Democracy in a Dim Light: Milquetoast Local Newspapers, Votes for Only Looking the Part, and Online News Cycles

by

Michael Colin Dougal

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Political Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Gabriel Lenz, Chair
Professor Eric Schickler
Professor Jasjeet Sekhon
Professor Kevin Quinn

Spring 2017

Democracy in a Dim Light: Milquetoast Local Newspapers, Votes for Only Looking the Part, and Online News Cycles

Abstract

Democracy in a Dim Light: Milquetoast Local Newspapers, Votes for Only Looking the Part, and Online News Cycles

by

Michael Colin Dougal

Doctor of Philosophy in Political Science

University of California, Berkeley

Associate Professor Gabriel Lenz, Chair

This dissertation examines the ability of the media to monitor politicians and the ability of voters to acquire politically relevant information. The dissertation is primarily made up of three separate papers. The first paper (Chapter 2) asks why citizens routinely fail to vote out of step representatives out of office and what institutions can help voters hold politicians accountable. To the extent that politicians exploit voters' lack of information to win at the ballot box despite shirking in Congress, the press could foster democratic accountability by sounding the alarm on out of step representatives and alerting otherwise inattentive voters that it is time for change. In the first paper in my dissertation I collect an original dataset of local newspaper coverage of candidates in the 2010 House election in order to find out whether newspapers play this role for voters and act as a watchdog of incumbent representatives. After working with research assistants to provide human classification of a random subset of these articles, I use a text as data machine learning approach to measure the content of the much larger volume of articles that we cannot read. After validating an ensemble "SuperLearner" by demonstrating out-of-sample classification accuracy that for many features approaches human inter-coder agreement, I show that challengers receive less coverage than incumbents in competitive districts, horse race coverage displaces policy coverage, and newspapers do not sound the alarm on out of step incumbents. Newspapers do provide a whiff of scandal when representatives are referred to the House Ethics Committee for potential ethics violations, but they do not criticize representatives accused of some form of corruption at significantly higher rates. Even in congressional districts that closely correspond to newspaper markets, journalists act as neither watchdog nor lapdog, but instead provide overwhelmingly neutral coverage, failing to criticize incumbents who vote against a majority of their constituents on landmark legislation.

The second paper (Chapter 3) provides experimental evidence that candidate appearance influences vote choice. According to numerous studies, candidates' looks predict voters' choices—a finding that raises concerns about voter competence and about the quality of elected officials. This potentially worrisome finding, however, is observational and therefore

vulnerable to alternative explanations. To better test the appearance effect, we conducted two experiments. Just before primary and general elections for various offices, we randomly assigned voters to receive ballots with and without candidate photos. Simply showing voters these pictures increased the vote for appearance-advantaged candidates. Experimental evidence therefore supports the view that candidates' looks could influence some voters. In general elections, we find that high-knowledge voters appear immune to this influence, while low-knowledge voters use appearance as a low-information heuristic. In primaries, however, candidate appearance influences even high-knowledge and strongly partisan voters.

The third paper (Chapter 4) examines which major events in the 2016 U.S. presidential campaign saw only a brief spike in coverage and which became a more permanent feature of campaign coverage. In particular, I analyze coverage of six major events in the presidential campaign to test the hypothesis that news outlets of all persuasions will cover major events as news, but only partisan outlets will continue to discuss negative stories about their opponents long after the event that made the topic news. Broadly, I find that all outlets do indeed pick up major stories temporarily, but that the more traditional news organization in my study does not stick with a higher level of coverage of any topic after a seven day window following the related event. Partisan outlets, in contrast, either continue to cover negative stories about the opposing candidate at a higher rate or were already on the story before a related event caused everyone else to temporarily pick up the story.

Finally, Chapter 5 discusses the implications of my findings for democratic accountability and the health of American democracy. I conclude that for the most part democracy is conducted in a dim light.

To Caitlin.

Contents

List of Figures

List of Tables

Acknowledgments

Special thanks to my dissertation committee, Gabe Lenz, Eric Schickler, Jas Sekhon, and Kevin Quinn for all their help from conception to completion. Laura Stoker, Jack Citrin, Rob Van Houweling, and Sean Gailmard also contributed greatly to my graduate education and provided valuable feedback on my dissertation in one form or another. At Berkeley, I benefited from the friendship and advice of numerous fellow graduate students, particularly Doug Ahler, David Broockman, Devin Caughey, and Ryan Hübert.

The title of my dissertation was inspired by John Zaller's suggestion of the title "Democracy in the Dark." While that would make a much snappier title, I am happy to report that there is still a dim light on. I am grateful to Jack Citrin for keeping the lights on for this project through the financial support of IGS. And while I mostly enjoyed teaching at Berkeley, I am deeply appreciative of Gabe Lenz and Eric Schickler for making sure that my summers at Berkeley were free to focus on research.

The machine learning used in this project would not be possible without human beings to learn from. A long list of Berkeley undergraduates worked on this project in one way or another through the Undergraduate Research Apprentice Program (URAP) at Berkeley. Thank you Luke Ward, Chloe Hunt, Lindsey Lohman, Jeff Landa, Kira Day, Mackenzie O'Holleran, Meghan Babla, Chloee Weiner, Rachel Tanabe, Gabriella Armato, and Chelsea Sektnan. I am also grateful for the long distance help of Ege Özhan who helped me cross the finish line.

I would like to thank by name a few of the many teachers that helped me make it to Berkeley. Thank you to my undergraduate thesis adviser Richard Ellis for his advice and support. Thank you Kelley Strawn for introducing me to quantitative social science. Thank you Karen Utter for your kindness and for introducing a boy obsessed with math & sports to a novel about philosophy.

They say it takes a village to raise a child, but that is probably not giving my parents enough credit. I am grateful for the love, patience, and support of my parents, Connie and Gregor. Thank you to my entire family for your love and support, particularly my sister Laura and my grandparents, Ann, Bob, and Marianne. Last, but not least, thank you Caitlin for listening to me talk about this project from the day I picked my dissertation topic to the day I turned it in. It's finally done!

Chapter 1

Introduction

This dissertation examines the ability of the media to monitor politicians and the ability of voters to acquire politically relevant information. Institutions like the press have the potential to foster democratic accountability by sounding the alarm on incumbent representatives who fail to faithfully represent their constituents, but what information do journalists provide voters in practice? How do voters make voting decisions with little information? And when the media does shine its spotlight on a major event in a presidential campaign, do important stories become a more permanent feature of campaign coverage or do they quickly fade away? Each chapter in my dissertation was written as a separate paper that tries to address one of these questions.[1] Together, they show that the media does not play the role of third party watchdog effectively, low-information voters will cast ballots for candidates who simply look the part, and in presidential elections journalists spin their spotlight from breaking news story to breaking story without providing a sustained focus on important topics. For the most part, democracy is conducted in a dim light.

To preview my results, Chapter 2 examines how local newspapers cover congressional candidates, particularly incumbent representatives. I find that while local newspapers do often mention a policy position of a House candidate, even someone diligently reading their local newspaper will rarely read criticism of their incumbent representative. In House elections, horse race coverage displaces policy coverage and incumbents who vote against a majority of their constituents on landmark legislation receive the same milquetoast coverage as more faithful representatives. Even representatives referred to the House Ethics Committee receive only a whiff of a scandal and no significant increase in criticism. In total, incumbents of all stripes receive an overwhelmingly neutral tone of coverage that rarely provides substantive criticism.

When voters have little information about political candidates, they often fall back on low-information heuristics that are uninformative. Chapter 3 examines whether candidate appearance influences vote choice.[2] Prior research had shown that candidates' looks predict

[1]Because each chapter comes from a separate paper, the chapter begins with a brief explanation of how the chapter fits into the broader dissertation, followed by the paper's abstract.

[2]This chapter is drawn from a paper previously published in *Political Behavior* titled "Face Value?

election outcomes, but this observational finding is vulnerable to an alternative explanation: candidates who work hard to get elected also work hard to get a better photograph. This chapter, however, uses two randomized experiments to show that candidates really do get votes just for looking the part. In particular, actual voters who vote on mock ballots just prior to an election are more likely to vote for the appearance advantaged candidates in a treatment condition that includes candidate photographs on the ballot than in control condition that does not. Absent other information, low-information voters will use candidate appearance to make vote choices.

If Chapter 3 looks at what can happen when voters fail to acquire even the most basic political information, Chapter 4 looks at online media coverage in the most saturated media environment: the 2016 U.S. presidential election. In particular, the chapter examines when media outlets ignore an event, when they pick up a story temporarily, and when they make a topic a more permanent feature of their campaign coverage. I find that newspapers will provide information about and even uncover major scandals about presidential candidates, but even these stories quickly fade away. In contrast, more partisan websites were already covering negative topics about the opposing candidate before these events or continue to cover many of these negative topics about the opposing candidate for the remainder of the campaign. Finally, Chapter 5 offers a brief conclusion on the implications of my findings for democratic accountability and the health of American democracy.

Experimental Evidence that Candidate Appearance Influences Electoral Choice." The paper was coauthored with Gabriel Lenz, Douglas J. Ahler, and Jack Citrin.

Chapter 2

Out of Step, but in the News? The Milquetoast Coverage of Incumbent Representatives

2.1 Paper Abstract

Why do citizens routinely fail to vote out of step representatives out of office and what institutions can help voters hold politicians accountable? To the extent that politicians exploit voters' lack of information to win at the ballot box despite shirking in Congress, the press could foster democratic accountability by sounding the alarm on out of step representatives and alerting otherwise inattentive voters that it is time for change. In this paper I collect an original dataset of local newspaper coverage of candidates in the 2010 House election in order to find out whether newspapers play this role for voters and act as a watchdog of incumbent representatives. After working with research assistants to provide human classification of a random subset of these articles, I use a text as data machine learning approach to measure the content of the much larger volume of articles that we cannot read. After validating an ensemble "SuperLearner" by demonstrating out-of-sample classification accuracy that for many features approaches human inter-coder agreement, I show that challengers receive less coverage than incumbents in competitive districts, horse race coverage displaces policy coverage, and newspapers do not sound the alarm on out of step incumbents. Newspapers do provide a whiff of scandal when representatives are referred to the House Ethics committee for potential ethics violations, but they do not criticize representatives accused of some form of corruption at significantly higher rates. Even in congressional districts that closely correspond to newspaper markets, journalists act as neither watchdog nor lapdog, but instead provide overwhelmingly neutral coverage, failing to criticize incumbents who vote against a majority of their constituents on landmark legislation.

2.2 Introduction

Voters face a classic principal-agent problem in controlling their elected representatives. Absent strong incentives outside politics to learn about the political process, voters are often uniformed about the most basic political facts, let alone the policy positions and performance of each of their many elected representatives (Aidt 2000). Citizens thus vote to delegate power knowing that they cannot constantly track their elected representatives once in office. To the extent that politicians exploit voters' lack of information to win at the ballot box despite shirking in Congress, any institution like the press that can act as a third-party monitor on behalf of voters can play a crucial role in the democratic process by sounding the alarm on out of step representatives and alerting otherwise inattentive voters that it is time for change (Lupia and McCubbins 1998). Journalists thus have the potential to foster democratic accountability by highlighting the behavior of out of step representatives for their constituents. This paper uses text-based machine learning to determine what information newspapers regularly provide voters and assess whether journalists play this watchdog role when incumbent representatives vote against a majority of their constituents on important legislation.

Newspapers could help voters hold politicians accountable in a variety of different ways. Zaller (2003, 119) provides a useful comparison between a Full News standard and a Burglar Alarm standard. The Full News standard "does not mean all news, which is obviously impossible. The name is intended to capture an aspiration of sober, detailed, and comprehensive coverage of public affairs." On the other hand, if journalists follow the Burglar Alarm standard they "should routinely seek to cover non-emergency but important issues by means of coverage that is intensely focused, dramatic, and entertaining and that affords the parties and responsible interest groups, especially political parties, ample opportunity for expression of opposing views."

My research design more directly tests whether newspapers meet the Burglar Alarm standard. Foremost, the paper tests whether newspapers have a "critical bite" when incumbents vote against their constituents "on controversial elements of the president's agenda," including votes on landmark legislation (Zaller 2003, 125). I also test whether newspapers have this critical bite when incumbents are referred to the House Ethics committee over corruption allegations. Finally, the paper also considers the possibility that newspapers do not provide useful coverage of House candidates in general, but do provide more politically relevant information as Election Day approaches. In particular, I focus on measuring the types of information newspapers provide, such as horse race, policy, or pork coverage, whether newspaper articles are more likely to include politically relevant information as the Election approaches, and whether horse race coverage displaces policy coverage.

While my research design most directly tests whether newspapers have the "critical bite" of the Burglar Alarm standard, in principle newspapers could adhere to either standard and I would find that they fulfill their role as a third-party watchdog, supplying voters with useful information about the candidates and sounding the alarm on out of step incumbents. Because I define criticism as criticism from any source, a newspaper could sound the alarm

on an out of step incumbent and still adhere to the objective tone of the Full News standard. To preview my results, I find that newspapers do adhere to that objective tone, but that the typical story fails to provide in-depth coverage of House candidates. What's worse, that seemingly neutral tone — no matter the candidate — typically avoids criticism of any kind in a Faustian bargain to appear moderate and unbiased at the expense of informing voters about incumbent politicians. The result: journalists fail to sound the alarm when incumbents vote against a majority of their constituents.

2.3 Data and Methods

While the press is sometimes referred to as the 4th Estate or the 4th branch of government, we know very little about the actual content of candidate coverage because each member of Congress can generate over 100 articles per year in their local newspaper, making reading and coding all candidate articles cost prohibitive. Indeed, Clarke and Evans (1983) provide one of the only studies to analyze newspaper coverage of congressional candidates. They find a significant incumbent bias, with already elected representatives receiving both more policy coverage and more personal coverage about their qualifications for office than a challenger: "The concept of bias implies an ideological compatibility between officeseeker and media management. We suspect this pales in comparison to the advantages of incumbency that our study has already illuminated" (Clarke and Evans 1983, 83). But because reading all candidate coverage is cost prohibitive they are limited to coverage in the 6 weeks prior to the election in a random sample of 71 congressional districts. They also could only look at a few characteristics of candidate coverage. And as Arnold (2004) notes, "A subsequent audit revealed that the clipping service missed about two-thirds of the articles (Goldenberg and Traugott 1984, 133)." While Arnold (2004) was able to look at the full two years prior to the 1994 election, he only searched for articles about the incumbent and could only do content analysis for 25 candidates. To overcome these resource constraints, I train a text-based machine learning algorithm on a random subset of coverage read and classified by research assistants in order to measure the content of all articles that mention a major party House candidate in their local newspaper in the two years prior to the 2010 election.

Using zip code level newspaper circulation data for 2000 to 2010 from the Audit Bureau of Circulations (ABC) and census data that identifies zip codes' congressional districts, I identify the primary newspaper by circulation in each House district. Newsbank's database includes over 90% of these newspaper for the 2009-2010 period, (see Supplementary Information (SI) for more details). I scraped all articles mentioning a major party candidate for the U.S. House of Representatives in their district's primary newspaper for 2009 and 2010. An article was included in my dataset if it was in the district's primary newspaper and included both the candidate's first and last names as listed in Congressional Quarterly. In total, there are 111,333 candidate-articles in my dataset for the 2009-2010 period. Because articles were retrieved by searching for each candidate's name, articles that mention both major party candidates are included twice in my dataset, once under the Democratic candidate and once

under the Republican candidate. All candidate coverage in the local newspaper, including both news and editorials, are included in this dataset.

