014)
race_binary	−0.032**	−0.004	−0.017	−0.032**	−0.016	0.002
	(0.014)	(0.015)	(0.016)	(0.015)	(0.013)	(0.017)
education	−0.010**	−0.006	0.0001	−0.005	0.004	0.004
	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)
age	0.001***	0.002***	0.001**	0.002***	0.001	−0.0005
	(0.0005)	(0.001)	(0.001)	(0.001)	(0.0005)	(0.001)
PID	0.012**	0.011	0.021***	0.008	0.015**	0.022***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.006)	(0.008)
Constant	−0.030	−0.068*	−0.076**	−0.053	−0.082***	−0.047
	(0.030)	(0.035)	(0.034)	(0.035)	(0.030)	(0.037)
Observations	353	352	353	350	348	352
R^2	0.948	0.910	0.904	0.928	0.936	0.896
Adjusted R^2	0.947	0.909	0.902	0.926	0.935	0.894
Residual Std. Error	0.104 (df = 346)	0.120 (df = 345)	0.119 (df = 346)	0.117 (df = 343)	0.103 (df = 341)	0.128 (df = 345)
F Statistic	1,046.309*** (df = 6; 346)	582.852*** (df = 6; 345)	543.512*** (df = 6; 346)	732.353*** (df = 6; 343)	834.854*** (df = 6; 341)	494.327*** (df = 6; 345)

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.2: Hypothesis 2 - Estimates and SEs

	Dependent variable:		
	prop_PID		
	(1)	(2)	(3)
treat_PID_one	−0.189***		
	(0.033)		
treat_PID_three		−0.177***	
		(0.035)	
treat_PID_five			−0.171***
			(0.035)
gender	0.063*	0.052	0.064*
	(0.034)	(0.035)	(0.036)
race_binary	−0.068*	−0.048	−0.084**
	(0.040)	(0.040)	(0.042)
education	0.013	−0.008	−0.002
	(0.012)	(0.012)	(0.012)
age	0.004***	0.001	0.001
	(0.001)	(0.001)	(0.001)
PID	0.075***	0.078***	0.064***
	(0.019)	(0.019)	(0.020)
Constant	0.414***	0.611***	0.626***
	(0.094)	(0.094)	(0.095)
Observations	346	345	348
R^2	0.162	0.134	0.110
Adjusted R^2	0.147	0.118	0.095
Residual Std. Error	0.310 (df = 339)	0.321 (df = 338)	0.329 (df = 341)
F Statistic	10.905*** (df = 6; 339)	8.700*** (df = 6; 338)	7.054*** (df = 6; 341)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table A.3: Hypothesis 6 - Estimates and SEs

	Dependent variable:						
	prop_correct						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
treat_noPID_zero	−0.165***						
	(0.019)						
treat_noPID_one		−0.088***					
		(0.018)					
treat_noPID_three			−0.011				
			(0.020)				
treat_noPID_five				−0.003			
				(0.020)			
treat_PID_one					−0.061***		
					(0.020)		
treat_PID_three						0.025	
						(0.020)	
treat_PID_five							0.020
							(0.021)
gender	0.022	0.047**	0.037*	0.025	0.046**	0.052**	0.033
	(0.019)	(0.019)	(0.020)	(0.020)	(0.020)	(0.020)	(0.021)
race_binary	0.035	0.033	0.019	0.008	−0.008	0.007	0.028
	(0.023)	(0.022)	(0.023)	(0.024)	(0.024)	(0.023)	(0.025)
education	0.002	−0.001	0.003	0.0004	−0.0001	0.004	0.006
	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
age	0.002**	0.002***	0.003***	0.002***	0.003***	0.002***	0.001*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
PID	0.029***	0.065***	0.062***	0.061***	0.081***	0.058***	0.063***
	(0.011)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)
Constant	0.562***	0.469***	0.444***	0.474***	0.395***	0.455***	0.480***
	(0.052)	(0.050)	(0.053)	(0.052)	(0.056)	(0.054)	(0.056)
Observations	349	350	350	349	346	345	348
R^2	0.228	0.181	0.127	0.109	0.198	0.124	0.104
Adjusted R^2	0.215	0.167	0.112	0.093	0.184	0.108	0.088
Residual Std. Error	0.178 (df = 342)	0.173 (df = 343)	0.183 (df = 343)	0.187 (df = 342)	0.184 (df = 339)	0.184 (df = 338)	0.193 (df = 341)
F Statistic	16.871*** (df = 6; 342)	12.663*** (df = 6; 343)	8.334*** (df = 6; 343)	6.952*** (df = 6; 342)	13.990*** (df = 6; 339)	7.957*** (df = 6; 338)	6.585*** (df = 6; 341)

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.4: Hypothesis 4 - Estimates and SEs

	Dependent variable:						
	posttreat.turnout						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
treat_noPID_one	1.884***						
	(0.296)						
treat_noPID_three		1.906***					
		(0.275)					
treat_noPID_five			1.960***				
			(0.281)				
treat_PID_zero				0.600*			
				(0.320)			
treat_PID_one					1.629***		
					(0.290)		
treat_PID_three						2.181***	
						(0.287)	
treat_PID_five							2.161***
							(0.283)
gender	0.515*	0.267	0.548*	−0.057	0.234	0.058	0.591**
	(0.296)	(0.278)	(0.282)	(0.323)	(0.291)	(0.292)	(0.289)
race_binary	1.113***	1.486***	1.272***	0.999***	1.271***	1.259***	0.881**
	(0.366)	(0.332)	(0.339)	(0.383)	(0.353)	(0.343)	(0.346)
education	0.011	0.055	−0.090	0.069	−0.033	0.006	−0.007
	(0.116)	(0.106)	(0.113)	(0.118)	(0.116)	(0.115)	(0.110)
age	−0.044***	−0.037***	−0.026**	−0.029**	−0.040***	−0.036***	−0.031***
	(0.012)	(0.011)	(0.011)	(0.013)	(0.012)	(0.013)	(0.012)
PID	−0.331**	−0.375**	−0.290*	−0.147	−0.326**	−0.169	−0.353**
	(0.164)	(0.152)	(0.157)	(0.175)	(0.161)	(0.159)	(0.161)
Constant	1.709**	1.374*	1.326*	0.792	1.817**	1.246	1.356*
	(0.793)	(0.745)	(0.739)	(0.847)	(0.814)	(0.779)	(0.762)
Observations	334	336	339	335	333	331	338
R^2	0.163	0.212	0.180	0.050	0.152	0.209	0.187
Adjusted R^2	0.147	0.198	0.165	0.032	0.136	0.194	0.173
Residual Std. Error	2.665 (df = 327)	2.503 (df = 329)	2.552 (df = 332)	2.901 (df = 328)	2.627 (df = 326)	2.604 (df = 324)	2.588 (df = 331)
F Statistic	10.583*** (df = 6; 327)	14.753*** (df = 6; 329)	12.156*** (df = 6; 332)	2.855** (df = 6; 328)	9.717*** (df = 6; 326)	14.276*** (df = 6; 324)	12.721*** (df = 6; 331)

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.5: Hypothesis 3 - Estimates and SEs

	Dependent variable:						
	posttreat.satisfaction						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
treat_noPID_one	2.057***						
	(0.341)						
treat_noPID_three		2.365***					
		(0.316)					
treat_noPID_five			2.821***				
			(0.320)				
treat_PID_zero				0.822**			
				(0.365)			
treat_PID_one					1.887***		
					(0.337)		
treat_PID_three						2.425***	
						(0.325)	
treat_PID_five							2.895***
							(0.326)
gender	0.441	0.502	0.490	−0.029	0.587*	0.083	0.585*
	(0.341)	(0.319)	(0.321)	(0.367)	(0.338)	(0.332)	(0.332)
race_binary	1.610***	1.554***	1.564***	1.685***	1.651***	1.553***	0.657*
	(0.417)	(0.379)	(0.386)	(0.436)	(0.410)	(0.385)	(0.397)
education	0.034	0.042	−0.169	0.033	−0.067	0.017	0.023
	(0.134)	(0.121)	(0.129)	(0.132)	(0.133)	(0.130)	(0.127)
age	−0.067***	−0.064***	−0.044***	−0.042***	−0.056***	−0.048***	−0.043***
	(0.014)	(0.013)	(0.013)	(0.015)	(0.014)	(0.015)	(0.014)
PID	−0.433**	−0.391**	−0.362**	−0.377*	−0.353*	−0.436**	−0.284
	(0.188)	(0.176)	(0.180)	(0.199)	(0.187)	(0.181)	(0.186)
Constant	6.680***	6.442***	6.586***	5.826***	6.489***	6.238***	5.676***
	(0.911)	(0.851)	(0.844)	(0.954)	(0.936)	(0.888)	(0.875)
Observations	349	349	346	345	344	340	345
R^2	0.179	0.237	0.241	0.085	0.168	0.216	0.219
Adjusted R^2	0.164	0.223	0.228	0.069	0.153	0.202	0.205
Residual Std. Error	3.138 (df = 342)	2.925 (df = 342)	2.941 (df = 339)	3.352 (df = 338)	3.099 (df = 337)	2.988 (df = 333)	3.006 (df = 338)
F Statistic	12.410*** (df = 6; 342)	17.680*** (df = 6; 342)	17.984*** (df = 6; 339)	5.231*** (df = 6; 338)	11.351*** (df = 6; 337)	15.314*** (df = 6; 333)	15.801*** (df = 6; 338)
Note:							*p<0.1; **p<0.05; ***p<0.01

Appendix B

Electoral Accountability and Exaggerated Credit Claiming

B.1 Proof of Proposition 1

In this section, I first prove the key equilibrium of analysis in the paper (i), as well as the constraints that must be satisfied for its existence (ii). I also address reasonable reconceptualizations of p, including: (iii) high-type Incumbents being open to successful challenges, and (iv) the success of challenges to claims being higher than that for promises.

B.1.1 (i) Proof of Semi-Separating Equilibrium

Since high Incumbent types always have an incentive to claim, we know that the class of semi-separating equilibria that is of interest is the one in which high types choose claim and low types mix between claim and promise. Assume that $x = \Pr[a_I = \text{Claim} \mid \theta_I = \text{Low}]$, and $(1 - x) = \Pr[a_I = \text{Promise} \mid \theta_I = \text{Low}]$.

To solve, we first consider how the Challenger updates their beliefs given this strategy profile:

- $\beta_C(\theta_I = Hi | a_I = Claim) = \frac{\mu}{\mu + x(1-\mu)}$
- $\beta_C(\theta_I = Hi | a_I = Promise) = 0$

From here, we define the expected utilities for the Challenger, and determine the value of x that would make the Challenger indifferent between choosing *Challenge* and *Promise* upon observing *Claim*. The expected utilities for the Challenger are as follows:

- EU_C(Challenge | a_I = Claim, θ_I = Hi) = $1 - q - \delta_{max}$
- EU_C(Challenge | a_I = Claim, θ_I = Lo) = $p(1 - q + \delta_{max}) + (1 - p)(1 - q - \delta_{max})$
- EU_C(Promise | a_I = Claim, θ_I = Hi) = $1 - q - \delta_{min}$
- EU_C(Promise | a_I = Claim, θ_I = Lo) = $1 - q - \delta_{min}$

Thus, the overall utilities for the Challenger, upon observing *Claim*, and given the strategy profile, are as follows:

- EU_C(Challenge | a_I = Claim, x) = $(\mu)(1)(1 - q - \delta_{max}) + (1 - \mu)(x)[p(1 - q + \delta_{max}) + (1 - p)(1 - q - \delta_{max})]$
- EU_C(Promise | a_I = Claim, x) = $1 - q - \delta_{min}$

Now we solve for the x^* that would keep the Challenger indifferent between choosing *Challenge* and choosing *Promise*:

$$(\mu)(1)(1 - q - \delta_{max}) + (1 - \mu)(x^*)[p(1 - q + \delta_{max}) + (1 - p)(1 - q - \delta_{max})] = 1 - q - \delta_{min}$$

$$\mu(1 - q - \delta_{max}) + (1 - \mu)(x^*)(p)(1 - q + \delta_{max}) + (1 - \mu)(x^*)(1 - p)(1 - q - \delta_{max}) = 1 - q - \delta_{min}$$

$$x^* = \frac{1 - q - \delta_{min} - \mu(1 - q - \delta_{max})}{(1-\mu)(p)(1-q+\delta_{max}) + (1-\mu)(1-p)(1-q-\delta_{max})}$$

$$x^* = \frac{1-q-\delta_{min}-\mu(1-q-\delta_{max})}{(1-\mu)(1-q-\delta_{max}+2p\delta_{max})}$$

To check whether this is an equilibrium, we need to check when low Incumbent types would be indifferent between choosing *Claim* and choosing *Promise*. Let y = Pr[a_C = Challenge | a_I = Claim]. The expected utilities for the Incumbent are as follows:

- EU_I(Promise | y, θ_I = Hi) = $(q+\delta_{min})$
- EU_I(Promise | y, θ_I = Lo) = $p(q-\delta_{min})+(1-p)(q+\delta_{min})$
- EU_I(Claim | y, θ_I = Hi) = $y(q+\delta_{max})+(1-y)(q+\delta_{min})$
- EU_I(Claim | y, θ_I = Lo) = $y[p(q-\delta_{max})+(1-p)(q+\delta_{max})]+(1-y)(q+\delta_{min})$

For low Incumbent types, EU_I(Promise | $\cdot$) = EU_I(Claim | $\cdot$) when:

$$p(q-\delta_{min})+(1-p)(q+\delta_{min}) = y^*[p(q-\delta_{max})+(1-p)(q+\delta_{max})]+(1-y^*)(q+\delta_{min})$$

$$p(q-\delta_{min})+(1-p)(q+\delta_{min})-q-\delta_{min} = y^*[p(q-\delta_{max})+(1-p)(q+\delta_{max})]-y^*q-y^*\delta_{min}$$

$$y^* = \frac{p(q-\delta_{min})+(1-p)(q+\delta_{min})-q-\delta_{min}}{p(q-\delta_{max})+(1-p)(q+\delta_{max})-q-\delta_{min}}$$

$$y^* = \frac{2p\delta_{min}}{2p\delta_{max}-\delta_{max}+\delta_{min}}$$

Thus, x^* and y^* denote the semi-separating equilibrium strategies for this game. □

B.1.2 (ii) Proof of Equilibrium Conditions

The value of y^* implies two constraints, since it must be the case that $y^* \in [0,1]$. First, for $y^* < 1$, it must be the case that:

- $2p\delta_{min} < 2p\delta_{max} - \delta_{max} + \delta_{min}$
- $\delta_{min}(2p-1) < \delta_{max}(2p-1)$
- This always holds, since $\delta_{max} > \delta_{min}$ by assumption.