From this set of candidate-articles a random subset were read and classified by four undergraduate research assistants. To give each candidate an equal probability of being included in the learning set, I stratify by district and party before randomly drawing the subset for human classification.[1] Articles selected to be read are randomly assigned to two research assistants with each two person pair assigned the same number of articles. After using the 1,480 candidate-article classifications by these research assistants to determine whether articles included in-depth coverage, I drew a second stratified random sample of in-depth coverage that was read and dual classified by a second set of five undergraduate research assistants. In this paper I combine the two training sets. In total, research assistants provided a training set of $N = 3,728$ (dual classified articles are counted twice) that at least mentioned a candidate.

Prepossessing Text

Quantitative text analysis tends to treat a text as a "bag of words" in which word order, punctuation, and capitalization do not matter (Grimmer and Stewart 2013). I employ a similar approach, constructing an n document by k term matrix with counts for all 1,2, and 3 word n-grams for each article regardless of their position in the text (after stripping out punctuation, removing stop words, and stemming the text). While this technique has been successful in the text analysis literature, if an article discusses two candidates, one positively and one negatively, a model trained on this document term matrix would yield the same estimate for the tone of coverage towards both candidates. In order to overcome this problem, I implement two additional text preprocessing steps.

First, I tokenize important concepts, including candidate name, opponent name, and state, so that these important concepts will be represented the same across districts and articles that mention both candidates will be distinct when viewed from the perspective of Candidate A vs. Candidate B. Second, I create a second document term matrix that includes counts of 1,2, and 3 word n-grams that appear within the 50 characters of the candidate's last name, in either direction, anywhere their name appears. This allows for the possibility that text closer to the candidate's name is more predictive of an article's coverage of that

[1] If we were to draw a random sample of articles without stratifying, candidates with large volumes of coverage would be over-represented in the learning set relative to candidates who received only a small amount of coverage. While an unstratified design would obtain a representative sample of the typical article, the stratified design obtains a representative sample of the coverage of the typical candidate. Ultimately, we are interested in knowing what the coverage of a typical candidate looks like, not the contents of the typical article. This subtle distinction is particularly important if candidates with a large volume of coverage receive atypical coverage such that the text in these articles might be a poor predictor of the content of an article in a more typical district. For example, Nancy Pelosi receives a large volume of coverage in her district, but the features of the text that predict a positive article for Nancy Pelosi may be very different from the features that predict a positive article for the typical candidate. Indeed, Republican candidates regularly associate their opponents with Nancy Pelosi.

candidate, which may be particularly true in a long article that discusses several candidates. For full details, see the Preprocessing Text section of the SI.

Before preceding, one additional prepossessing step was necessary. As mentioned, articles were downloaded if they were in a candidate's local newspaper and included both their first and last name. Because some candidates go by a middle name or a nickname that in a written publication is squeezed between the candidate's first and last names, I did not require that the first and last names be next to each other in the text. However, this led to false positives where articles not about a candidate were included in the dataset. While there are no other "Ron Paul"s in Ron Paul's district, their are plenty of "Paul"s and plenty of "Ron"s that sometimes end up in the news together. Because a significant percentage of the articles in the initial dataset were not actually about the candidate (around 30 % in the first random sample), all subsequent analyses first classify whether or not the candidate in question is actually mentioned in the article and then analyze the remaining articles classified as including as least a mention of the candidate.

The SuperLearner Algorithm

Several different possible models could be used to classify candidate coverage. I use an ensemble "SuperLearner" algorithm (van der Laan, Polley, and Hubbard 2007). This algorithm takes a set of machine learning algorithms, applies them to the training set, measures their out-of-sample performance within the training set using V-fold cross-validation, and then creates a weighted ensemble "SuperLearner" that is a combination of all the machine learning algorithms tested, weighted by their out-of-sample performance. More specifically, within the cross validation stage each candidate learner is trained on the set of observations not in the V-fold and makes out-of-sample predictions for the V-fold of observations left out such that every candidate learner makes an out-of-sample prediction for each observation in the training set. The candidate learners are then trained on the entire training set and the SuperLearner algorithm is a weighted ensemble of these algorithms, weighted by regressing the actual values of the dependent variable on the out-of-sample predictions of each algorithm in the cross-validation stage.[2] This method can be applied to text as data in order to classify documents and has the virtue that it uses the method of prediction that in practice performs best out-of-sample (see van der Laan, Polley, and Hubbard 2007 for a full discussion of the algorithm and its properties, including a proof of the rate of asymptotic convergence of the estimator to "the best possible estimator given the set of candidate learners considered").

I implement the SuperLearner algorithm in Python using the scikit-learn package as a library of machine learning algorithms. For this paper, the underlying candidate algorithms used by the SuperLearner include OLS, logit, lasso, multinomial naive Bayes, Guassian naive Bayes, support vector machine, decision tree, Random Forest, ridge regression, gradient

[2]To avoid problems with overfitting in highly colinear data, I use a Lasso with $\alpha = 0.001$ penalty on large coefficients. A Lasso with $\alpha = 0$ is equivalent to Ordinary Least Square (OLS) regression so this tiny penalty on large coefficients only shrinks model coefficients significantly in cases of extreme multi-collinearity.

boosting, and AdaBoost. For each feature classified by research assistants, these algorithms make up an ensemble weighted by their out-of-sample performance on that feature.

Out of Sample Validation

I randomly divide the candidate articles classified by research assistants (dual classified articles are counted twice) in the learning set into a training set (80% of the sample) and a test set (20% of the sample) against which to measure the algorithm's out-of-sample performance. When two research assistants read and classified the same candidate-article, the vast majority of candidate-articles in the learning set, I group both classifications together in either the training or the test set by randomizing at the level of the candidate-article.

The ability of the algorithm to accurately classify out-of-sample candidate articles is a first test of the validity of the SuperLearner model. One complication, however, is that most articles have been classified twice by two different readers. If what we cared about were the individual person's assessment, for example if they were rating a product, then we might include demographic variables about that individual in the model and offer a different classification of each article for each individual. However, given that the candidate-article is the unit of analysis and what we care about is describing the content of candidate coverage, the model should only offer one predicted value for each candidate-article. Because two human beings can and frequently do disagree on the classification of an article (and we incorporate such disagreement because it helps avoid overfitting), a classifier by construction must provide the wrong classification according to one of the two human research assistants in these instances. In Table 2.1, I report the model's out-of-sample mean squared error, classification accuracy, and inter-coder agreement of human beings. As Table 2.1 shows, the algorithm performs best on binary choice variables and struggles most to predict 7 point ideology scales. For these dichotomous variables, the SuperLearner's average out-of-sample classification accuracy for the learning set lands on average within 3% of human performance on dual classified articles. Figure 2.1 plots the performance of the SuperLearner against human performance for key variables.

Table 2.1: SuperLearner Out-of-Sample Performance

	SuperLearner Accuracy	Inter-Coder Agreement	MSE	Categories	Dual Classified N	Classified N
Endorsement	0.97	0.97	0.03	2	1656	3728
Scandal	0.96	0.96	0.03	2	1656	3728
Out of Step	0.96	0.95	0.04	2	1091	2257
Pork	0.91	0.90	0.07	2	1656	3728
Healthcare	0.88	0.91	0.09	2	537	1419
Editorial	0.87	0.88	0.10	2	1091	2257
Criticism	0.82	0.88	0.13	2	1656	3728
Has Party	0.80	0.89	0.14	2	1656	3728
Horse	0.79	0.80	0.15	2	1656	3728
Policy	0.78	0.85	0.15	2	1656	3728
Tone	0.64	0.68	0.56	5	1656	3728
Criticism Type	0.63	0.69	0.24	2	219	684
Primay Focus	0.51	0.55	0.57	2	1656	3728
Ideology	0.37	0.41	1.61	7	1656	3728
Policy Ideology	0.17	0.43	2.82	7	537	1419

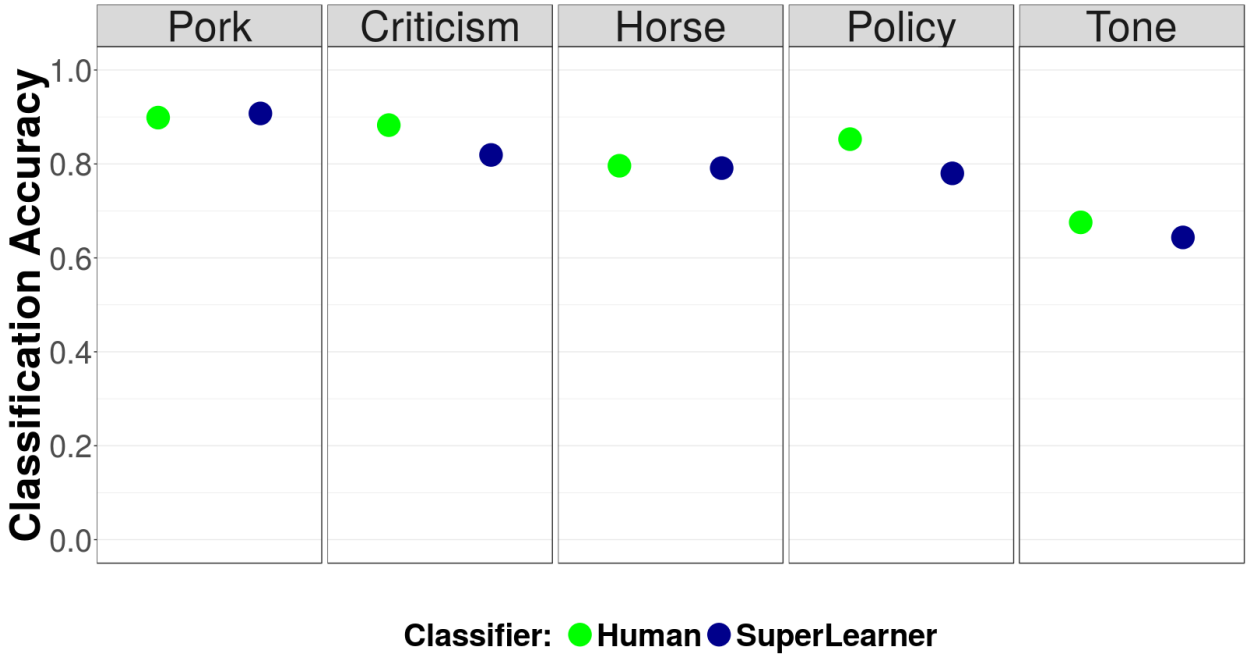


Figure 2.1: Super-Learner vs. Human Out-of-Sample Performance

Incorporating Uncertainty in Machine Learning

While machine learning algorithms can make very accurate predictions, it is important to avoid overfitting on the training set and incorporate the uncertainty in these estimates into subsequent analyses. To obtain accurate estimates of candidate coverage that better incorporate the uncertainty of the machine learning estimates into my subsequent analyses, I do two things. First, as discussed in the previous section, most articles in the learning set are classified by two different research assistants. Two human beings can and frequently do disagree on the classification of an article. By including both (sometimes conflicting) classifications in the training set, I incorporate such disagreement into the weights placed by the machine learning algorithms on different terms in the text. So, for example, if two human beings agree that an article criticizes a candidate then the algorithm will make a more definitive prediction based on the word terms in that article than the word terms in an article where two human beings disagreed about whether or not it included criticism.

While using dual classification helps avoid overfitting on word terms, the SuperLearner estimates may still be influenced by the inclusion or exclusion of particular articles in the training set. While resampling different batches of articles for human research assistants to read would be cost prohibitive, I can repeatedly sample from the articles that were read in order to provide some bounds on the extent to which including or excluding a certain set of the articles influences the estimated model. For each outcome variable, I obtain multiple estimates by training the SuperLearner on a random sample of 80% of the classified candidate

articles $m = 100$ times. For each of the 100 estimates, I draw an 80% sample without replacement from the full learning set, use that sample to train the SuperLearner, produce estimates of coverage for all articles, and then analyze the data using those estimates. After repeating this process 100 times, I follow King et al.'s (2001) outline for how to incorporate analyses involving multiple datasets into a single point estimate with standard errors that incorporate the variance across the different samples (Rubin 1987 as cited by King et al. 2001). For any quantity of interest Q the point estimate $\overline{q}$ is simply the mean across the m samples:

$$\overline{q} = \frac{1}{m} \sum_{j=1}^{m} q_j \tag{2.1}$$

So, for example, all regression coefficients reported are the mean of the coefficient as estimated in each of the $m = 100$ samples. If we were to only look at one sample and the estimated standard error $SE(q_j)$, we would not incorporate the sample variance S_q^2 across the m point estimates:

$$S_q^2 = \sum_{j=1}^{m} (q_j - \overline{q})^2 / (m-1) \tag{2.2}$$

To calculate the standard errors then for the multiple dataset estimate, we take the square root of the average variance within datasets plus the variance across datasets "(multiplied by a factor that corrects for bias because $m < \infty$)":

$$SE(q) = \sqrt{\frac{1}{m} \sum_{j=1}^{m} SE(q_j)^2 + S_q^2(1 + 1/m)} \tag{2.3}$$

Because the learning set — articles classified by human beings — is only drawn once, the estimates could still ultimately be biased if the randomly sampled learning set is very atypical of the larger population of articles, but if the learning set is representative of the population then this process more accurately reflects estimate uncertainty based on random choice of learning set. Note that this procedure adds the variance across datasets S_q^2 to the average variance within the multiple datasets, so the procedure is conservative relative to performing the analysis on a single sample.

2.4 Results

To preview my results, first I show that the machine learning results have convergent validity and newspapers supply basic political information. For example, competitive districts receive a greater share of horse coverage and newspapers typically provide readers with the candidate's political party. Next, I show that that while challengers receive a greater share of coverage as the election approaches, they still receive less coverage than incumbents in competitive districts. Then I show that horse race coverage displaces policy coverage. Finally, I test whether newspapers have the critical bite of the Burglar Alarm standard. First and foremost, I examine whether newspapers sound the alarm on incumbents who vote against a majority of their constituents and show that out of step incumbents do not receive a greater share of criticism, but instead get the same milquetoast coverage as more faithful representatives. Lastly, I show that while there is a whiff of a scandal when members are referred to the House Ethics committee, they do not receive significantly more criticism.

Throughout this section, I analyze both the entire set of candidate articles for the 2009-2010 period and the 90 days prior to the 2010 general election, examining all articles and in-depth articles alone. I define in-depth coverage as articles where the candidate is (4) a major focus or (5) the primary focus of the article, as classified by the SuperLearner algorithm on a five point scale (see SI for full question wording). All regression analyses use probability weights such that each congressional district receives equal weight in the model, with standard errors clustered at the district-level. Reported results reflect the average of the analyses using the $m = 100$ training sets. Unless otherwise noted, analyses use articles' predicted values from the SuperLearner rather than classifications in order to avoid bias from classifying a disproportionate portion of articles into a particular category (for example, at .55 on a 0-1 scale an article would be classified as including a policy stance from the candidate, but this is inconsistent with the model's predicted probability of a 45% chance that the article does not contain policy coverage). Results are based on a model trained on both the stratified random sample and the in-depth stratified random sample. Time is measured in months, with larger values closer to Election Day. District competitiveness is measured as the .5 minus the absolute distance of the district's 2008 presidential vote share from .5 (i.e. 50/50), such that larger values indicate a more competitive district, i.e. $.5 - |2008 \text{ Democratic Presidential Vote Share} - .5|$. The analyses in this paper broadly follow a preanalysis plan. The results for the exact analyses proposed and a discussion of the changes made in the paper are included in the SI.