Meanwhile, for y^* to exist (denominator $\neq 0$) and be positive, it must be the case that:

- $2p\delta_{max} - \delta_{max} + \delta_{min} > 0$
- $p > \frac{\delta_{max} - \delta_{min}}{2\delta_{max}}$

Likewise, $x^* \in [0,1]$ implies two constraints. First, for $x^* < 1$, it must be the case that:

- $1 - q - \delta_{min} - \mu(1 - q - \delta_{max}) < (1-\mu)(1 - q - \delta_{max} + 2p\delta_{max})$
- $\mu(1 - q - \delta_{max} + 2p\delta_{max} - 1 + q + \delta_{max}) < 1 - q - \delta_{max} + 2p\delta_{max} - 1 + q + \delta_{min}$
- $\mu < \frac{2p\delta_{max} - \delta_{max} + \delta_{min}}{2p\delta_{max}}$

For x^* to exist and be positive, it must be the case that:

- $(1-\mu)(1 - q - \delta_{max} + 2p\delta_{max}) > 0$
- $(1-\mu) > 0$ and $(1 - q - \delta_{max} + 2p\delta_{max}) > 0$
- The second part holds, since $(1 - q - \delta_{max}) > 0$ by assumption
- Thus, it must be the case that $\mu < 1$

In this case, the first constraint subsumes the second, since $\frac{2p\delta_{max} - \delta_{max} + \delta_{min}}{2p\delta_{max}} < 1$. □

B.1.3 (iii) Robustness to Pr[Challenge Success | Hi Type] $\neq$ 0

In terms of the specification of p, one may reasonably wonder why it is the case that challenges against high-type incumbents are never successful. It may be the case that they can be successful, but just at lower rates than that for low-type incumbents. Put more succinctly, are the equilibrium results robust to a version in which Pr[Challenge Success | Hi Type] $\neq 0$? This section walks through the model with this specification, and shows that the results do carry through. Thus, the simplification made in the paper is made without loss of generality.

For this version of the model, let p_H = Pr[Challenge Successful | a_I = Claim, θ_I = Hi], and p_L = Pr[Challenge Successful | a_I = Claim, θ_I = Lo], such that $p_H < p_L$. As before, assume that x = Pr[a_I = Claim | θ_I = Low], and $(1-x)$ = Pr[a_I = Promise | θ_I = Low].

To solve, we first consider how the Challenger updates their beliefs given the strategy profile:

- $\beta_C(\theta_I = Hi | a_I = Claim) = \frac{\mu}{\mu + x(1-\mu)}$
- $\beta_C(\theta_I = Hi | a_I = Promise) = 0$

From here, we define the expected utilities for the Challenger, and determine the value of x that would make the Challenger indifferent between choosing *Challenge* and *Promise* upon observing *Claim*. The expected utilities for the Challenger are as follows:

- EU_C(Challenge | a_I = Claim, θ_I = Hi) = $p_H(1-q+\delta_{max}) + (1-p_H)(1-q-\delta_{max})$
- EU_C(Challenge | a_I = Claim, θ_I = Lo) = $p_L(1-q+\delta_{max}) + (1-p_L)(1-q-\delta_{max})$
- EU_C(Promise | a_I = Claim, θ_I = Hi) = $1-q-\delta_{min}$
- EU_C(Promise | a_I = Claim, θ_I = Lo) = $1-q-\delta_{min}$

Thus, the overall utilities for the Challenger, upon observing *Claim*, and given the strategy profile, are as follows:

- $EU_C(\text{Challenge} \mid a_I = \text{Claim}, x) = (\mu)(p_H(1-q+\delta_{max}) + (1-p_H)(1-q-\delta_{max})) + (1-\mu)(x)[p_L(1-q+\delta_{max}) + (1-p_L)(1-q-\delta_{max})]$
- $EU_C(\text{Promise} \mid a_I = \text{Claim}, x) = 1-q-\delta_{min}$

Now we solve for the x^* that would keep the Challenger indifferent between choosing *Challenge* and choosing *Promise*:

$$(\mu)(p_H(1-q+\delta_{max}) + (1-p_H)(1-q-\delta_{max})) + (1-\mu)(x)[p_L(1-q+\delta_{max}) + (1-p_L)(1-q-\delta_{max})] = 1-q-\delta_{min}$$

$$x^* = \frac{1-q-\delta_{min}-\mu(1-q-\delta_{max}+2p_H\delta_{max})}{(1-\mu)(1-q-\delta_{max}+2p_L\delta_{max})}$$

To check whether this is an equilibrium, we need to check when low Incumbent types would be indifferent between choosing *Claim* and choosing *Promise*. Let $y = \Pr[a_C = \text{Challenge} \mid a_I = \text{Claim}]$. The expected utilities for the Incumbent are as follows:

- $EU_I(\text{Promise} \mid y, \theta_I = \text{Hi}) = p_H(q-\delta_{min}) + (1-p_H)(q+\delta_{min})$
- $EU_I(\text{Promise} \mid y, \theta_I = \text{Lo}) = p_L(q-\delta_{min}) + (1-p_L)(q+\delta_{min})$
- $EU_I(\text{Claim} \mid y, \theta_I = \text{Hi}) = y[p_H(q-\delta_{max}) + (1-p_H)(q+\delta_{max})] + (1-y)(q+\delta_{min})$
- $EU_I(\text{Claim} \mid y, \theta_I = \text{Lo}) = y[p_L(q-\delta_{max}) + (1-p_L)(q+\delta_{max})] + (1-y)(q+\delta_{min})$

For low Incumbent types, $EU_I(\text{Promise} \mid \cdot) = EU_I(\text{Claim} \mid \cdot)$ when:

$$p_L(q-\delta_{min}) + (1-p_L)(q+\delta_{min}) = y^*[p_L(q-\delta_{max}) + (1-p_L)(q+\delta_{max})] + (1-y^*)(q+\delta_{min})$$

$$y^* = \frac{2p_L\delta_{min}}{2p_L\delta_{max}-\delta_{max}+\delta_{min}}$$

To verify that the equilibrium exists when $p_H < p_L$, we must check and ensure that $x^* \in (0,1)$ when this relationship between the probabilities holds. For $x^* \in (0,1)$, it must be the case that:

$$1 - q - \delta_{min} - \mu(1 - q - \delta_{max} + 2p_H\delta_{max}) < (1-\mu)(1 - q - \delta_{max} + 2p_L\delta_{max})$$

$$\mu < \frac{\delta_{min} - \delta_{max}(1-2p_L)}{2\delta_{max}(p_L - p_H)}$$

Since we know that $\mu \in (0,1)$ by assumption, it must also be the case that: (1) $\delta_{min} - \delta_{max}(1-2p_L) > 0$, and (2) $\delta_{min} - \delta_{max}(1-2p_L) < 2\delta_{max}(p_L - p_H)$. The first case holds when $p_L > \frac{\delta_{max} - \delta_{min}}{2\delta_{max}}$, and the second case holds when $p_H < \frac{\delta_{max} - \delta_{min}}{2\delta_{max}}$. Thus, not only do we find that this equilibrium holds, but it holds iff $p_H < p_L$. □

B.1.4 (iv) Robustness to Pr[Challenge Success | Claim] ≠ Pr[Challenge Success | Promise]

One may reasonably wonder why the model equates the probability of successful challenges to claims and this probability for promises. If promises are more difficult to falsify, shouldn't the probability of successful challenges against them be lower than that against claims? As mentioned in the paper, combining these probabilities is a simplification made without loss of generality. We can disaggregate this probability, and show that the equilibrium results still hold. This section of the appendix does just that.

For this version of the model, let p_P = Pr[Challenge Successful | a_I = Promise], and p_C = Pr[Challenge Successful | a_I = Claim], such that $p_P < p_C$. As before, assume that x = Pr[a_I = Claim | θ_I = Low], and $(1-x)$ = Pr[a_I = Promise | θ_I = Low].

To solve, we first consider how the Challenger updates their beliefs given the strategy profile:

- $\beta_C(\theta_I = Hi | a_I = Claim) = \frac{\mu}{\mu + x(1-\mu)}$
- $\beta_C(\theta_I = Hi | a_I = Promise) = 0$

From here, we define the expected utilities for the Challenger, and determine the value of x that would make the Challenger indifferent between choosing *Challenge* and *Promise* upon observing *Claim*. The expected utilities for the Challenger are as follows:

- EU_C(Challenge | a_I = Claim, θ_I = Hi) = $1 - q - \delta_{max}$
- EU_C(Challenge | a_I = Claim, θ_I = Lo) = $p_C(1 - q + \delta_{max}) + (1 - p_C)(1 - q - \delta_{max})$
- EU_C(Promise | a_I = Claim, θ_I = Hi) = $1 - q - \delta_{min}$
- EU_C(Promise | a_I = Claim, θ_I = Lo) = $1 - q - \delta_{min}$

Thus, the overall utilities for the Challenger, upon observing *Claim*, and given the strategy profile, are as follows:

- EU_C(Challenge | a_I = Claim, x) = $(\mu)(1)(1 - q - \delta_{max}) + (1 - \mu)(x)[p_C(1 - q + \delta_{max}) + (1 - p_C)(1 - q - \delta_{max})]$
- EU_C(Promise | a_I = Claim, x) = $1 - q - \delta_{min}$

Now we solve for the x^* that would keep the Challenger indifferent between choosing *Challenge* and choosing *Promise*:

$$(\mu)(1)(1-q-\delta_{max}) + (1-\mu)(x^*)[p_C(1-q+\delta_{max}) + (1-p_C)(1-q-\delta_{max})] = 1-q-\delta_{min}$$

$$\mu(1 - q - \delta_{max}) + (1 - \mu)(x^*)(p_C)(1 - q + \delta_{max}) + (1 - \mu)(x^*)(1 - p_C)(1 - q - \delta_{max}) = 1 - q - \delta_{min}$$

$$x^* = \frac{1-q-\delta_{min} - \mu(1-q-\delta_{max})}{(1-\mu)(p_C)(1-q+\delta_{max}) + (1-\mu)(1-p_C)(1-q-\delta_{max})}$$

$$x^* = \frac{1-q-\delta_{min}-\mu(1-q-\delta_{max})}{(1-\mu)(1-q-\delta_{max}+2p_C\delta_{max})}$$

To check whether this is an equilibrium, we need to check when low Incumbent types would be indifferent between choosing *Claim* and choosing *Promise*. Let y = Pr[a_C = Challenge | a_I = Claim]. The expected utilities for the Incumbent are as follows:

- EU_I(Promise | y, θ_I = Hi) = $(q+\delta_{min})$
- EU_I(Promise | y, θ_I = Lo) = $p_P(q-\delta_{min})+(1-p_P)(q+\delta_{min})$
- EU_I(Claim | y, θ_I = Hi) = $y(q+\delta_{max})+(1-y)(q+\delta_{min})$
- EU_I(Claim | y, θ_I = Lo) = $y[p_C(q-\delta_{max})+(1-p_C)(q+\delta_{max})]+(1-y)(q+\delta_{min})$

For low Incumbent types, EU_I(Promise | $\cdot$) = EU_I(Claim | $\cdot$) when:

$p_P(q-\delta_{min})+(1-p_P)(q+\delta_{min}) = y^*[p_C(q-\delta_{max})+(1-p_C)(q+\delta_{max})]+(1-y^*)(q+\delta_{min})$

$p_P(q-\delta_{min})+(1-p_P)(q+\delta_{min})-q-\delta_{min} = y^*[p_C(q-\delta_{max})+(1-p_C)(q+\delta_{max})]-y^*q-y^*\delta_{min}$

$y^* = \frac{p_P(q-\delta_{min})+(1-p_P)(q+\delta_{min})-q-\delta_{min}}{p_C(q-\delta_{max})+(1-p_C)(q+\delta_{max})-q-\delta_{min}}$

$$y^* = \frac{2p_P\delta_{min}}{2p_C\delta_{max}-\delta_{max}+\delta_{min}}$$

Comparing these results with that from Proposition 1(i) and 1(ii), we see that a semi-separating equilibrium exists under similar conditions: (1) $p_C > \frac{\delta_{max}-\delta_{min}}{2\delta_{max}}$, and (2) $\mu < \frac{2p_C\delta_{max}-\delta_{max}+\delta_{min}}{2p_C\delta_{max}}$. For the major relationships of the paper to hold, it must be the case that

$\frac{\partial x^*}{\partial p_C} < 0$, and $\frac{\partial y^*}{\partial p_C} < 0$. These both clearly hold, since p_C only appears in the denominator of both values, the numerator and denominator of which are positive when the equilibrium exists. □

B.2 Proof of Proposition 2

Here, I prove both parts of Proposition 2, which argue the relationship between x^* and p, as well as y^* and p.

B.2.1 (i) Relationship between x^* and p

We have our equilibrium value for x^* below:

$$x^* = \frac{1-q-\delta_{min}-\mu(1-q-\delta_{max})}{(1-\mu)(1-q-\delta_{max}+2p\delta_{max})}$$

To determine the relationship between x^* and p, we take the first derivative of x^* with respect to p:

$$\frac{\partial x^*}{\partial p} = \frac{-2\delta_{max}(1-q-\delta_{min}-\mu(1-q-\delta_{max}))}{(1-q-\delta_{max}+2p\delta_{max})^2(1-\mu)}$$

If $\mu > \frac{(1-q-\delta_{min})}{(1-q-\delta_{max})}$, then x^* is increasing in p. Otherwise, x^* is decreasing in p. Since we know that $\delta_{max} > \delta_{min}$ by assumption, this means that $\frac{(1-q-\delta_{min})}{(1-q-\delta_{max})} > 1$. Since $\mu < 1$ by assumption, this means that our inequality of interest never holds, and thus x^* is decreasing with p. This means that, as the likelihood of successful challenges increases, low incumbent types are less likely to pick claim. □

B.2.2 (ii) Relationship between y^* and p

We have our equilibrium value for y^* below:

$$y^* = \frac{2p\delta_{min}}{2p\delta_{max}-\delta_{max}+\delta_{min}}$$

To determine the relationship between y^* and p, we take the first derivative of y^* with respect to p:

$$\frac{\partial y^*}{\partial p} = \frac{-2\delta_{min}(\delta_{max} - \delta_{min})}{((2p-1)\delta_{max} + \delta_{min})^2}$$

Since the denominator is always positive, $(\delta_{max} - \delta_{min}) > 0$ by assumption, and $\delta_{min} > 0$ by assumption, this value is always negative. □

Appendix C

Measuring Bureaucratic Responsiveness Using Text from the Appropriations Process

C.1 Data Collection Protocols

The following directions were given to each researcher that was involved in the data collection process.

Project Purpose

This project seeks to understand the relationship between the wording and tone of congressional appropriations reports and compliance within executive agencies. As appropriations committees determine the level of funding for the agencies of government, they draft a bill that, if passed, continues along the legislative process. Alongside this bill, the committees also write a report that details various aspects of the funding given, which can include: (i) the rationale behind the funding level, (ii) restrictions on the funding provided, (iii) directives tied to the funding, and/or (iv) praises or reprimands directed at the organi-

zation.