Convergent Validity: Newspapers Provide More Horse Race Coverage in Competitive Races and Closer to Election Day

Before moving to other hypotheses that may very well be false, I examine a feature of candidate newspaper coverage where we have extremely strong theoretical expectations that verge on common sense. I expect that newspapers will provide more horse race coverage in competitive races and closer to Election Day. If we did not see that newspapers include

horse race coverage in a greater share of articles under these circumstances, then we should be extremely skeptical about other machine learning results. As expected, however, I find that newspapers provide horse race coverage in a greater share of articles both in competitive districts and closer to Election Day. For in-depth coverage, I estimate that newspapers discuss "any aspect of an election or an electoral campaign" in 26 percent more articles on Election Day than three months prior to the election (Table 2.2, Column 4). While unsurprising, this result helps validate the machine learning based results. Similarly, I estimate that candidates in 50/50 districts receive 4% more horse race coverage than candidates in 60/40 districts (Table 2.2, Column 4).

Table 2.2: Horse Coverage By Days to Election and District Competitiveness

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Date (in months)	0.015***	0.086***	0.018***	0.085***
	(0.001)	(0.008)	(0.001)	(0.012)
District Competitiveness	0.29***	0.489***	0.229***	0.4***
	(0.055)	(0.071)	(0.088)	(0.117)
Observations	62185	13499	19592	5280
Districts	333	319	314	272

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Political Party: An Available Heuristic

If voters want to know how a candidate will vote in Congress, the candidate's political party serves as an excellent heuristic. While more in-depth policy coverage could be valuable to voters, candidates may strategically take public positions that (mis)represent their general ideology (Henderson 2013). To the extent that candidates can influence which issue positions get covered, providing a candidate's political party may be particularly valuable to voters trying to determine whether a candidate represents their views in general. I find that newspaper coverage explicitly and consistently identifies candidates' respective political parties in 86 percent of articles [95 % CI (0.84,0.88)]. Thus, while candidates may be able to leave their party out of television commercials or highlight issues designed to distance themselves from their party's brand, newspapers by style convention provide readers with candidates' political party.

While newspaper style guides do not require the mention of a candidate's party, the Associated Press' Style Guide for Party Affiliation (2016) provides a prototypical example of when party identification will 'naturally occur": an article covering "two senators that are vying for a single senate seat." Thus, we would expect that journalists will be most

likely to mention a candidate's political party in a horse race story. Indeed, I find that horse race coverage virtually guarantees that an article about a candidate mentions their political party, increasing the percent of articles including party identification by 31-37% depending on the specification (Table 2.3). Policy coverage also increases party identification by 8-19% (Columns 1-4), and pork coverage increases party identification by roughly 16% for in-depth coverage (Column 3 and Column 4). Finally, I had expected that articles would be more likely to mention a candidate's party as the election approached, but I find no clear pattern.[3] Overall though, the machine learning results correspond with what we would expect based on journalistic style convention: newspapers will almost always identify the candidate's political party, particularly in campaign coverage.

Table 2.3: Political Party Identified?

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Date (in months)	0.000	-0.02***	-0.001	-0.009
	(0.001)	(0.005)	(0.001)	(0.008)
District Competitiveness	0.025	-0.091	0.018	-0.107
	(0.046)	(0.056)	(0.05)	(0.072)
Pork	0.103**	0.082	0.156***	0.171*
	(0.04)	(0.065)	(0.048)	(0.095)
Policy	0.186***	0.122***	0.138***	0.082
	(0.031)	(0.04)	(0.037)	(0.05)
Horse	0.368***	0.359***	0.312***	0.32***
	(0.044)	(0.053)	(0.057)	(0.082)
Primary Focus	-0.002	-0.001	0.035	0.039
	(0.011)	(0.014)	(0.03)	(0.046)
Challenger	-0.021	-0.015	-0.04**	-0.019
	(0.013)	(0.013)	(0.016)	(0.017)
Open Seat	-0.01	-0.008	-0.029	-0.009
	(0.017)	(0.02)	(0.021)	(0.024)
Observations	66730	15421	21174	6172
Districts	367	353	346	302

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

[3]In Table 2.3 I find no effect, or even a very small negative one. One potential explanation for this finding could be that the other aspects of coverage in the model, such as Horse Coverage, increase over time and thus we might worry that party identification also increases over time in conjunction with these other features. To rule this out I also run the analysis in Table 2.3 without the other features of the article content and find that the effect is always positive, but small (see SI Table 2.14).

Challengers Receive Less Coverage

As we would expect, newspaper coverage of both incumbents and challengers picks up significantly as the election approaches, particularly in the 30 days prior to the election (see Figure 2.2).[4] For voters to make informed decisions between an incumbent and his challenger, they need information about both candidates. Coverage of the challenger makes up only 20% of articles in the full sample. However, challengers received a much larger share of candidate coverage as the election approached: articles written in the 90 days prior to the election were roughly 19% more likely to be about the challenger if they were written on Election Day in November than if they were written at the beginning of August (see Table 2.4, Column 2). The challengers' share of in-depth coverage increases by 24% over this same period (Table, 2.4, Column 4). Thus, both the total volume and the challenger's relative share of coverage increases as the election approaches. But as the local estimate of the challenger's share of coverage by days to election shows, the challenger approaches, but never achieves parity in the volume of coverage (Figure 2.3).[5]

Table 2.4: Challenger Share of Coverage By Days to Election and District Competitiveness

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Date (in months)	0.018***	0.063***	0.021***	0.081***
	(0.001)	(0.011)	(0.002)	(0.017)
District Competitiveness	0.501***	0.744***	0.46***	0.765***
	(0.1)	(0.134)	(0.116)	(0.178)
Observations	62185	13499	19592	5280
Districts	333	319	314	272

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Like anything that might influence an election, if the challenger's share of coverage influences voters it could only sway the election in close races. In addition to receiving more coverage as the election approaches, challengers also receive significantly more coverage in competitive districts. Going from a 60/40 district for 2008 Democratic presidential vote, to a 50/50 district (increasing the measure of district competitiveness by .1) increases the challenger's share of general coverage by 5% (Table 2.4, Column 2) and in-depth coverage by roughly 8% (Table 2.4, Column 4) in the 90 days before the election. As the local estimate of the impact of district competitiveness on the challenger's share of coverage shows, however,

[4]When analyzing incumbents and challengers, I look only at contested races in which the incumbent runs for reelection.

[5]Blue line represents the local estimate, with a shaded 95% confidence interval. Points plot bin averages, with larger points for bins with more observations.

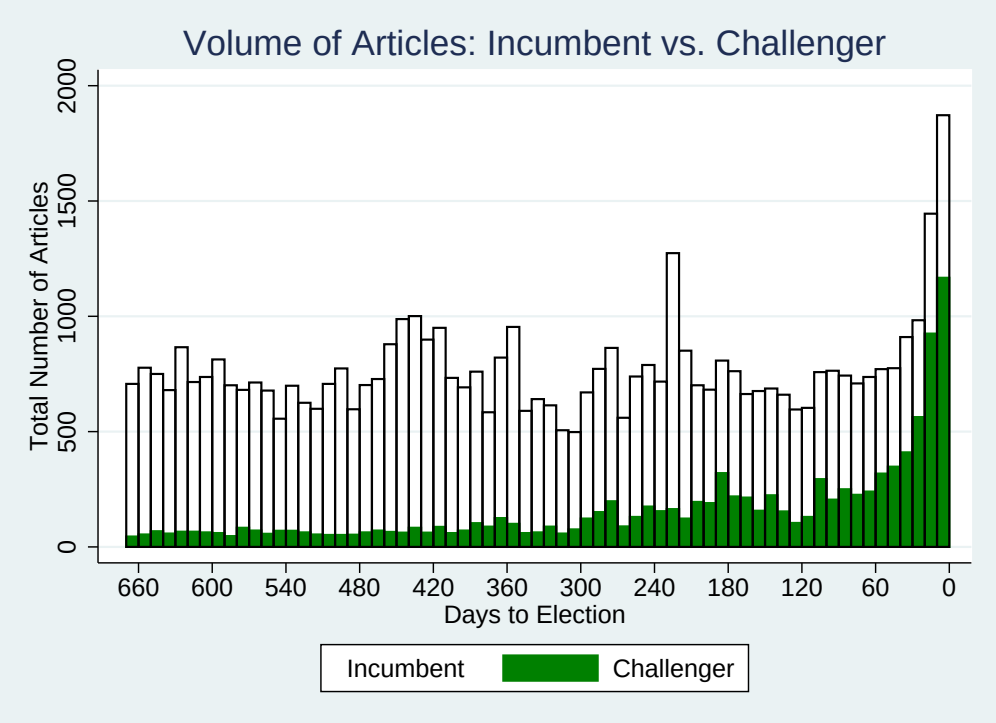


Figure 2.2: Histogram of Number of Articles by Days to Election Day. Counts of articles by days to Election Day 2010 for Incumbents and Challengers.

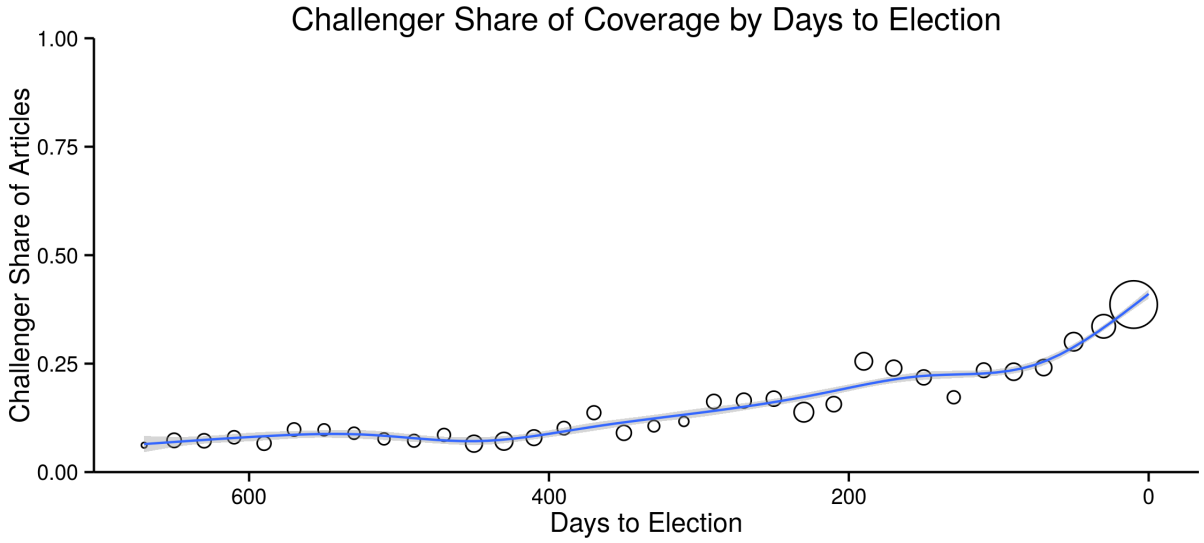


Figure 2.3: Local Estimate of Challenger Share of Coverage By Days to Election Day

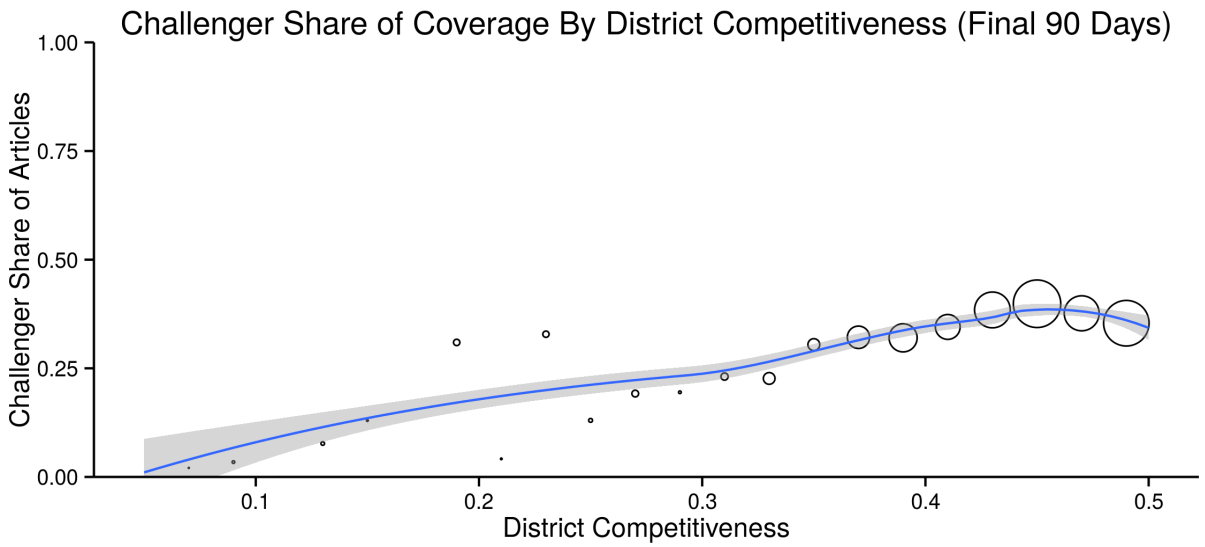


Figure 2.4: Local Estimate of Challenger Share of Coverage By District Competitiveness in Final 90 Days.

the regression estimates in Table 2.4 may overestimate the spike in challenger coverage in competitive districts. When going from a relatively safe 60/40 district to a 50/50 district (see Figure 2.4) coverage barely increases.

More Policy Coverage Closer to the Election, but Horse Race Coverage Displaces Policy Coverage

While the total volume of articles, the challenger's share of articles, and horse race coverage increase as the election approaches, the share of articles that include a policy stance shows no clear trend. In the full sample and in the final 90 days policy coverage decreases as the election approaches. The largest shift occurs for articles in the final 90 days where the share of policy articles is estimated to decrease by 7% between early August and Election Day (see Table 2.5, Column 2). This pattern, however, does not hold for in-depth articles in the final 90 days (Table 2.5, Column 4). Across all specifications in Table 2.5, newspapers cover candidate's policy stances slightly more often in competitive races, but the effect is substantively small. In the final 90 days, a 50/50 district will see policy positions in about 4% more of in-depth articles than in a 60/40 district (Table 2.5, Column 4).

While district competitiveness may increase policy coverage, providing readers with certain policy stances of a candidate, horse race coverage displaces policy coverage (see Table 2.6). Note that both horse race coverage and policy coverage were very broadly defined and non-exclusive categories: the codebook for the project instructs research assistants that "An article includes a policy stance if it describes a belief, vote, statement, or any other action of

Table 2.5: Policy Coverage By Days to Election and District Competitiveness

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Date (in months)	-0.006***	-0.023***	0.000	0.011
	(0.001)	(0.008)	(0.001)	(0.011)
District Competitiveness	0.192***	0.295***	0.246***	0.375***
	(0.064)	(0.065)	(0.083)	(0.111)
Observations	62185	13499	19592	5280
Districts	333	319	314	272

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

a candidate that explicitly or implicitly identifies a candidate's position on a policy" and "An article includes "horse race" coverage if any aspect of an election or an electoral campaign is discussed." Thus, there was no inherent reason that policy coverage and horse race coverage had to displace each other within an article, but horse race coverage clearly does take the place of policy coverage. In the full sample, horse race coverage decreases policy coverage by 39% (Table 2.6, Column 1). Even for in-depth articles in the final 90 days, where the effect is smallest, an article that includes horse coverage is 12% less likely to include a policy stance of the candidate (Table 2.6, Column 4). We see a similar pattern in Figure 2.5 for local non-parametric estimates.