Although only the bill has the force of law, Kirst (1969) and Schick (2008) have found that the wording in appropriations reports affects whether agencies comply with legislative committees. The research on this phenomenon has been limited and largely qualitative in nature, and this project seeks to determine whether there is quantitative support behind this relationship. This will be done by examining: (i) the word count of committee report directives and appraisals directed towards agencies, and (ii) the tone of the wording directed towards agencies.

Project Role

At this point, we are in the data collection phase of this project. This involves skimming through the House and Senate appropriations reports for each congressional committee (1998-2017), and recording the appropriations levels and word counts of directives/appraisals for executive agencies and bureaus. This information will be recorded in an Excel document.

Project Procedure

Accessing the Reports

1. Go to the following website:
https://www.congress.gov/resources/display/content/Appropriations+and+Budget

2. Click on the appropriate fiscal year (FY).

3. Click on the link “Regular Appropriations.” You may need to scroll down.

4. You will see the Appropriations Committee Reports within the table that opens. We

will be collecting data from all of the committees except "Legislative Branch," starting with the House committees and then continuing to the Senate committees.

5. When you open a committee report, you will be taken to a webpage version of the report.

a. It is beneficial to open a separate tab with the summary table of the appropriations (most often, but not always, located at the end of the PDF version of the report). This will help you to distinguish between agencies, bureaus, and accounts, as well as to easily access the appropriations amounts. If you open this as a separate PDF file, you will be able to rotate the document for easier reading of the charts.

b. You may use the webpage text or the pdf version of the report, but be sure to ensure that your information is copying accurately in Excel and Word.

Collecting Data - Regular Procedure

For most agencies and bureaus, the reports follow a set structure, making it simple to skim the report and collect data. The examples below are taken from the 2016 THUD House Report.

1. Skip the introduction and begin at the body of the report, this is most easily done by finding "Title I" of the report. The picture below represents the start of the body of this report:

TITLE I--DEPARTMENT OF TRANSPORTATION

Office of the Secretary

SALARIES AND EXPENSES

Appropriation, fiscal year 2015	$105,000,000
Budget request, fiscal year 2016	113,657,000
Recommended in the bill	105,000,000
Bill compared with:	
Appropriation, fiscal year 2015	- - -
Budget request, fiscal year 2016	-8,657,000

COMMITTEE RECOMMENDATION

The bill provides $105,000,000 for the salaries and expenses of the offices comprising the Office of the Secretary of Transportation (OST). The Committee's recommendation is the same as the 2015 enacted level and $8,657,000 below the request. The Committee's recommendation includes individual funding for each of these offices as has been done in prior years. The following table (dollars in thousands) compares the fiscal year 2015 enacted level to the fiscal year 2016 budget request and the Committee's recommendation by office. The Committee strongly urges the Department to manage hiring and attrition in 2015 to meet these levels for 2016. Reductions are also encouraged in the areas of travel and contracts.

2. For this report, the first agency listed is the Department of Transportation, and the bureau is the Office of the Secretary. Copy this information into the respective columns in the Excel sheet.

3. Refer to the summary table to quickly locate the aggregate appropriations for this bureau. On page 154 of this report, the amounts are: 802,624,000 for FY 2015, 1,612,387,000 for FY 2016 Request, and 400,413,000 for FY 2016 Bill. Record these numbers in the appropriate columns in the Excel sheet (note that table numbers are amounts in thousands).

4. To begin recording the word count, skim through this section of the report for language that indicates directives or appraisals. For example, in the image above, you would copy: "The Committee strongly urges the Department to manage hiring and attrition in 2015 to meet these levels for 2016. Reductions are also encouraged in the areas of travel and contracts." Copy this text and paste it into a Word document to keep an ongoing word count.

5. General considerations for selecting wording:

a. The following phrases tend to signal sections of the report to copy and include:

Satisfaction/Dissatisfaction	Binding/Non-binding
Is encouraged…	Is directed/Directs…
Is frustrated…	Encourages…
Believes…	Assumes…
Is/Remains Concerned…	Has included language…
Recognizes…	Advises…
Notes…	(Strongly) Encourages…
Understands…	Believes…
Reminds…	Reminds…
Is supportive of/Supports…	Expects/Is expected…
Appreciates…	Urges…
Is pleased…	Requires…
Is troubled…	Would benefit from…
Is aggrieved by…	Includes the requirement that…
Is disturbed by…	
Applauds…	
Agrees/Disagrees…	

b. Sections in which the committee breaks down appropriations more specifically than the bill should also be included.

c. Sections in which the committee mentions that they are **not funding** a program should also be included.

d. When in doubt, compare the language of the report to that in the bill, and copy language that goes beyond what is in the bill.

6. Once you have a complete word count, enter that number in the column titled "Word Count." Copy the entire word selection and paste that into the column titled "Word Text."

a. If the text entry is long, (typically beyond 4000+ words – you will notice that the entry will not paste within Excel), the text will need to be entered in a txt file. Save that file using a title that includes the year, committee, and agency/bureau, and type that file name in the

"Word Text" column.

7. In addition to copying words in the report, copy tables that specify how appropriations are to be spent. Record the word count of these tables in the excel spreadsheet.

8. When in doubt as to whether an entry is an agency, a bureau, or an account, consult the layout of the tables, as well as the wording in the report.

Collecting Data - Exceptions to Procedures

1. For each fiscal year, the Department of Defense (DoD) committee will appear in the chart as only one entry. Take note that DoD entries will appear for other committees.

2. Some agencies (for example, FDA) appear in multiple committees in a given year. Keep these as separate entries in the Excel document.

3. For the "State, Foreign Operations, and Related Programs" committee, only gather data for the following agencies/programs:

a. Department of State (usually the only separated bureau is the inspector general)
b. US Agency for International Development
c. Overseas Private Investment Corporation
d. Export-Import Bank of the United States
e. Inter-American Foundation
f. African Development Foundation
g. Asia Foundation
h. US Institute of Peace
i. Broadcasting Board of Governors

j. International Boundary Water Commission

k. Commission for the Preservation of America's Heritage Abroad

l. US Commission on International Religious Freedom

m. Commission on Security and Cooperation in Europe

n. US-China Economic and Security Review Commission

o. Peace Corps

p. Millennium Challenge Corporation

q. Trade and Development Agency

4. The "Executive Office of the President" will appear throughout different committees for each fiscal year. Altogether, the bureaus that you should collect data from include:

a. Office of Management and Budget

b. Office of National Drug Control Policy

c. Office of Science and Technology Policy

d. Council on Environmental Quality

5. You can skip any appropriations that are given to the legislative or judicial branches, as this project will only be focused on executive agencies.

C.2 Logistic Regression Weights for Ensemble

This table displays the results of the logistic regression used to weight the labels from each algorithm in the ensemble.

Table C.1: Logistic Regression to Weight Algorithm Codes

	Model 1
Support Vector Machine	1.307
	(0.730)
Maximum Entropy	0.104
	(0.405)
SLDA	0.658
	(0.471)
Boosting	3.038
	(0.595)
Bagging	-1.420
	(0.753)
Random Forest	1.801
	(0.503)
Regression Tree	0.520
	(1.043)
N	2,200

C.3 Additional Examples of Ensemble Scoring

Table C.2 displays the ensemble scoring for an excerpt of the 2016 Senate Interior and Environment appropriations subcommittee, addressing the Environmental Protection Agency.

Table C.2: Ensemble Scores for 2016 Senate IE Excerpt

Sentence	Score
This Committee is extremely concerned with the Agency's rulemaking regarding health and environmental protection standards for uranium and thorium mill tailings (Part 192).	0
The Committee is concerned the Agency has failed to justify this rulemaking or consult with State regulators or the industry that will be regulated by this rule.	0
The Agency acknowledges it is not aware of a single example of groundwater contamination caused by an in-situ uranium recovery project.	1
The Committee encourages EPA to withdraw this rulemaking and instead work with State regulators, the Nuclear Regulatory Commission, and the uranium recovery industry to collect sufficient data to determine if any updates are needed to the existing generally applicable standards.	1

Note: 0 indicates "non-responsive," while 1 indicates "responsive" or "no valence"

Table C.3 displays the ensemble scoring for an excerpt of the 2010 House Labor, Health and Human Services, and Education appropriations subcommittee, addressing the Occupational Safety and Health Administration (OSHA).

Table C.4 displays the ensemble scoring for an excerpt of the 2009 Homeland Security appropriations subcommittee, addressing Immigration and Customs Enforcement (ICE).

Table C.3: Ensemble Scores for 2010 House LHHSE Excerpt

Sentence	Score
The Committee is concerned about a 2009 audit by the DOL Inspector General (IG) that identifies serious deficiencies in the implementation of OSHA's Enhanced Enforcement Program (EEP) from 2003 through 2008.	0
The audit states that OSHA did not comply with its own program requirements in 97 percent of cases sampled by the IG.	0
The worst failures included a lack of follow-up inspections at worksites that had been cited for violations to ensure that problems had been corrected, and the lack of any inspections at related company-wide worksites even when safety and health issues indicated that workers at these sites were at risk for serious injuries or death.	0
In addition, the previous administration issued a revised directive in 2008 that sharply reduced the number of serious cases being referred to the EEP.	0

Note: 0 indicates "non-responsive," while 1 indicates "responsive" or "no valence"

Table C.4: Ensemble Scores for 2009 House HS Excerpt

Sentence	Score
The Committee is concerned about recent reports issued by the OIG and GAO concerning the Department's failure to comply with standards for providing safe, secure and humane treatment of those detained in ICE custody.	0
The Committee directs the ICE Office of Professional Responsibility to continue to expand its detention oversight programs in 2009.	1

Note: 0 indicates "non-responsive," while 1 indicates "responsive" or "no valence"

Bibliography

Abramowitz, A.I. and K.L. Saunders. 2008. "Is Polarization a Myth?" *The Journal of Politics* 70(2):542–555.

Abramowitz, Alan. 2010. *The Disappearing Center: Engaged Citizens, Polarization, and American Democracy*. Yale University Press.

Abramowitz, Alan. 2012. *The Polarized Public? Why American Government is so Dysfunctional*. Pearson Higher Ed.

Achen, Christopher H and Larry M Bartels. 2004. "Blind retrospection: Electoral responses to drought, flu, and shark attacks.".

Adcock, Robert and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95(3):529–546.

Althaus, S. L. 2003. *Collective Preferences in Democratic Politics.* Cambridge University Press.

Angrist, J. D. and J. S. Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press.

Ansolabehere, Stephen, Jonathan Rodden and James M Snyder. 2008. "The strength of issues: Using multiple measures to gauge preference stability, ideological constraint, and issue voting." *American Political Science Review* 102(2):215–232.

Arnold, R. Douglas. 1979. *Congress and the Bureaucracy*. New Haven, CT: Yale University Press.

Aronow, P. M. and B. T. Miller. 2015. *Theory of Agnostic Statistics*. Yale University.

Barber, Michael and Jeremy C Pope. 2019. "Does party trump ideology? Disentangling party and ideology in America." *American Political Science Review* 113(1):38–54.

Bendor, Jonathan and Adam Meirowitz. 2004. "Spatial Models of Delegation." *American Political Science Review* 98(2):293–310.

Bergan, D. E., A. S. Gerber, D. P. Green and C. Panagopoulos. 2005. "Grassroots mobilization and voter turnout in 2004." *Public Opinion Quarterly* 69(5):760–777.

Berinsky, A. J., G. A. Huber and G. S. Lenz. 2012. "Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk." *Political Analysis* 20(3):351–368.

Bickers, Kenneth N and Robert M Stein. 1996. "The electoral dynamics of the federal pork barrel." *American Journal of Political Science* pp. 1300–1326.

Bolton, Alexander and Sharece Thrower. 2019. "The Constraining Power of the Purse: Legislative Appropriations and Executive Discretion." *Journal of Politics* .

Brady, Henry E and Paul M Sniderman. 1985. "Attitude attribution: A group basis for political reasoning." *American Political Science Review* 79(4):1061–1078.

Burkov, Andriy. 2019. *The Hundred-Page Machine Learning Book*. Quebec City, Canada: Andriy Burkov.

Cain, Bruce, John Ferejohn and Morris Fiorina. 2013. The personal vote. In *The Personal Vote*. Harvard University Press.

Campbell, Angus, Philip E Converse, Warren E Miller and Donald E Stokes. 1980. *The american voter*. University of Chicago Press.

Carpenter, Daniel and George A. Krause. 2014. "Transactional Authority and Bureaucratic Politics." *Journal of Public Administration Research and Theory* 25(2):5–25.

Carpenter, Daniel P. 2001. *The Forging of Bureaucratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies, 1862-1928*. Princeton, NJ: Princeton University Press.

Carpini, Michael X Delli and Scott Keeter. 1996. *What Americans know about politics and why it matters*. Yale University Press.

Claassen, Ryan L, David B Magleby, J Quin Monson and Kelly D Patterson. 2008. ""At Your Service" Voter Evaluations of Poll Worker Performance." *American Politics Research* 36(4):612–634.

Converse, Philip E. 2006. "The nature of belief systems in mass publics (1964)." *Critical review* 18(1-3):1–74.

Copeland, Curtis W. 2008. Congressional Influence on Rulemaking and Regulation Through Appropriations Restrictions. Technical Report RL34354 Congressional Research Service Washington, DC: .

Dancey, Logan and Geoffrey Sheagley. 2013. "Heuristics behaving badly: Party cues and voter knowledge." *American Journal of Political Science* 57(2):312–325.

Dodd, Lawrence C. and Richard L. Schott. 1979. *Congress and the Administrative State*. New York, NY: John Wiley & Son, Inc.

Downs, Anthony et al. 1957. "An economic theory of democracy.".

Druckman, James N, Donald P Green, James H Kuklinski and Arthur Lupia. 2006. "The growth and development of experimental research in political science." *American Political Science Review* 100(4):627–635.

Durant, Robert F. 1995. "Public Policy, Overhead Democracy, and the Professional State Revisited." *Administration & Society* 27(2):165–202.

Dziuda, Wioletta and William G Howell. 2021. "Political scandal: a theory." *American Journal of Political Science* 65(1):197–209.

E., Levay K., J. Freese and J. N. Druckman. 2016. "The Demographic and Political Composition of Mechanical Turk Samples." *Sage Open* 6(1).

Eisner, Marc Allen and Kenneth J. Meier. 1990. "Presidential Control versus Bureaucratic Power: Explaining the Reagan Revolution in Antitrust." *American Journal of Political Science* 34(1):269–287.

Enelow, James M and Melvin J Hinich. 1984. *The spatial theory of voting: An introduction*. CUP Archive.

Epstein, David and Peter Zemsky. 1995. "Money talks: Deterring quality challengers in congressional elections." *American Political Science Review* 89(2):295–308.

Epstein, David and Sharyn O'Halloran. 1999. *Delegating Powers: A Transaction Cost Politics Approach to Policy Making under Separate Powers*. Cambridge University Press.