Table 2.6: Policy Coverage By Days to Election, District Competitiveness, and Horse Coverage

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Date (in months)	0.000	-0.005	0.005***	0.021*
	(0.001)	(0.008)	(0.001)	(0.012)
District Competitiveness	0.304***	0.4***	0.312***	0.421***
	(0.06)	(0.065)	(0.079)	(0.114)
Horse Coverage	-0.386***	-0.215***	-0.288***	-0.118
	(0.037)	(0.044)	(0.058)	(0.078)
Observations	62185	13499	19592	5280
Districts	333	319	314	272

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

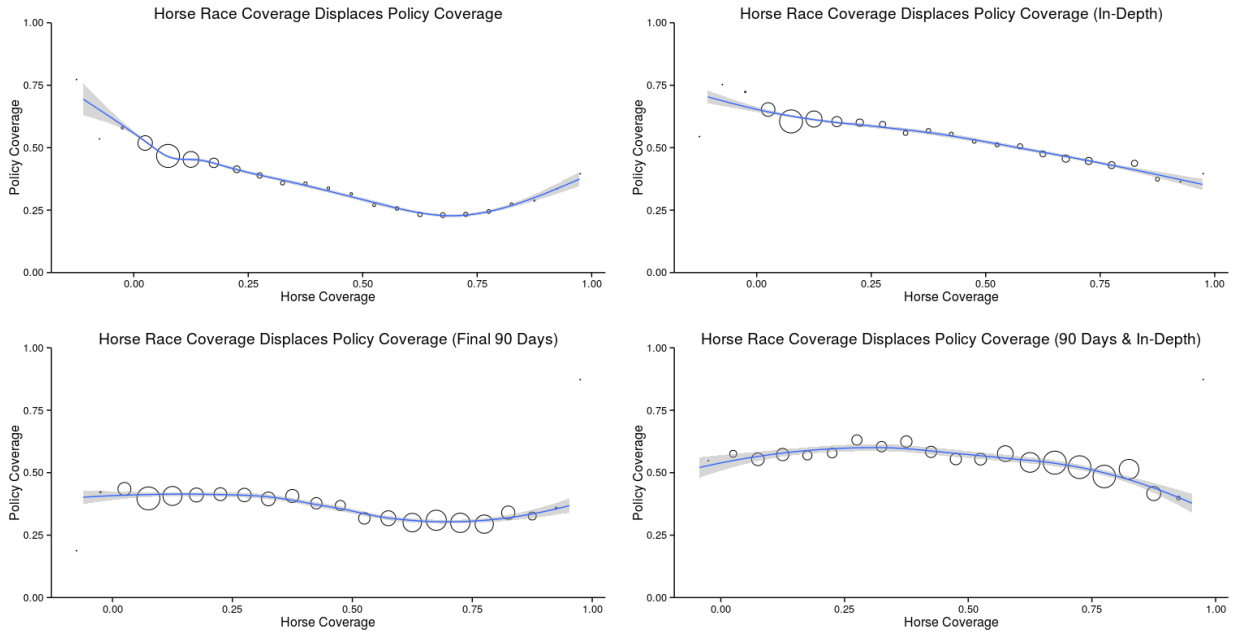


Figure 2.5: Local Estimate of Policy Coverage (Predicted Probability) by Horse Race Coverage (Predicted Probability).

Do Newspapers Act As Watchdogs and Sound the Alarm? Out of Step But in Office

Over 90% of incumbents continue to win reelection despite incumbents often voting against their constituents on key votes. If voters lack information about incumbents or have biased information provided by campaigns, they cannot be expected to vote out of step incumbents out of office. Do newspapers sound the alarm and provide negative coverage of out of step incumbents?

In order to assess the quality of policy representation we must place members of Congress and their districts in the same policy space. I present results using MRP-like estimates of district preference on particular bills, which that are derived from the 2010 Cooperative Congressional Election Study (CCES) using a hierarchical model weighted to validated turnout (see Hill 2015 for full details). These bills include the 2009 Stimulus, The Affordable Care Act, Dodd-Frank, SCHIP expansion, Cap and Trade, the repeal of Don't Ask Don't Tell. Taken together, these votes represent the core of what Barack Obama did and did not accomplish legislatively during his presidency. Thus, the paper analyzes the impact of incumbents' "roll call votes on controversial elements of the president's agenda," including votes on landmark legislation (Zaller 2003, 125). Pooling together these six important votes, I examine the impact of Votes Cast With Constituents, Abstentions, and District Competitiveness on candidate criticism. Unless otherwise noted, the analysis of incumbent coverage includes only districts in which the incumbent ran for reelection and the race was contested by both

major parties.

Table 2.7: Candidate Criticism and Out of Step Voting in Congress

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Votes Cast With Constituents	-0.004	-0.006	-0.002	-0.003
	(0.003)	(0.005)	(0.006)	(0.008)
Abstentions	0.007	-0.002	0.006	0.011
	(0.007)	(0.013)	(0.011)	(0.025)
District Competitiveness	0.059	0.237***	0.128	0.383***
	(0.059)	(0.077)	(0.082)	(0.133)
Observations	47412	8184	14106	2935
Districts	307	279	283	229

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

If incumbents were rewarded for voting with their constituents, we would expect that candidates would receive less criticism the more key votes they cast in step with their constituents. Instead, I find little to no effect. In the full sample, casting an additional important vote with their constituents decreases criticism by an estimated -.004 [95% CI (-.010, 0.002)] (Table 2.7, Column 1). In other words, a candidate who voted against their constituents on all six important bills would receive criticism in only 2.4% more articles than a candidate perfectly in step with the majority of their constituents on all six bills. And indeed for in-depth coverage in the final 90 days, if anything, in-step representatives receive a smaller reward than in general coverage (Table 2.7, Column 4).

I obtain similar results examining the general tone of coverage.[6]. Finally, I find that abstaining on these important bills has no significant impact on the amount of criticism a candidate receives (Table 2.7) in this analysis.[7]

One potential concern could be that I find a null effect because of errors in the machine learning. I obtain similar results, however, when I replicate my analyses directly on the learning set using human classifications instead of machine learning (see SI, Learning Set Replication). Another concern is that measurement error in the estimates of constituent preferences could lead to attenuation bias. However, I find a similar lack of significant crit-

[6]I present results on the impact of incumbent voting in Congress using candidate criticism as the dependent variable for two reasons. First, newspapers adopting the Full News standard might balance criticism from one source with praise from another, leading to a neutral tone. Second, the machine learning algorithm more accurately measures candidate criticism than tone of coverage. However, I find broadly similar results when using tone of coverage as the dependent variable instead of candidate criticism (See SI for full details).

[7]I do find some evidence of criticism for abstentions in the final 90 days when directly analyzing the training set, possibly because candidate's opponents criticize them for missing key votes in their campaigns (see SI for more details).

icism for out of step incumbents when I perform the same analysis using the relationship between districts' presidential voting and DW-NOMINATE to measure MC's relative extremity.[8] In the main analysis for in-depth coverage in the final 90 days presented here (Table 2.7, Column 4), I can reject the hypothesis that casting an important vote against a majority of your constituents increases criticism by more than 3% at $\alpha = 0.001$.

It is possible that newspapers do not generally punish out of step incumbents, but do publish critical coverage directly after an important vote on which a representative cast a vote against a majority of their constituents. To test this hypothesis, I examine whether candidates who vote against the majority of their constituents receive significantly more criticism in a 7 day window after an out of step vote. As Table 2.8 shows, however, candidates who cast an out of step vote receive, if anything, less criticism than in step representatives in the 7 days following a vote.

Table 2.8: Out of Step Voting in Congress and Incumbent Criticism Following the Vote

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Votes Cast With Constituents	-0.004	-0.002		
	(0.003)	(0.006)		
Abstentions	0.007	0.005		
	(0.007)	(0.011)		
Vote Window	0.007	0.008		
	(0.021)	(0.032)		
Out of Step in Vote Window	-0.014	-0.051		
	(0.035)	(0.042)		
District Competitiveness	0.059	0.128		
	(0.059)	(0.081)		
Observations	47412	14106		
Districts	307	283		

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

[8] In that analysis, I only find a significant effect for in-depth coverage in the final 90 days where a one standard deviation increase in relative extremism predicts a 4.4% increase in criticism of the incumbent representative. See SI, Using DW-NOMINATE and Presidential Vote To Measure Extremity for full details. In short, as measured with DW-NOMINATE the punishment for relative extremity is if anything smaller in high congruence districts and the significant effect in the full sample represents only a 1/4th of standard deviation change in criticism.

2.5 Overlapping Markets and Democratic Accountability

In general, I find that newspapers provide an overwhelmingly neutral tone of coverage, even for out of step representatives. Overall, a mere 8.0% [95 % CI (5.8,10.3)] of incumbent coverage includes criticism of any kind and 92.2% [95 % CI (89.9,94.5)] is classified as neutral in overall tone. One explanation could be that most newspapers do not have a significant stake in any one race if their readers primarily reside in other congressional districts. Snyder and Strömberg (2010) find that members of Congress do more constituency work and are less likely to vote the party line when newspaper markets and congressional districts are highly congruent. They find a greater volume of press coverage in districts with high congruence and believe that differences in coverage drive their results. If members of Congress behave differently in highly congruent districts, a possible mechanism could be the content of newspaper coverage. With both measures of article content and data on which newspapers have the greatest share of their readership in a congressional district, I can estimate whether greater congruence between newspaper markets and congressional districts leads to more critical coverage of out of step incumbents.

For each newspaper-district pair I calculate the share of the newspaper's readers in that district. For the primary newspaper in a district, the average congruence is roughly .2 or 20%. I thus use congruence >= .2 as a cutoff and analyze coverage in the districts with high congruence by this metric. Broadly, I find that newspapers in highly congruent markets also fail to sound the alarm on out of step incumbents. In contrast, incumbents in more competitive districts do receive significantly more criticism. For in-depth coverage in the final 90 days, candidates in 50/50 2008 presidential vote share districts are estimated to receive about 6% more criticism than incumbents in districts 60/40 districts (Table 2.9, Column 4).

Table 2.9: Candidate Criticism and Out of Step Voting in Congress (High Congruence)

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Votes Cast With Constituents	-0.002	0.000	0.004	0.006
	(0.004)	(0.007)	(0.007)	(0.008)
Abstentions	0.001	-0.018**	0	-0.009
	(0.006)	(0.009)	(0.01)	(0.024)
District Competitiveness	0.195***	0.431***	0.41***	0.629***
	(0.061)	(0.13)	(0.12)	(0.186)
Observations	33447	5757	10401	2124
Districts	163	153	155	140

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

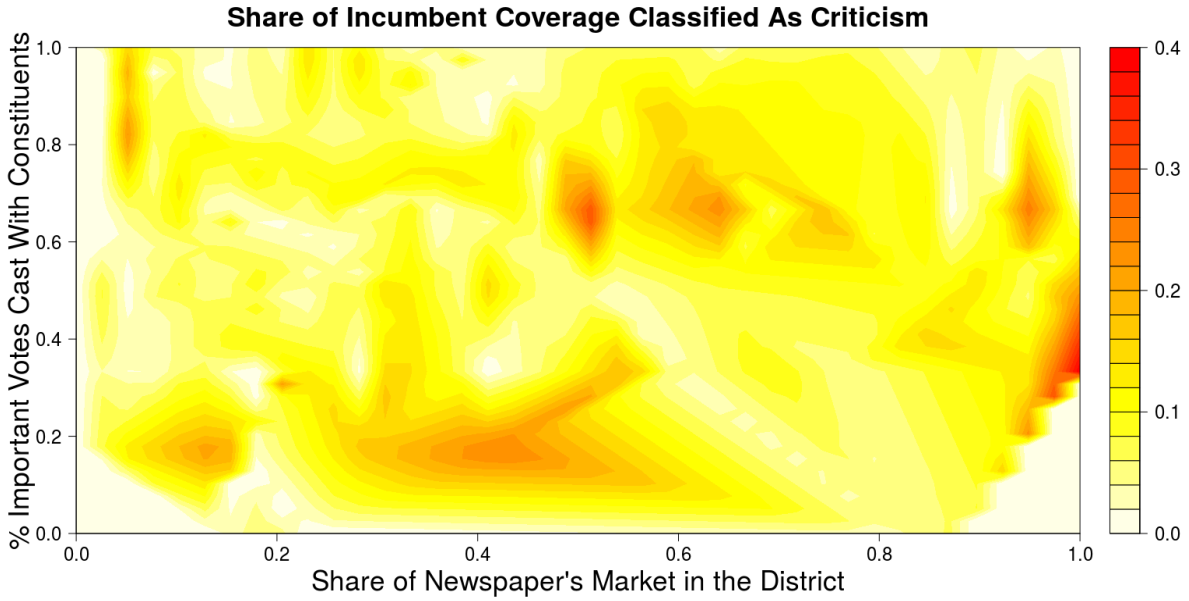


Figure 2.6: Filled Contour Plot of the Share of Incumbent Articles Classified as Criticism by Share of Newspaper's Readers in the District and % Important Votes Cast With Constituents.

As we can see in the filled contour plot in Figure 2.6, newspapers across the board fail to provide a significant share of incumbent criticism. Even newspapers with most of their readers concentrated in a single district provide little criticism when an incumbent consistently votes against his constituents. Thus, on the whole I find little evidence that newspapers criticize incumbents who fail to represent their constituents' policy preferences.

2.6 What Makes a Positive Article and Could Out of Step Representatives be Compensating with Pork?

Representatives who are out of step with their district on policy may compensate in other ways. If representatives who are out of step on policy get better coverage because they are good at other things like bringing home money for projects in their district, then these other efforts may mask the effect of voting against their constituents. When I analyze which charactertics lead to a positive tone of coverage, more than any other feature articles that cover pork projects provide positive candidate coverage.[9] Pork coverage improves the tone

[9]The actual question presented to research assistants reads "Does this article discuss a local project for the district? (A particularized good for constituents, e.g. specific spending for a bridge or health clinic in

of coverage for a candidate by .24 (see Table 2.10, Column 1) for all articles and by .31 for in-depth articles (Table 2.10, Column 3). In contrast, policy coverage and horse race coverage have no significant impact on the tone of coverage.

Table 2.10: What Makes a Positive Article?

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
District Competitiveness	0.073	0.03	0.077	-0.112
	(0.065)	(0.09)	(0.099)	(0.152)
Pork Coverage	0.235***	0.236**	0.314**	0.319
	(0.064)	(0.098)	(0.14)	(0.232)
Policy Coverage	-0.073	-0.052	-0.018	0.043
	(0.05)	(0.063)	(0.063)	(0.085)
Horse Coverage	-0.047	-0.043	0.056	0.056
	(0.061)	(0.075)	(0.078)	(0.112)
Date (in months)	0.001	0.035***	0.000	0.003
	(0.001)	(0.011)	(0.002)	(0.016)
Primary Focus	-0.031**	-0.043**	-0.068	-0.077
	(0.015)	(0.02)	(0.042)	(0.063)
Challenger	0.005	-0.008	0.022	0.02
	(0.019)	(0.022)	(0.028)	(0.036)
Open Seat	0.021	0.038	0.034	0.059
	(0.023)	(0.036)	(0.031)	(0.044)
Observations	66730	15421	21174	6172
Districts	367	353	346	302

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Because out of step incumbents may compensate voters with pork to make up for a policy disconnect and coverage of distributive goods improves the overall tone of coverage, I reanalyze whether representatives who vote with their constituents receive less criticism controlling for pork coverage. As Table 2.11 shows, however, even when the measure of pork coverage is included in the model newspapers still fail to significantly reward incumbents for voting with their constituents or punish incumbents out of step with their districts. While candidates get positive coverage from articles about specific distributive goods provided for their district, such coverage is rare (the mean predicted value for pork is .11). One explanation could be that newspapers rarely publish such pieces because they require a candidate to actually produce distributive goods for their district.

the district)"

Table 2.11: Compensating with Pork? Candidate Criticism and Out of Step Voting in Congress

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Votes Cast With Constituents	-0.003	-0.006	-0.002	-0.003
	(0.003)	(0.005)	(0.006)	(0.008)
Abstentions	0.006	-0.002	0.002	0.009
	(0.007)	(0.014)	(0.012)	(0.024)
District Competitiveness	0.058	0.232***	0.119	0.364***
	(0.059)	(0.077)	(0.08)	(0.131)
Pork Coverage	-0.129***	-0.118**	-0.248***	-0.286***
	(0.027)	(0.048)	(0.053)	(0.089)
Observations	47412	8184	14106	2935
Districts	307	279	283	229

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

2.7 When Congress Investigates: The Whiff of a Scandal

While this paper focuses on policy accountability, that is not the only form of democratic accountability. Voters and newspapers may care more about political corruption and so I examine whether newspapers cover incumbents differently when they are investigated by the House Ethics committee. Using the committee's summary of its activities in the 111th Congress, I identify MCs referred to the committee for some sort of ethics violation. I coded all referrals for allegations of corruption, excluding referrals for minor legal problems (expired driver's license, arrested at a protest). In the 111th Congress, Rep. Charles Rangel was censured for numerous ethics violations, but no other member was found to have violated House ethics rules. Both Rep. Rangel and other incumbents who were referred to the House Ethics committee received 7% more scandal coverage (Table 2.12, Column 4). Yet while Rep. Rangel receives significantly more criticism, the other members merely referred to the Ethics Committee do not. This suggests that when an article discusses a referral to the House Ethics Committee it has a hint of scandal, but but does not necessarily explicitly criticize the incumbent representative.

Table 2.12: Do Ethics Investigations Lead to More Scandal Coverage?

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Corruption Allegation	0.013**	0.024*	0.05**	0.072**
	(0.006)	(0.013)	(0.024)	(0.035)
Censure	0.119***	0.128***	0.106***	0.092**
	(0.026)	(0.033)	(0.04)	(0.045)
District Competitiveness	-0.037*	-0.007	-0.091**	-0.068
	(0.02)	(0.021)	(0.037)	(0.053)
Observations	47412	8184	14106	2935
Districts	307	279	283	229

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.13: Do Ethics Investigations Lead to More Criticism?

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Corruption Allegation	-0.001	0.003	0.036	0.039
	(0.018)	(0.029)	(0.049)	(0.046)
Censure	0.17***	0.186***	0.106**	0.13**
	(0.024)	(0.037)	(0.052)	(0.057)
District Competitiveness	0.106**	0.313***	0.171**	0.446***
	(0.048)	(0.06)	(0.069)	(0.112)
Observations	47412	8184	14106	2935
Districts	307	279	283	229

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

2.8 Conclusion

When incumbents vote with their constituents, newspapers reward them with roughly the same share of criticism and general tone of coverage as out of step representatives. In the main analysis for in-depth coverage in the final 90 days, I can reject the hypothesis that casting an important vote against a majority of your constituents increases criticism by more than 3% at $\alpha = 0.001$. Journalists do provide readers with basic information about candidates, at the most basic level they consistently provide candidates' party affiliation. They also provide some coverage of candidates' policy stances and an increasing (though not equal) share of coverage of the challenger as the election approaches. Yet, newspapers do not provide significantly more negative coverage or greater criticism of out of step incumbents. Instead, newspapers provide incumbents of all stripes with overwhelmingly neutral coverage of day to day events that rarely provides substantive criticism: 92.2% [95 % CI (89.8,94.5)] of incumbent coverage is neutral in overall tone. Even in congressional districts that closely correspond to newspaper markets, journalists do not sound the alarm on out of step incumbents.

Zaller (2003) criticizes the Full News standard for having an unreasonably high expectation of what information swing voters, often the most uninformed voters, will actually acquire and use to make voting decisions. One potential criticism of this paper is that it hues too closely to a "Progressive" civics textbook model of policy accountability in which "citizens should take each election as an occasion to examine the record of their MC to decide whether she or he deserves another term." (Zaller 2003, 124). By examining whether newspapers run more criticism when incumbents vote against a majority of their constituents, this paper does assume a model in which voters have preferences to which politicians can be held accountable. However, I do not expect that newspapers adhere to anything approaching a Full News standard and the conclusion that newspapers provide milquetoast coverage to incumbents of all stripes does not rest on any particular model of democratic accountability.

In many respects, this paper assesses the ability of newspapers to meet the Burglar Alarm standard. It does not focus its analysis on coverage of wonkish committee work or minor legislation, but on the "roll call votes on controversial elements of the president's agenda" that Zaller (2003, 125) thinks newspapers should focus on and Arnold (2004) finds that most newspapers cover. In addition to examining whether out of step incumbents receive more criticism in general, i.e. "critical bite," the paper tests whether newspapers include more criticism of an incumbent just after they vote against a majority of their constituents on controversial legislation. I find that even in that narrow window where journalists could mount a "feeding frenzy," they fail to do so. I also show that incumbents referred to the House Ethics Committee over alleged corruption do not receive significantly more criticism but only a whiff of a scandal. In short, newspapers fail to help voters meet "The needs of democracy... by scrutinizing the records of those incumbents whose achievements are in doubt and reelecting the rest with minimal fuss" (Zaller 2003, 124).

Zaller might point to the finding of increased criticism of incumbents in competitive districts as evidence that journalists do indeed meet the Burglar Alarm standard. He argues

that "if party activists and the strongest potential challengers scrutinize an MC's record and decide that, even after giving it their best shot, the incumbent could not be beaten, there is no need for an expensive, time-consuming contest" (2003, 124). Even for in-depth coverage in high congruence markets though, incumbents in 50/50 districts only see 6% more criticism than incumbents in 60/40 districts. Moreover, the low level of incumbent criticism could discourage high quality candidates from challenging incumbent representatives. As Bennet (2003, 135) argues, "Since most incumbents and many interest groups are served by the non-competitive democracy, they are not likely to sound alarms. Given the way in which journalists must advance stories either through indexing or through finding developments that push the narrative, there is no automated or routine way to 'arm' the system for this story." So this could even be a self-reinforcing pattern in which the general lack of incumbent criticism leads to fewer high quality challengers who can successfully supply journalists with these critical story lines.

Overall, a mere 7.85% [95 % CI (5.60,10.18)] of incumbent coverage includes criticism of any kind. Because that number includes criticism from any source and my dataset includes editorials, this result does not simply reflect a preference for an objective tone of coverage in the news section, but a widespread failure to offer critiques of incumbent representatives. As Zaller (2003) emphasizes, voters often lack incentives to acquire political information. In the canonical spatial model, politicians must faithfully represent their constituents' preferences (in equilibrium) because competitive elections allow informed citizens (with perfect information) to replace out of step representatives with in step challengers (Downs 1957). While this works in theory because the model assumes perfect information (and binding policy platforms), for voters to exercise anything approaching this policy accountability in practice, they need to learn when representatives fail to faithfully represent their policy preferences. In Zaller's (1992) terms, a voter must first receive a consideration before they can accept a consideration. But before a voter can receive the consideration that their representative is failing to represent their policy preferences, someone must actively provide this information. Challengers have an incentive to provide this information, but they might also simply lie about the incumbent, making them difficult to trust (Minozzi 2011). While the press could step in as a more credible third party monitor, their failure to sound the alarm on out of step incumbents makes it more difficult for otherwise inattentive voters to exercise control over their many representatives.

2.9 Supplementary Information

Data

The ABC circulation data set and Newsbank's database of coverage lack a common newspaper identifier. Newspapers were first matched based on an exact city, state, and name match. Newspapers without an exact match were then matched if they shared the same state, had similar city names (an exact match of the shorter city name could be found in the longer city name e.g. "Minneapolis" found in "Minneapolis-St. Paul"), and shared at least one seven character string in common in their name excluding their state and city names (e.g. "Star Tr" in "Star Tribune" and "Minneapolis Star Tribune").

Prepossessing Text Detailed Explanation

As described in the paper, I do two additional preprocessing steps before creating the document term matrix. First, I replace word terms that represent important concepts but which differ across articles with a term representing that concept. For example, the text 'Barbara Lee' will be included in an article on Congressperson Barbara Lee in the San Francisco Chronicle, but will not be present in an article on Congressperson Keith Ellison in the Minneapolis Star Tribune. However, the same concept — the candidate's first and last name — will appear in the form of 'Keith Ellison'. Thus, I replace the name of the candidate being analyzed in each respective candidate-article with tokens representing the concepts CANDIDATEFIRST and CANDIDATELAST. I similarly replace the first and last names of the candidate's opponent with tokens for OPPONENTFIRST and OPPONENTLAST. Thus, (a) the same concept is represented similarly across districts and (b) an article that mentions both candidates in a race with have different term counts when the article is analyzed from the perspective of candidate A vs. candidate B. In addition to replacing candidate names, I also replace state names and state abbreviations in a similar fashion in order to allow for the possibility that mentioning the state predicts a positive article in the average district, but to avoid these terms becoming fixed effects for states. Similarly, I strip out all digits to avoid estimating parameters for districts.

Second, because articles often include only a small amount of coverage on a candidate within a much larger article, I create a second document term matrix of counts for 1,2, and 3 word n-grams that appear within the 50 characters of the candidate's last name, in either direction, anywhere their name appears. This allows for the possibility that a model should place greater weight on terms that appear near a candidate's name than if they appeared elsewhere in the article. Last, because a large number of terms occur in only a small number of documents, we must employ some sort of feature selection to reduce the number of terms included in the model. I drop all terms that occur in less than 3% of candidate articles in the training set from the analysis.

Political Party Identification Increases Over Time?

When I exclude other features of the article and focus only on characteristics of the candidates, newspapers identify candidates' political party at slightly higher rates as the election approaches, but the effect is only statistically significant in the full sample (Table 2.14, Column 1 and Column 3). Proximity to Election Day influences coverage the most for in-depth coverage in the final 90 days (Column 4), but a roughly 3% increase over the final three months of the came is statistically insignificant and substantively rather small. Without other article features included in the model, I find that newspapers mention challengers' political party and open candidates' political party at slightly higher rates, small but statistically significant effects.

Table 2.14: Political Party Identified? (No Article Features)

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Date (in months)	0.003***	0.001	0.003***	0.01
	(0.001)	(0.005)	(0.001)	(0.009)
District Competitiveness	0.125**	0.061	0.084	-0.009
	(0.049)	(0.054)	(0.053)	(0.073)
Challenger	0.042***	0.05***	0.024*	0.036**
	(0.011)	(0.011)	(0.013)	(0.014)
Open Seat	0.041**	0.048**	0.021	0.046*
	(0.019)	(0.021)	(0.022)	(0.025)
Observations	66730	15421	21174	6172
Districts	367	353	346	302

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Results for Tone of Incumbent Coverage

Tone is measured on a five point scale ranging from 1 = very negative to 5 = very positive. Similar to candidate criticism, however, MCs see no substantively or statistically signficant reward for voting with their constituents (see Table 2.15). Similarly, candidates do not receive a worse tone of coverage in the 7 days following a vote against their constituents (Table 2.16). In High Congruence districts where a substantial share of a newspaper's readers resides, newspapers provide a more negative tone of coverage when they cover more competitive districts, but while statistically significant the effect of moving from a 60/40 district to a 50/50 district is only a 5% shift towards a more negative tone of coverage for in-depth articles in the final 90 days (Table 2.17, Column 4). Newspapers do, however, provide incumbents with a better tone of coverage when articles include pork coverage. Nevertheless,

92.1% [95 % CI (89.9,94.5)] of incumbent coverage is classified as neutral in overall tone. Overall then, the results using tone are consistent with the main analyses using candidate criticism.

Table 2.15: Tone of Coverage and Out of Step Voting in Congress

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Votes Cast With Constituents	0.003	0.002	0.005	0.004
	(0.005)	(0.008)	(0.009)	(0.011)
Abstentions	0.002	-0.004	0.008	-0.006
	(0.01)	(0.019)	(0.016)	(0.033)
District Competitiveness	-0.011	-0.177	0.105	-0.163
	(0.091)	(0.134)	(0.141)	(0.195)
Observations	47412	8184	14106	2935
Districts	307	279	283	229

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.16: Tone of Coverage and Out of Step Voting in Congress

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Votes Cast With Constituents	0.004	0.006		
	(0.005)	(0.009)		
Abstentions	0.002	0.008		
	(0.01)	(0.016)		
Vote Window	-0.024	-0.049		
	(0.037)	(0.048)		
Out of Step in Vote Window	0.023	0.057		
	(0.043)	(0.058)		
District Competitiveness	-0.011	0.104		
	(0.09)	(0.14)		
Observations	47412	14106		
Districts	307	283		

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.17: Tone of Coverage and Out of Step Voting in Congress (High Congruence)

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Votes Cast With Constituents	0.003	-0.004	0.000	-0.002
	(0.005)	(0.01)	(0.01)	(0.013)
Abstentions	0.007	0.019	0.015	0.013
	(0.008)	(0.015)	(0.013)	(0.031)
District Competitiveness	-0.203**	-0.473**	-0.339*	-0.523*
	(0.096)	(0.236)	(0.18)	(0.314)
Observations	33447	5757	10401	2124
Districts	163	153	155	140

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.18: Compensating with Pork? Tone of Coverage and Out of Step Voting in Congress

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Votes Cast With Constituents	0.003	0.001	0.005	0.003
	(0.005)	(0.008)	(0.008)	(0.011)
Abstentions	0.004	-0.004	0.012	-0.003
	(0.01)	(0.019)	(0.016)	(0.032)
District Competitiveness	-0.009	-0.17	0.117	-0.142
	(0.089)	(0.133)	(0.138)	(0.197)
Pork	0.225***	0.188*	0.309**	0.314
	(0.061)	(0.096)	(0.135)	(0.225)
Observations	47412	8184	14106	2935
Districts	307	279	283	229

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Using DW-NOMINATE and Presidential Vote To Measure Extremity

A major difficulty in assessing the quality of policy representation is placing members of Congress and their districts in the same policy space. In the paper, I use MRP-like estimates of district-level opinion on specific bills voted on in the 111th Congress. This strategy has the advantage that it does not assume an ideological dimension, but examines congruence on particular bills. It also has the advantage that it does not confuse consistently liberal (conservative) positions with extremely liberal (conservative) positions (Broockman 2016). In this particular instance, however, this approach has two disadvanatages. First, no polls measure public opinion prior to the vote. Second, the CCES lacks a sufficiently large sample to measure public opinion in every district, so it is necessary to use MRP to estimate district opinion.