Erikson, Robert S. 1971. "The advantage of incumbency in congressional elections." *Polity* 3(3):395–405.

Evans, Diana. 1994. "Policy and pork: the use of pork barrel projects to build policy coalitions in the House of Representatives." *American Journal of Political Science* pp. 894–917.

Fearon, James D. 1999. "Electoral accountability and the control of politicians: selecting good types versus sanctioning poor performance." *Democracy, accountability, and representation* 55:61.

Fenno, Richard F. 1966. *The Power of the Purse: Appropriations Politics in Congress*. New York, NY: Little, Brown, and Company.

Fenno, Richard F. 1978. *Home Style: House Members in their Home Districts*. Harper-Collins.

Ferejohn, John A. 1974. *Pork barrel politics: Rivers and harbors legislation, 1947-1968*. Stanford University Press.

Fiorina, Morris P, Samuel J Abrams et al. 2008. "Political polarization in the American public." *ANNUAL REVIEW OF POLITICAL SCIENCE-PALO ALTO-* 11:563.

Franchino, Fabio and Francesco Zucchini. 2015. "Voting in a multi-dimensional space: A conjoint analysis employing valence and ideology attributes of candidates." *Political science research and methods* 3(2):221–241.

Gailmard, Sean. 2012. "Accountability and principal-agent models." *Chapter prepared for the Oxford Handbook of Public Accountability* .

Gailmard, Sean and John W Patty. 2019. "Giving advice versus making decisions: Transparency, information, and delegation." *Political Science Research and Methods* 7(3):471–488.

Galston, William A. 2001. "Political knowledge, political engagement, and civic education." *Annual review of political science* 4(1):217–234.

Gates, Scott and Brian D Humes. 1997. *Games, information, and politics: Applying game theoretic models to political science*. University of Michigan Press.

Gerber, Alan S and Donald P Green. 2000. "The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment." *American political science review* 94(3):653–663.

Gerber, Alan S and Donald P Green. 2005. "Do phone calls increase voter turnout? An update." *The Annals of the American Academy of Political and Social Science* 601(1):142–154.

Gerber, Alan S, Donald P Green and Christopher W Larimer. 2008. "Social pressure and voter turnout: Evidence from a large-scale field experiment." *American political Science review* 102(1):33–48.

Gomez, Brad T, Thomas G Hansford and George A Krause. 2007. "The Republicans should pray for rain: Weather, turnout, and voting in US presidential elections." *The Journal of Politics* 69(3):649–663.

Goodfellow, Ian, Yoshua Bengio and Aaron Courville. 2016. *Deep Learning*. Cambridge, MA: MIT Press.

Goodliffe, Jay. 2005. "When do war chests deter?" *Journal of Theoretical Politics* 17(2):249–277.

Gordon, Sanford C. 2011. "Politicizing Agency Spending Authority: Lessons from a Bush-era Scandal." *American Political Science Review* 105(4):717–734.

Gordon, Sanford C, Gregory A Huber and Dimitri Landa. 2007. "Challenger entry and voter learning." *American Political Science Review* 101(2):303–320.

Green, Donald P, Bradley Palmquist and Eric Schickler. 2004. *Partisan hearts and minds: Political parties and the social identities of voters*. Yale University Press.

Grimmer, Justin, Sean J Westwood and Solomon Messing. 2014. The impression of influence. In *The Impression of Influence*. Princeton University Press.

Grimmer, Justin, Solomon Messing and Sean J Westwood. 2012. "How words and money cultivate a personal vote: The effect of legislator credit claiming on constituent credit allocation." *American Political Science Review* 106(4):703–719.

Haeder, Simon F. and Susan Webb Yackee. 2015. "Influence and the Administrative PRocess: Lobbying the U.S. President's Office of Management and Budget." *American Political Science Review* 109(3):507–522.

Hainmueller, Jens and Daniel J Hopkins. 2015. "The hidden American immigration consensus: A conjoint analysis of attitudes toward immigrants." *American Journal of Political Science* 59(3):529–548.

Hainmueller, Jens, Daniel J Hopkins and Teppei Yamamoto. 2014. "Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments." *Political analysis* 22(1):1–30.

Hammond, Thomas H. and Jack H. Knott. 1996. "Who Controls the Bureaucracy?: Presidential Power, Congressional Dominance, Legal Constraints, and Bureaucratic Autonomy in a Model of Multi-Institutional Policy-Making." *Journal of Law, Economics, and Organization* 21(1):119–166.

Hara, Kotaro, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch and Jeffrey P Bigham. 2018. A data-driven analysis of workers' earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. pp. 1–14.

Harsanyi, John C. 1967. "Games with incomplete information played by "Bayesian" players, I–III Part I. The basic model." *Management science* 14(3):159–182.

Healy, Andrew, Neil Malhotra et al. 2010. "Random events, economic losses, and retrospective voting: Implications for democratic competence." *Quarterly Journal of Political Science* 5(2):193–208.

Hill, Seth. 2022. *Frustrated Majorities: How Issue Intensity Enables Smaller Groups of Voters to Get What They Want*. Cambridge University Press.

Hollibaugh, Gary E. Jr. 2019. "The Use of Text as Data Methods in Public Administration: A Review and an Application to Agency Priorities." *Journal of Public Administration Research and Theory* 29(3):474–490.

Huber, Gregory A, Seth J Hill and Gabriel S Lenz. 2012. "Sources of bias in retrospective decision making: Experimental evidence on voters' limitations in controlling incumbents." *American Political Science Review* 106(4):720–741.

Huddy, Leonie and Nayda Terkildsen. 1993. "Gender stereotypes and the perception of male and female candidates." *American journal of political science* pp. 119–147.

Huff, Connor and Dustin Tingley. 2015. ""Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents." *Research & Politics* 2(3):2053168015604648.

Ingraham, Patricia W. and Carolyn Ban. 1988. "Politics and Merit: Can They Meet in a Public Service Model?" *Review of Public Personnel Administration* 8(1):7–19.

Jessee, Stephen A. 2012. *Ideology and spatial voting in American elections*. Cambridge University Press.

Jurka, Timothy P., Loren Collingwood, Amber E. Boydstun, Emiliano Grossman and Wouter van Atteveldt. 2013. "RTextTools: A Supervised Learning Package for Text Classification." *The R Journal* 5(1):6–12.

Key, Valdimer Orlando et al. 1966. *The responsible electorate*. Vol. 10 Harvard University Press Cambridge, MA.

Kinder, Donald R. 1998. "Opinion and Action in the Realm of Politics.".

Kirkland, Patricia A and Alexander Coppock. 2018. "Candidate choice without party labels." *Political Behavior* 40(3):571–591.

Kirst, W. Michael. 1969. *Government without Passing Laws*. Chapel Hill, NC: University of North Carolina Press.

Kramer, Gerald H. 1977. "A dynamical model of political equilibrium." *Journal of Economic Theory* 16(2):310–334.

Krause, George A. 1999. *A Two-Way Street: The Institutional Dynamics of the Modern Administrative State*. Pittsburgh, PA: University of Pittsburgh Press.

Krause, George A., David E. Lewis and James W. Douglas. 2006. *Political Appointments, Civil Service Systems, and Bureaucratic Competence: Organizational Balancing and Executive Branch Revenue Forecasts in the American States*. Vol. 50.

Kriner, Douglas L and Eric Schickler. 2016. *Investigating the President: Congressional Checks on Presidential Power*. Princeton University Press.

Lago, Ignacio and André Blais. 2019. "Early voting and satisfaction with vote choice." *Representation* 55(2):225–233.