So the second strategy I use to measure whether MCs are out of step with their districts is to project congressional districts onto the first dimension of DW-NOMINATE based on 2004 presidential vote, 2008 presidential vote, and MRP-IRT estimates of district preferences from Tausanovitch and Warshaw (2013) with regression.[10] The obvious limitation of this approach is that representative's voting behavior in Congress is, on average, assumed to be representative of their constituents. Nevertheless, we can still analyze the impact of an MC being to the left or the right of where we expect them to be given the relationship between district preferences and MC voting behavior in the broader chamber.

Table 2.19 reports results based on projecting districts into the DW-NOMINATE first dimension. Relatively extreme and relative centrist representatives receive slightly more criticism, but the effect is not significant in the full sample or in all coverage in the final 90 days (Table 2.19, Columns 1 and 2). Relative Extremism does significantly increase candidate criticism for in-depth coverage in the final 90 days. In the final 90 days, a one standard deviation increase in relative extremism predicts a 0.044 or 4.4% increase in candidate criticism (Table 2.19, Column 4). While statistically significant, substantively this is a small shift in the share of coverage that includes candidate criticism and represents only a 1/4 standard deviation change in criticism. I also do not find a significant effect in high congruence districts (see Table 2.20).

In an earlier version of this paper, before I had data on district preferences on major legislation, I used relative extremism and relative centrism as above. I used this approach on the theory that newspapers might punish relative extremism and reward relative centrism. The preanalysis plan, however, called for using the absolute difference between the district projected into the DW-NOMINATE space and the representative's own DW-NOMINATE score. I present that analysis below in Table 2.21 with the addition of the same control

[10] After taking the difference between the predicted DW-NOMINATE and the incumbents actual score I recode this difference into two variables. Relative Extremism is defined as distance to the left (right) of the district for Democrats (Republicans), with MCs closer to the the center than their district placed at zero. Relative Centrism is defined as the distance to the right (left) of the district for Democrats (Republicans), with MCs more extreme than their district placed at zero.

Table 2.19: Candidate Criticism and Out of Step Voting in Congress

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Relative Extremism	0.022	0.072	0.09	0.162*
	(0.03)	(0.049)	(0.056)	(0.088)
Relative Centrism	0.046	0.103	0.038	0.101
	(0.065)	(0.077)	(0.09)	(0.149)
District Competitiveness	0.094	0.279**	0.057	0.277
	(0.075)	(0.113)	(0.121)	(0.196)
Observations	47412	8184	14106	2935
Districts	307	279	283	229

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.20: Candidate Criticism and Out of Step Voting in Congress (High Congruence)

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Relative Extremism	-0.009	0.067	0.044	0.104
	(0.031)	(0.05)	(0.062)	(0.082)
Relative Centrism	0.02	0.098	-0.039	0.112
	(0.081)	(0.169)	(0.163)	(0.19)
District Competitiveness	0.23***	0.367**	0.3**	0.448*
	(0.07)	(0.148)	(0.136)	(0.236)
Observations	33447	5757	10401	2124
Districts	163	153	155	140

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

for district competitiveness. The difference has a statistically significant effect in every specification except the full sample. However, the effect size is similarly small with a standard deviation change in the distance between a representative and their district resulting in only a 2.4% increase in criticism for in-depth article in the final 90 days. Taken together then, the results using DW-NOMINATE to measure ideology are consistent with the vote based results: when a representative is out of step with their district they receive little additional criticism.

Table 2.21: Hypothesis 6B With Controls: Ideologically Distant Incumbents Receive Same Level of Criticism as More Representative Colleagues.

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
MC-District Ideological Distance	0.027	0.078*	0.079*	0.147**
	(0.027)	(0.04)	(0.048)	(0.073)
District Competitiveness	0.075	0.253***	0.098	0.33***
	(0.052)	(0.064)	(0.068)	(0.11)
Observations	47412	8184	14106	2935
Districts	307	279	283	229

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Learning Set Replication

This section replicates the main findings of the paper on the lack of criticism of out of step incumbents using only the candidate newspaper articles with human classifications provided by research assistants. One potential concern could be that the findings in this paper might be influenced by biases, praticularly attenuation, introduced in the machine learning results. This section demonstrates that I also find no punishment for out of step incumbents without using machine learning, but with instead relying on the stratified random sample of coverage included in the human classified learning set. As Table 2.22 shows, casting an important vote with your constituents is at best rewarded with a 2% reduction in criticism. I do find more evidence in the learning set that newspapers heavily punish abstentions, particularly in coverage during the final 90 days of the election where a single abstention is estimated to increase an incumbent's criticism by 33% (Table 2.22, Column 2). Abstentions on such important bills, however, are rare. And while I present results using the learning set in order to show that there is not some major disconnect between the training data and the machine learning results, it is important to remember that in the learning set there are only a handful of random articles per district. This is particularly true for coverage in the final 90 days, so these results are likely heavily influenced by the random draw of articles.

Table 2.22: Training Set Only Candidate Criticism and out of step Voting in Congress

	Incumbent Criticism			
	All	90 Days	In-Depth	90 In-Depth
	(1)	(2)	(3)	(4)
Votes Cast With Constituents	-0.017^{*}	-0.016	-0.015	-0.012
	(0.009)	(0.021)	(0.019)	(0.031)
Abstensions	0.023	0.333^{*}	0.017	0.387^{***}
	(0.019)	(0.185)	(0.034)	(0.132)
District Competitiveness	-0.064	0.303	0.017	1.007^{*}
	(0.143)	(0.296)	(0.278)	(0.590)
Constant	0.297^{***}	0.195	0.442^{***}	0.116
	(0.082)	(0.180)	(0.163)	(0.313)

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

Table 2.23: Training Set Only Candidate Criticism and out of step Voting in Congress (High Congruence

	Incumbent Criticism			
	All	90 Days	In-Depth	90 In-Depth
	(1)	(2)	(3)	(4)
Votes Cast With Constituents	−0.020	−0.033	−0.009	−0.035
	(0.013)	(0.027)	(0.025)	(0.036)
Abstensions	0.005	0.220	−0.017	0.176
	(0.015)	(0.296)	(0.036)	(0.219)
District Competitiveness	0.049	0.523	0.162	1.532**
	(0.220)	(0.529)	(0.426)	(0.686)
Constant	0.276**	0.199	0.346	−0.016
	(0.121)	(0.272)	(0.238)	(0.375)

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

Preanalysis Plan Analyses

Many of the analyses in the paper closely follow the analyses proposed in the in the pre-analysis plan. Three major changes were made between the preanalysis plan and the paper. First, for many of the analyses district competitiveness, as measured by distance from a 50/50 presidential vote, was added as a control variable in order to account for differences in coverage based on district competitiveness. Second, the preanlaysis plan and the first draft of this paper focused primarily on tone rather than criticism, though all proposed analyses of tone in the preanalysis plan were also proposed for criticism. The appendix shows similar results using tone, but the main body of the paper uses candidate criticism rather than tone because the SuperLearner is much more accurate out of sample for criticism than for tone. Third, in the paper I use estimates of voter preferences on individual votes rather than placing voters on a more general ideological measure like DW-NOMINATE. The advantages of using individual issues/votes in Congress over a more general measure of ideology are discussed in detail in the appendix on results using DW-NOMINATE, but I did not have the vote data when I wrote the preanalysis plan and substantively the two approaches produce similar results. Finally, I also added additional analyses, most notably of high congruence districts and examining the impact of a referral to the House Ethics Committee Below are all of the analyses proposed in the preanalysis plan as they were proposed.

Table 2.24: Hypothesis 1: Challenger's Relative Share of Coverage Increases as Election Approaches

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Date (in months)	0.018***	0.071***	0.022***	0.087***
	(0.001)	(0.011)	(0.002)	(0.016)
Observations	62185	13499	19592	5280
Districts	333	319	314	272

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.25: Hypothesis 2: Articles Mores Likely to Mention A Candidate's Party as the Election Approaches

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Date (in months)	0.004***	0.005	0.004***	0.013
	(0.001)	(0.005)	(0.001)	(0.009)
Observations	66730	15421	21174	6172
Districts	367	353	346	302

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.26: Hypothesis 3: Articles About Challenger More Likely to Include a Policy Stance as the Election Approaches

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Date (in months)	0.001	0.014	0.013***	0.046***
	(0.002)	(0.011)	(0.002)	(0.016)
Observations	10707	4475	3853	1980
Districts	286	263	226	204

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.27: Hypothesis 4A: More Negative Tone of Coverage as the Election Approaches, Particularly in Close Elections

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Date (in months)	0.006*	0.111***	0.011**	0.043
	(0.003)	(0.039)	(0.005)	(0.062)
Race Competitiveness	0.183	4.578**	0.429*	2.132
	(0.128)	(1.932)	(0.219)	(3.378)
Date (in months) * Race Competitiveness	-0.016*	-0.226**	-0.031**	-0.108
	(0.008)	(0.094)	(0.014)	(0.162)
Observations	66730	15421	21174	6172
Districts	367	353	346	302

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.28: Hypothesis 4B: More Candidate Criticism as the Election Approaches, Particularly in Close Elections

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Date (in months)	-0.005***	0.002	-0.005	0.012
	(0.002)	(0.017)	(0.003)	(0.037)
Race Competitiveness	-0.14*	-0.05	-0.159	-0.09
	(0.081)	(0.987)	(0.122)	(2.009)
Date (in months) * Race Competitiveness	0.019***	0.017	0.025***	0.02
	(0.005)	(0.047)	(0.008)	(0.096)
Observations	66730	15421	21174	6172
Districts	367	353	346	302

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.29: Hypothesis 5: More Horse Coverage as the Election Approaches, Particularly in Close Elections

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
Date (in months)	0.000	0.105***	0.002	0.146***
	(0.002)	(0.026)	(0.004)	(0.049)
Race Competitiveness	-0.145*	2.431*	-0.186	4.865*
	(0.083)	(1.409)	(0.164)	(2.655)
Date (in months) * Race Competitiveness	0.045***	-0.083	0.046***	-0.203
	(0.006)	(0.067)	(0.011)	(0.125)
Observations	66730	15421	21174	6172
Districts	367	353	346	302

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.30: Hypothesis 6A: Ideologically Distant Incumbents Receive Same Tone of Coverage as More Representative Colleagues.

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
MC-District Ideological Distance	-0.052	-0.158**	-0.058	-0.122
	(0.04)	(0.078)	(0.067)	(0.093)
Observations	47412	8184	14106	2935
Districts	307	279	283	229

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.31: Hypothesis 6B: Ideologically Distant Incumbents Receive Same Level of Criticism as More Representative Colleagues.

	(1) All	(2) 90 Days	(3) In-Depth	(4) 90 In-Depth
MC-District Ideological Distance	0.042*	0.124***	0.1**	0.217***
	(0.024)	(0.038)	(0.044)	(0.072)
Observations	47412	8184	14106	2935
Districts	307	279	283	229

Robust standard errors clustered by congressional district in parentheses.
Estimates weighted by congressional district.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.32: Descriptive Statistics on Candidate Coverage

	In-Depth Articles	Tone	Policy Position	Pork Coverage
All	0.259	3.050	0.345	0.050
Incumbent	0.251	3.050	0.384	0.062
Challenger	0.284	3.050	0.226	0.016
Close Election	0.298	3.050	0.320	0.042
Open Seat	0.303	3.070	0.188	0.012

Table 2.33: Descriptive Statistics on Candidate Coverage (continued)

	Horse Coverage	Candidate Criticism	Policy Criticism	Personal Criticism
All	0.277	0.082	0.041	0.041
Incumbent	0.180	0.084	0.041	0.043
Challenger	0.608	0.092	0.047	0.045
Close Election	0.405	0.097	0.052	0.045
Open Seat	0.590	0.078	0.035	0.042

Human Content Analysis

Research assistants received their assigned articles in an HTML file that they could open in their web browser and leave open as they completed the survey. Above each article was a unique Article ID and the candidate's name. The first and last name of the candidate being analyzed were both in bold anywhere they appeared in the text of the article in order to help the reader identify passages about the candidate. Because the unit of analysis is the candidate-article every article is read and classified from the perspective of the candidate named above the article. All answers are entered via the online survey platform Qualtrics. When reading an article, research assistants first identify who is classifying the article and enter both the Article ID and the candidate's name into Qualtrics. After entering the ID, the survey asks them to "Please read the newspaper article carefully before continuing. When you are done reading the article, hit next." After reading the article and hitting next, the research assistant is asked whether the article was primarily about the candidate. If the article did not appear to mention a political candidate the survey terminated. If the article did mention a political candidate, the research assistant was asked about the tone of coverage, whether the article mentions a policy stance of the candidate, (and if so the policy and perceived ideology on that issue), whether the article discusses a local pork project, includes horse race coverage, includes criticism of the candidate, the source of any candidate criticism, whether the article is an editorial endorsement, who receives an endorsement, the perceived ideology of the candidate, and the political party of the candidate according to the article. The second set of research assistants who read only in-depth articles were also asked if the article portrayed the candidate as out of step with the district and whether an article was in the news or editorial section. For articles that did at least mention the political candidate, research assistants took, on average, one minute to read the article and another 90 seconds to complete the survey.

The full codebook of questions, answers, and instructions is included below. This codebook was developed in consultation with the undergraduate research assistants by reading and classifying roughly 100 articles per person with three discussions of how to improve the codebook over a three week period. During this time I was able to look at inter-coder reliability and focus on the phrasing and coding instructions for questions with low inter-coder reliability.

Candidate Coverage Codebook

Screener Question

Was this article primarily about the candidate?

1) This article does not appear to mention a political candidate.
2) The article only briefly mentions the candidate.
3) The article discusses the candidate in some detail.
4) The candidate is a major focus of the article.
5) The candidate is the primary focus of the article.

Tone

Was the tone of coverage towards the candidate very negative, negative, neutral, positive, or very positive?
1) Very Negative
2) Negative
3) Neutral
4) Positive
5) Very Positive

Policy

Does this article mention a policy stance of the candidate?
1) Yes
2) No

An article includes a policy stance if it describes a belief, vote, statement, or any other action of a candidate that explicitly or implicitly identifies a candidate's position on a policy.

Policy Area (if yes to policy)

Which category below best describes the policy area in which the candidate took a position? [list of options]
If a candidate took a position in more than one policy area please select the policy area that played a bigger part in the article.

Policy Ideology

Place the candidate's policy stance as conveyed in this article on an ideological scale:
1) Very Liberal
2) Liberal
3) Somewhat liberal
4) Moderate
5) Somewhat Conservative
6) Conservative
7) Very Conservative

Pork

Does this article discuss a local project for the district? (A particularized good for constituents, e.g. specific spending for a bridge or health clinic in the district)
1) Yes
2) No
An article discusses a local project if it discusses a local spending project that delivers a concentrated good for constituents within the district, e.g. specific spending for a bridge or health clinic in the district.

Horse Race

Does the article include "horse race" coverage about the state of the electoral campaign?
1) Yes
2) No
An article includes "horse race" coverage if any aspect of an election or an electoral campaign is discussed.

State of Race (if Horse Race yes)

What is the state of the race according to the article?
1) The candidate is widening their lead.
2) The candidate is holding a steady lead.
3) The candidate is in the lead but losing ground.
4) The two candidates are tied.
5) The candidate is in second place but gaining ground.
6) The candidate is holding steady in second place.
7) The candidate is in second place and losing ground.
8) Unclear

Criticism

Does the article include criticism of the candidate?
1) Yes
2) No
An article includes criticism of a candidate if any portion of the article criticizes a trait, action, or position implicitly or explicitly linked to the candidate.