Lau, Richard R and David P Redlawsk. 2001. "Advantages and disadvantages of cognitive heuristics in political decision making." *American journal of political science* pp. 951–971.

Lau, Richard R and David P Redlawsk. 2008. "Older but wiser? Effects of age on political cognition." *The Journal of Politics* 70(1):168–185.

Layman, Geoffrey C, Thomas M Carsey and Juliana Menasce Horowitz. 2006. "Party polarization in American politics." *Annual Review of Political Science* 9:83–110.

Lazarsfeld, Paul F, Bernard Berelson and Hazel Gaudet. 1968. The people's choice. In *The people's choice*. Columbia University Press.

Lazarus, Jeffrey and Shauna Reilly. 2010. "The electoral benefits of distributive spending." *Political Research Quarterly* 63(2):343–355.

Levitt, Steven D and James M Snyder Jr. 1997. "The impact of federal spending on House election outcomes." *Journal of political Economy* 105(1):30–53.

Lewis, David E. 2008. *The Politics of Presidential Appointments: Political Control and Bureaucratic Performance*. Princeton, NJ: Princeton University Press.

Lodge, Milton, Marco R Steenbergen and Shawn Brau. 1995. "The responsive voter: Campaign information and the dynamics of candidate evaluation." *American political science review* 89(2):309–326.

Lowande, Kenneth. 2018*a*. "Politicization and Responsiveness in Executive Agencies." *The Journal of Politics* 81(1):33–48.

Lowande, Kenneth. 2018*b*. "Who Policies the Administrative State?" *American Political Science Review* 112(4):874–890.

Lowi, Theodore J. 1969. *The End of Liberalism: The Second Republic of the United States*. New York, NY: Norton.

Lupia, Arthur. 1994. "Shortcuts versus encyclopedias: Information and voting behavior in California insurance reform elections." *American Political Science Review* 88(1):63–76.

Lupia, Arthur, Mathew D McCubbins, Lupia Arthur et al. 1998. *The democratic dilemma: Can citizens learn what they need to know?* Cambridge University Press.

MacDonald, Jason A. 2010. "Limitation riders and congressional influence over bureaucratic policy decisions." *American Political Science Review* 104(4):766–782.

Marvel, John D. and Robert J. McGrath. 2016. "Congress as Manager: Oversight Hearings and Agency Morale." *Journal of Public Policy* 36(3):489–520.

Mashaw, Jerry L. 1983. *Bureaucratic Justice: Managing Social Security Disability Claims*. New Haven, CT: Yale University Press.

Mayhew, David R. 1974. *Congress: The Electoral Connection*. New Haven, CT: Yale University Press.

McDermott, Monika L. 1998. "Race and gender cues in low-information elections." *Political Research Quarterly* 51(4):895–918.

Merrifield, John. 1993. "The institutional and political factors that influence voter turnout." *Public Choice* 77(3):657–667.

Miller, George A. 1956. "The magical number seven, plus or minus two: Some limits on our capacity for processing information." *Psychological review* 63(2):81.

Moe, Terry M. 1984. "The New Economics of Organization." *American Journal of Political Science* 28(4):739–777.

Moe, Terry M. 1985*a*. "Control and Feedback in Economic Regulation: The Case of the NLRB." *American Political Science Review* 79(4):1094–1116.

Moe, Terry M. 1985*b*. The Politicized Presidency. In *The New Direction in American Politics*, ed. John E. Chubb and Paul E. Peterson. Washington, DC: Brookings Institution Press pp. 235–271.

Mutz, Diana C. 2011. Population-based survey experiments. In *Population-Based Survey Experiments*. Princeton University Press.

Nathan, Richard P. 1983. *The Administrative Presidency*. Hoboken, NJ: Wiley.

Nicholson, Stephen P. 2012. "Polarizing cues." *American journal of political science* 56(1):52–66.

Niskanen, William A. 1971*a*. *Bureaucracy and Representative Government*. Aldine.

Niskanen, William A. 1971*b*. *Bureaucracy and Representative Government*. Chicago, IL: Aldine-Atherton.

Ogden, Benjamin and Alejandro Medina. 2020. "Strategic Opposition Research." *Available at SSRN 3631386* .

Panagopoulos, Costas, Christopher W Larimer and Meghan Condon. 2014. "Social pressure, descriptive norms, and voter mobilization." *Political Behavior* 36(2):451–469.

Petersen, Michael Bang. 2015. "Evolutionary political psychology: On the origin and structure of heuristics and biases in politics." *Political Psychology* 36:45–78.

Polsby, Nelson W, Aaron B Wildavsky and David A Hopkins. 2008. *Presidential elections: Strategies and structures of American politics*. Rowman & Littlefield.

Poole, Keith T and Howard Rosenthal. 1984. "The polarization of American politics." *The journal of politics* 46(4):1061–1079.

Popkin, Samuel L and Samuel L Popkin. 1991. *The reasoning voter: Communication and persuasion in presidential campaigns*. University of Chicago Press.

Potter, Rachel Augustine. 2019. *Bending the Rules: Procedural Politicking in the Bureaucracy*. Chicago, IL: University of Chicago Press.

Rahn, Wendy M. 1993. "The role of partisan stereotypes in information processing about political candidates." *American Journal of Political Science* pp. 472–496.

Randall, Ronald. 1979. "Presidential Power versus Bureaucratic Intransigience: The Influence of the Nixon Administration on Welfare Policy." *American Political Science Review* 73(3):795–810.

Resh, William G. 2015. *Rethinking the Administrative Presidency: Trust, Intellectual Capital, and Appointee-Careerist Relations in the George W. Bush Administration*. Baltimore, MD: Johns Hopkins University Press.

Rudalevige, Andrew. 2002. *Managing the President's Program: Presidential Leadership and Legislative Policy Formulation*. Princeton, NJ: Princeton University Press.

Saltzstein, Grace Hall. 1992. "Bureaucratic Responsiveness: Conceptual Issues and Current Research." *Journal of Public Administration Research and Theory* 2(1):63–88.

Schick, Allen. 2008. *The federal budget: Politics, policy, process*. Brookings Institution Press.

Shotts, Kenneth and Scott Ashworth. 2011. Challengers, democratic contestation, and electoral accountability. In *APSA 2011 Annual Meeting Paper*.

Stein, Robert M and Kenneth N Bickers. 1994. "Congressional elections and the pork barrel." *The Journal of Politics* 56(2):377–399.

Tomz, Michael and Robert P Van Houweling. 2008. "Candidate positioning and voter choice." *American Political Science Review* 102(3):303–318.

Tversky, Amos and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty." *science* 185(4157):1124–1131.

Washington, George. 1796. "Farewell address." *September* 19(1796):1931–1944.

Weingast, Barry R. and Mark J. Moran. 1983. "Bureaucratic Discretion or Congressional Control? Regulatory Policymaking by the Federal Trade Commission." *Journal of Political Economy* 91(5):765–800.

Weko, Thomas J. 1995. *The Politicizing Presidency: The White House Personnel Office*. Lawrence, KS: University of Kansas Press.

Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York, NY: Basic Books.

Wood, B. Dan and Richard W. Waterman. 1991. "The Dynamics of Political Control of the Bureaucracy." *American Political Science Review* 85(3):801–828.

Wood, B. Dan and Richard W. Waterman. 1994. *Bureaucratic Dynamics: The Role of Bureaucracy in a Democracy*. Boulder, CO: Westview Press.

Workman, Samuel. 2008. *The Dynamics of Bureaucracy in the U.S. Government: How Congress and Federal Agencies Process Information and Solve Problems*. Cambridge, UK: Cambridge University Press.