Type of Criticism (if criticized)

What best describes the type of criticism?
1) The article criticizes the candidate personally.
2) The article criticizes a policy position of the candidate.

Source of criticism (if criticized)

Who is criticizing the candidate? (check all that apply)
The article paraphrases others to criticize the candidate.
The article uses quote(s) from unnamed sources that criticize the candidate.
The article uses on the record quote(s) that criticize the candidate.
The article criticizes the candidate without attribution.
This article is a letter to the editor.

Source of criticism 2 (if criticized)

Is the candidate being criticized directly by their opponent, their opponent's campaign, or a member of the opposing party?
1) Opponent

2) Opponent's campaign
3) Opponent's party
4) None of the above

Out-of-Step
Does the article portray the candidate as out-of-step with their district?
1) Yes
2) No
A candidate is portrayed as out-of-step with their district if the article characterizes the candidate as having different political positions or values than the typical voter in their district.

Out-of-Step 2 (if yes)
Does the article portray the candidate as too liberal or too conservative for the voters in their district?
1) Too liberal
2) Too conservative
3) Unclear

Scandal
Does this article mention a political scandal that the candidate was involved in?
1) Yes
2) No

Type of Scandal
What type of scandal was the candidate involved in?
1) Corruption scandal
2) Sex scandal
3) Other [text box]

Endorsement
Was the article an editorial endorsement?
1) Yes
2) No
Endorsement Type (if endorsement)
1) Endorsed the candidate
2) Endorsed the candidate's opponent
3) Declined to make an endorsement

Ideology
Based on this newspaper article, place the candidate on the ideological scale below:
1) Very Liberal
2) Liberal

3) Somewhat liberal
4) Moderate
5) Somewhat Conservative
6) Conservative
7) Very Conservative

Political Party

What is the political party of the candidate according to the article?
1) Democratic
2) Republican
3) Third Party
4) This article does not explicitly identify the candidate's political party.

Section

Was the article in the news section or the editorial section?
1) News
2) Editorial

Letter to the editor [if editorial]

Was this article a letter to the editor?
1) Yes
2) No

Name Recognition

Do you recognize the candidate's name, [candidate name], from outside of this project?
1) Yes
2) No

Chapter 3

Face Value? Experimental Evidence that Candidate Appearance Influences Electoral Choice

If the previous chapter demonstrates that local newspapers rarely criticize incumbent representatives, even for voting against a majority of their constituents, this chapter examines what happens when voters lack information. In particular, it shows that when voters lack other information they sometimes vote based on candidate appearance. Prior research had shown that candidates' looks predict voters' choices, but this observational finding is vulnerable to an alternative explanation: candidates who work hard to get elected also work hard to get a better photograph. This chapter uses two experiments to rule out alternative explanations and show that candidate appearance causes candidates who look the part to get more votes.[1] The chapter is from a paper published in *Political Behavior* titled "Face Value? Experimental Evidence that Candidate Appearance Influences Electoral Choice." that was coauthored with Gabriel Lenz, Douglas J. Ahler, and Jack Citrin.[2]

3.1 Paper Abstract

According to numerous studies, candidates' looks predict voters' choices—a finding that raises concerns about voter competence and about the quality of elected officials. This potentially worrisome finding, however, is observational and therefore vulnerable to alternative explanations. To better test the appearance effect, we conducted two experiments. Just be-

[1]The studies were approved by the Committee for the Protection of Human Subjects (CPHS) at the University of California, Berkeley. All procedures performed in studies involving human participants were in accordance with the ethical standards of CPHS and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

[2]The chapter thus uses "we" to refer to the authors. The chapter also references an Online Appendix, which accompanied the paper. Both are available online at: `http://link.springer.com/article/10.1007/s11109-016-9348-6`

fore primary and general elections for various offices, we randomly assigned voters to receive ballots with and without candidate photos. Simply showing voters these pictures increased the vote for appearance-advantaged candidates. Experimental evidence therefore supports the view that candidates' looks could influence some voters. In general elections, we find that high-knowledge voters appear immune to this influence, while low-knowledge voters use appearance as a low-information heuristic. In primaries, however, candidate appearance influences even high-knowledge and strongly partisan voters.

3.2 Introduction

On what basis do voters decide? The answer bears directly on debates over citizens' competence and the quality of the officials they elect. Confronted by evidence of widespread ignorance, ideological innocence, and the paucity of issue voting (Converse 1964; Delli Carpini and Keeter 1996; Lenz 2012), scholars have looked to heuristics—simple rules of thumb—to haul voters onto the shores of rationality (Fiorina 1981; Key 1968; Lupia 1994; Popkin 1991). Voters undoubtedly do rely on informative heuristics, such as the state of the economy, party ties, and feelings about the incumbents. But how often do simple rules of thumb lead them astray from a more informed and appropriate choice (Kuklinski and Quirk 2000)?

One potentially worrisome heuristic is a candidate's appearance. Endorsing Mitt Romney for the 2012 Republican presidential nomination, Bob Dole declared, "So it looked to me like it would be either Romney or Newt [Gingrich] for the nomination, but Romney looks like a president" (Laskas 2012, 88). Whether candidates look like presidents may not be entirely uninformative, but seems unlikely to provide much information about the candidates—a point we return to in the conclusion. Nevertheless, research implies that some voters evaluate candidates as Dole did in 2012. They vote for politicians whose appearance in photographs is judged more competent or attractive by nave raters (those who neither know nor recognize the candidates) at higher rates in actual U.S. Senate and House elections, as well as in elections abroad (Atkinson, Enos, and Hill 2009; Ballew and Todorov 2007; Banducci et al. 2008; Berggren, Jordahl, and Poutvaara 2010; Hall et al. 2009; King and Leigh 2009; Lawson et al. 2010; Mattes et al. 2010; Olivola and Todorov 2010; Rosar, Klein, and Beckers 2008; Spezio et al. 2012; Todorov et al. 2005).

While these observational studies find that candidate appearance correlates with actual election results, these studies do not show that candidate appearance actually causes voters to change their minds in real-world elections. It is troubling for the quality of electoral choices if voters conflate mere physical appearance—stylish hair, straight teeth or a strong jaw—and actual competence, but the results of observational research cannot rule out several alternative explanations for the correlation between appearance and vote choice, leaving the causal mechanism ambiguous.

Do voters rely on the seemingly superficial heuristic of appearance when voting? Or is there an alternative explanation for the correlation between appearance and vote choice that reveals more competent voters? Foremost among the possible alternative explanations is

the influence of candidate effort. Candidates who campaign harder or have more resources may look better because they also put more effort into their press materials (including photographs) or because their superior resources can pay for professional photographers, image consultants, $600 haircuts, and the like. But if this is the case then the observationally estimated effects of candidate appearance could be entirely spurious, an artifact of the correlation between improved candidate appearance and other facets of campaign effort—voter mobilization, showing up at events, outreach and communication, etc.—that may be the real drivers of voters' choices.

Campaign effort—whether by the candidate or her party—could thus make candidates' looks appear to influence voters when they do not do so directly. Indeed, candidates who outspend their opponents do look better to nave raters. Examining 2006 Senate races, Lenz and Lawson (2011, 584-5) find a 0.59 correlation between a candidate's spending advantage and appearance advantage and a 0.56 correlation between incumbency and appearance advantage. However, after controlling for variables that might capture campaign effort, such as race competitiveness and candidate spending, Atkinson, Enos, and Hill (2009) find that candidate appearance still has a small effect on vote share.[3] Given these observational findings, how should we assess the possibility that the appearance-vote findings are spurious rather than causal?

Sorting out causation here is hard. Statistically controlling for variables such as competitiveness, spending, or incumbency is appropriate only if these variables are causes of candidate appearance, not consequences. If appearance is in fact causally prior to spending—e.g., better looking candidates can raise more money, win endorsements based on looking the part (like Governor Romney), or are likely to have won previous elections (and so become incumbents)—then researchers should not control for these variables because they could be consequences of candidate appearance (i.e., post-treatment).[4] Put differently, if candidates can raise more money or attract more volunteers because they are better looking, then controlling for such variables will bias estimates of the candidate appearance effect downward by incorrectly attributing part of the true effect of appearance to these variables.[5] Given

[3]Several other observational results are inconsistent with the alternative explanation emphasizing the causal influence of campaign effort entirely explaining the observed effect of candidate appearance. Specifically, the effect of the candidates appearance holds when professional photographers took the pictures in a standard format (Antonakis and Dalgas 2009; Klein and Rosar 2005), and when one statistically controls for differences in image quality and other aspects of the pictures, such as visible light (Lawson et al. 2010; Rosar, Klein, and Beckers 2008). Additionally, appearance-advantaged candidates win in competitive races, where the candidates should be more comparable in quality and in resources (Antonakis and Dalgas 2009; Benjamin and Shapiro 2009). They also perform disproportionately well in systems where legislators compete against members of the same party (Berggren, Jordahl, and Poutvaara 2010) and in non-partisan contests (Banducci et al. 2008; Martin 1978).

[4]In general, researchers should not control for variables that intervene between the treatment and the outcome, in this case, between candidate appearance and vote share. For a general discussion, see King (1991, 1049-50).

[5]When estimating the effect of challenger appearance, Atkinson et al. (2009) carefully try to avoid post-treatment bias by measuring district competitiveness at least one year before the general election, when the challenger's identity is less clear (using the Cook Political Report). Nevertheless, these experts may already

the possibility for complex causal relations among these variables, drawing firm inferences with observational data may be impossible.[6]

For these reasons, we test the influence of a candidate's appearance on voters with two experiments rather than with observational studies. We interviewed individuals just before an election in which they said they would likely participate and asked for their voting intention. Crucially, however, we randomly assigned participants to one of two conditions: (1) a control group received a ballot designed to resemble the one actually used, and (2) a treatment group received a ballot that also showed candidates' photos next to their names. To evaluate whether appearance directly influences votes, we simply compare the degree to which candidate appearance predicts vote intentions in the two conditions.

This research design sheds light on the appearance-vote relationship in a way that previous studies could not. It does so because random assignment rules out the alternative explanations. Since candidate effort—raising more money, shaking more hands, kissing more babies—cannot differentially influence voters in the photo condition, any effect we detect must be a result of viewing the photos of the candidates.[7] The experimental design can thus provide internally valid evidence that candidate appearance influences vote choice, even when real world voters have other information about the candidates. However, random assignment only provides clean causal identification within the experiment on the mock ballots, which differ from actual ballot results and therefore do not necessarily provide generalizable (externally valid) estimates. Thus, this design may not tell us how much candidate appearance really matters to election outcomes, but that it could matter. Given concerns that the entire appearance-vote correlation could be from omitted variable bias, our studies make an important contribution by providing experimental evidence that at least a portion of the appearance-vote correlation is causal, even if they do not tell us precisely how much of that correlation is due to this causal effect.

We ran this experiment on 14 House races in the 2012 California congressional primary and 44 statewide races across 18 states in the 2012 general election. Using nave raters' assessments of candidate appearance, we find that including candidates' photographs on the ballot does indeed lead participants to vote more often for appearance-advantaged candidates. Using actual candidates in the midst of an election as stimuli, we find that a substantively significant percentage of our participants (9% in the general election races) voted differently than they otherwise would because they saw candidate photos. Although our emphasis is on the existence of an appearance effect, not its magnitude, we note that at face value this shift would be large enough to change the outcome in roughly 29% of primary races and

know the likely challengers and so may be influenced by their looks (making these ratings post-treatment).

[6]Indeed, Atkinson et al. (2009, 236) are careful not to interpret their regression coefficient for incumbent appearance as a causal estimate. They suggest instead that appearance-advantaged incumbents (as challengers in a prior election) disproportionately select into competitive districts, which would bias their estimate of incumbent appearance downwards. This downward bias and, more generally, the causal complexity of observational studies on appearance provide reasons to turn to experimental studies such as ours.

[7]Of course, candidates' efforts to "improve" their appearance, as revealed through their photos, may contribute to any such causal effects.

about 14The rest of the paper proceeds as follows: First, we describe our research design more fully and present candidate-level results for House primaries in California. Second, we replicate these results in higher-salience, statewide general election races. Third, we consider external validity and assess the robustness of our findings. Fourth, we conduct individual-level analyses for both studies and show that candidate appearance most heavily influences low-information voters and matters more in the earlier stages of a campaign. This important nuance in our results helps us assess the ramifications of these studies for voter competence and democratic accountability.

3.3 Study 1: Appearance Advantage in the 2012 California House Primaries

Design and Procedures

Starting 10 days before the 2012 California primary, an Internet poll conducted by Survey Sampling International (SSI) interviewed 1,268 registered voters from 14 of California's 53 House districts. The sample adequately represents registered voters on age, party registration, and political ideology. Fifty-three candidates ran in these 14 districts—11 females and 42 males.[8] In terms of partisanship, there were 23 Republicans, 23 Democrats, and 10 no-party preference or other party candidates. Importantly, the survey's election results closely mirror the actual election results. (See the online appendix [OA], section 1.1, which also presents the demographic characteristics of participants. The OA is available with the online version of this article at the journal website.) As noted, we randomly assigned participants to one of two conditions. Those assigned to the control condition received a ballot identical to the one they would see in the actual June 5 election (a top-two primary ballot with all candidates listed, regardless of party). In the treatment condition, we gave participants the same choice of candidates but also displayed black-and-white photographs of the candidates' faces next to their names. Figure 3.1 provides an example.

We measured the appeal of candidates' appearances in a separate survey by showing U.S. workers on Amazon's Mechanical Turk the photos and asking, "How good of a congressperson do you think this person would be?" (See OA section 1.2 for survey details.) We use this general measure to sidestep the debate about which traits voters primarily respond to—e.g., competence (Todorov et al. 2005) or attractiveness (Banducci et al. 2008).[9] Our

[8]We also ran the experiment in six California State Senate races. We do not pool these races with the House primaries in the analysis because photograph quality was noticeably lower. Instead, we present these results in OA section 1.2. Including them in the main analysis leaves our key findings unchanged.

[9]No matter what measure of appearance we choose, that measure will also pick up other characteristics that correlate with it. One way to break these correlations would be to artificially alter candidate photos in order to experimentally vary these traits, but using altered photos would significantly reduce the external validity of our experiments. We thus use actual candidate photos and make no claim about the particular aspect of a candidates appearance that influences voters.

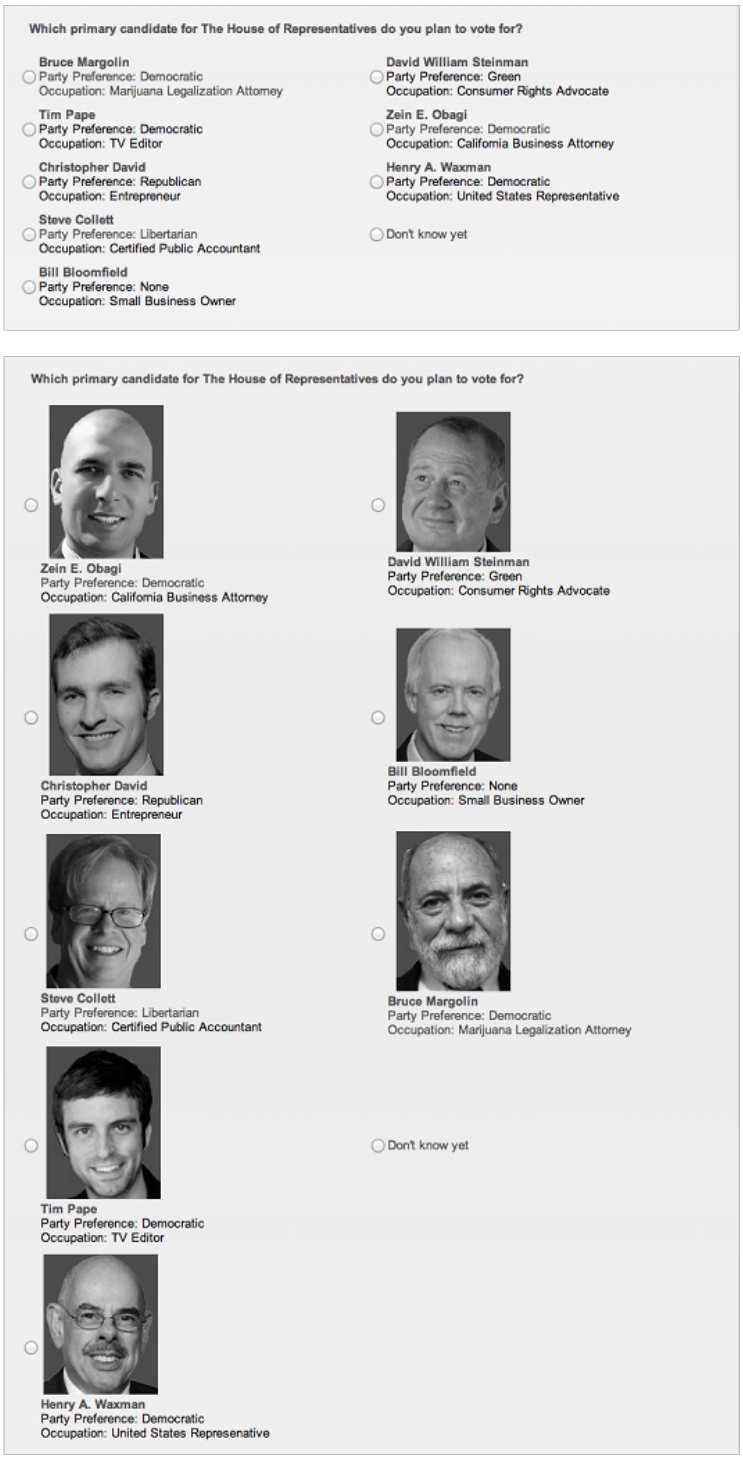


Figure 3.1: Example of Control (Top) and Treatment (Bottom) Ballots in California Primary Experiment (Study 1). Note: The ballots showed the same information as the California 2012 primary ballot, except of course for the photos. This example is from district 33.

measure, however, correlates more strongly with perceptions of facial competence (.61) and attractiveness (.60), than with perceptions of dominance (.19, see OA section 2.3 for more detail). This general appearance measure predicts election outcomes as well as specific trait-based measures (Lawson et al. 2010). To obtain nave ratings, we excluded ratings from California Mechanical Turk workers and from workers who reported recognizing the candidate (just 0.1% of 9174 ratings). We gave participants in the experiments five options for their evaluations of "how good of a congressperson do you think this person would be," ranging from "much better than average" to "much worse than average." Mean candidate ratings ranged from a low of 2.4 on the five-point scale (Jim Reed, a Democratic candidate from the 1st District) to a high of 3.5 (Mary Bono Mack, an incumbent Republican from the 36th District), with a mean of 2.9 and a standard deviation of 0.27. For the analysis below, we recode this measure so it captures Appearance Advantage within a district by subtracting from a candidate's (e.g. Mary Bono's) raw appearance rating the mean rating of all the candidates in her district. We rescale this variable to a 0-1 scale, but the results are robust to other coding procedures (see OA Section 1.5).

Our dependent variable is a candidate's vote share in the treatment condition minus her vote share in the control condition (Photo Condition minus No-Photo Condition Vote Share). This measures the degree to which a candidate performs better or worse when voters see her face and her opponents' faces on the ballot. If appearance matters, candidates should receive an increasing vote share in the treatment condition as their Appearance Advantage grows. About 70% of participants reported an intended vote choice—congressional primaries have low salience—so we lose 30% of participants when we calculate the dependent variable. (The voting rate was similar across conditions.) We also exclude an additional 2% because they said they would not vote in the actual election, leaving us with 851 participants. (The results are similar in the full sample.)

External Validity

The experimental design provides internally valid evidence that candidate appearance can influence vote choice. In US elections, of course, ballots do not show photos. So, to what extent do the findings help us understand the appearance effect in real-world elections? This question is difficult to answer. To make the findings as generalizable as possible, we took several steps: (1) we only interviewed registered voters, (2) we excluded respondents who said they would not vote, (3) we interviewed respondents close to Election Day, (4) and we showed respondents ballots with actual candidates from their own districts. Despite these efforts, however, there are reasons why we may be overestimating the real-world appearance effect, but there are also reasons why we may be underestimating it. We therefore devote a section of the paper, after presenting the results of both studies, to discussing external validity.

Candidate-Level Results

Figure 3.2 presents evidence that showing photos of the candidates to voters just before the election influences how they vote. The vertical axis presents the difference in candidate vote share between the photo (treatment) and no-photo (control) conditions, while the horizontal axis places candidates according to their within-district appearance advantage. The positive trend in Figure 3.2 indicates that appearance-advantaged candidates—that is, those rated highly by the nave Mechanical Turk raters—do receive more votes when voters see photos on the ballot.

For example, incumbent Democrat Henry Waxman (CA-33) suffers from a considerable appearance disadvantage, scoring poorly on the relative appearance measure (0.13). When participants saw the candidates' photos (which we show in the example ballot in Figure 1), support for Waxman dropped by about 10 percentage points. In contrast, Waxman's most appearance advantaged opponent, an unknown Democratic candidate named Tim Pape, scored near the top of the appearance measure (0.81) and received almost a 15-point boost in vote share when voters saw the photographs in the multicandidate primary. Waxman also lost out to "no party preference" (NPP) candidate Bill Bloomfield, who scored well on the appearance measure (0.62) and received a 10-point boost in vote share in the treatment group. Bloomfield went on to lose only narrowly to the powerful incumbent Waxman in the general election.

Critically, this effect cannot result from Pape or Bloomfield exerting more campaign effort or strategically choosing to enter this race. Since we randomly assigned voters to the photo and no-photo conditions, the candidates' campaigns—mostly absent in Pape's case anyway—could not have disproportionately influenced the treatment group.

This pattern holds on average across candidates, as the scatterplot in Figure 3.2 shows. The slope of the best-fit line through the data points is 0.21, indicating that if the most appearance-advantaged and appearance-disadvantaged candidates (across all districts) received equal vote shares in the control condition, we would expect to see a 21-point difference in vote shares in the appearance-advantaged candidate's favor in the photo condition. More typically, a one standard deviation increase in appearance advantage, coupled with the treatment ballot, yields a 5.4-point boost, an effect that could alter outcomes in several of these primaries. The bivariate regression analysis in Table 3.1 presents the estimated slope of the best-fit line in Figure 3.2 and shows that this 0.21 slope is unlikely to occur by chance alone. (The 95% confidence interval [95% CI] for the estimate is 0.10 to 0.31).This effect is robust across important categories of candidates. As Columns 2 and 3 in Table 3.1 show, appearance-advantaged incumbents and challengers both benefit from having their photo shown. The next two columns show that appearance affects the fortunes of both viable and nonviable candidates, indicating that the photos did not simply serve to remind participants of real political information they had previously learned about prominent, photogenic candidates.[10] Finally, Columns 6 and 7 show that appearance matters for both Democratic and

[10]The authors and a team of research assistants used endorsements, campaign finance data, previous office, number of competing co-partisans, and vote share in previous elections to classify candidates as viable

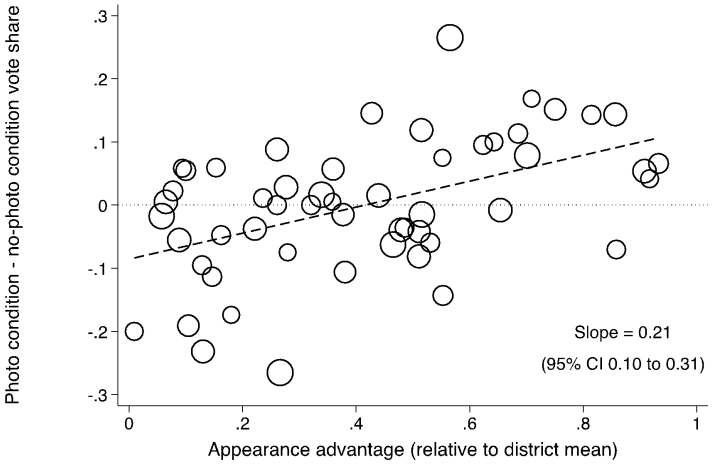


Figure 3.2: Appearance-Advantaged Candidates in House Primaries Benefit in Photo Condition. Note: Each point represents a candidate in the 2012 US House primaries in California. Observations weighted by the total number of respondents reporting a vote choice in the candidate's district.

Republican politicians: appearance-advantaged candidates from both parties fared significantly better in the treatment condition.

In addition, candidates' race and gender fail to explain the appearance effect. Column 8 of Table 3.1 presents the estimated effect of appearance-advantage, controlling for candidate race, gender, and incumbency status.[11] Interestingly, the photos appear to change the impact of these characteristics on voting intentions: incumbent and male candidates perform significantly worse, and white candidates significantly better on the photo ballot. Even so, neither the substantive nor the statistical significance of the appearance advantage effect changes when we control for these covariates.

A potential concern is that voters' reactions to especially unprofessional looking candidates may drive this result. For some candidates, the only pictures available showed them wearing T-shirts or with disheveled hair. Rather than reacting to candidates' physical features, voters could be inferring candidates' competence based on their unprofessional presentation. However, when we exclude races with the most poorly rated candidates, those below the 25th percentile on overall photograph ratings, we find a larger treatment effect, not a smaller one (Table 3.1, Column 10, see OA section 1.6 for scatterplots and analyses

or nonviable before the election. These assessments were largely holistic but are validated by the actual election results: A nonviable-classified candidate finished ahead of a viable-classified candidate in just 2 of the 13 races under study here, and neither of those candidates came close to advancing to the general election.

[11]Ideally, we would address race and gender not with controls but by restricting the analysis to candidates matched on race and gender, but only three of the 14 races in Study 1 were so matched. We are, however, able to conduct this analysis in the second study.

Table 3.1: Appearance-Advantaged Candidates in House Primaries Benefit in Photo Condition (Dependent variable: Photo condition minus no-photo condition vote share).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	All	Challengers	Inc.	Non-viable	Viable	Dem.	Rep.	All with controls	Min. Rating <= 25th percentile	Min. Rating > 25th percentile
Appearance advantage	0.21***	0.18***	0.30**	0.14**	0.19**	0.25***	0.19**	0.15***	0.11**	0.28***
	(0.05)	(0.05)	(0.14)	(0.06)	(0.09)	(0.08)	(0.08)	(0.05)	(0.05)	(0.08)
Incumbent								-0.09***		
								(0.03)		
White								0.05		
								(0.03)		
Male								-0.10***		
								(0.04)		
Constant	-0.09	-0.06	-0.17	-0.03	-0.10	-0.15	-0.04	0.01	-0.04	-0.12
	(0.03)	(0.03)	(0.07)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.02)	(0.04)
Candidates (N)	53	39	14	24	29	23	23	53	16	37
R2	0.23	0.24	0.28	0.18	0.14	0.30	0.23	0.41	0.25	0.26

Note: This table shows candidate-level regressions (each column showing a separate model) with standard errors in parentheses. The column 1 regression simply shows the 0.21 slope of the best-fit line in Figure 2. The number of participants in these regressions is 851, all of whom are registered voters. We do not cluster the standard errors at the district level because, with only 14 clusters, clustering is unreliable (Angrist and Pischke 2009, ch. 8), but we do cluster by candidate in the individual-level analysis shown in OA section 4.5. * p<0.1, ** p<0.05, *** p<0.01

with other cutoffs, which yield similar results). Thus, voters do not appear to be simply reacting negatively to candidates whose photos may signal incompetence.[12]

3.4 Study 2: Appearance Advantage in the 2012 General Elections

The results from the primary study imply that candidates' looks can directly affect voters. Our experimental design enables us to reject the argument that candidate effort or other confounding variables fully account for previous observational findings of the appearance effect. However, because voters often pay little attention to primary campaigns and know little about the candidates, they may be more likely to rely on appearance as a low-information heuristic in this electoral setting. Would these results also hold in general election races? We examine this in Study 2.

Design and Procedures

In the three weeks prior to the general election on November 6, 2012, we recruited 2,235 participants across 18 states through Amazon's Mechanical Turk. We asked these participants about their likely vote choices in anywhere between one and nine statewide races, which ranged from higher-salience races (i.e., gubernatorial or senatorial) to down ballot races (e.g., attorney general).[13] We selected the states before running the study because they possess sufficiently large populations of Mechanical Turk workers. The 44 contests included: 15 U.S. Senate races; three for governor; four each for attorney general and state treasurer; three each for lieutenant governor and secretary of state; two each for commissioner of insurance; state auditor, and superintendent of public insurance; and one for agricultural commissioner; labor commissioner; public land commissioner; railroad commissioner; presiding judge court of criminal appeals; and university board of regents. (See OA section 2.1.1 for a list of races.) The candidates included 26 females and 62 males. Of the 2,235 participants in the study, we excluded four percent because they failed an attention test and an additional 10 percent because they reported not intending to vote, leaving us with 1,933 participants. (The results are similar in the full sample; see OA section 2.2.)

As in Study 1, we randomly assigned participants to one of two conditions: a standard ballot or a ballot with candidate photographs. Participants assigned to the control condition

[12]Finally, to address separate concerns about candidate vote share not being independently distributed within district, OA section 4.6 shows that the photo-ballot respondents were 10 percentage points more likely to vote for the most appearance-advantaged candidate in their district compared to control ballot respondents ($p < 0.001$).

[13]Survey dates: October 17-November 2. Election Day was November 6. We also asked participants about a handful of multicandidate races and single-candidate judicial retention elections. Since analyzing races with only one or more than two candidates introduces complications, we relegate analysis of these races to the OA (see OA section 5.1). The results are consistent with the overall findings in the paper.

received a mock ballot that included the statewide races they would see on the actual ballot on Election Day. In the treatment condition, we gave participants the same ballot but added black-and-white candidate photographs. In both conditions, the ballot showed the candidate's party affiliation.

We then measured candidate appearance in a separate survey on Amazon's Mechanical Turk, asking participants "How good of an elected official (e.g. Sen. or Gov.) do you think this person would be?" (See OA section 2.3 for survey details). To obtain nave assessments, we excluded ratings when participants reported recognizing the candidate. Participants rated the candidates on the same 1-5 scale as used in the previous study, and their ratings have a mean of 3.1 and a standard deviation of 0.32. We measure Appearance Advantage in the direction of the Republican candidate (Republican Appearance minus Democratic Appearance), rescaled to a 0-1 range (see Figure 3.3 for the ratings). Since all these contests featured just two candidates—one Democrat and one Republican—we can conduct the analysis at the race-level as opposed to the candidate-level. Thus, our dependent variable is the Republican vote share in the photo condition minus the Republican vote share in the no-photo condition (Photo Condition minus No-Photo Condition Vote Share).

Race-Level Results

Even in general election races, voting intentions differed (in the aggregate) when the ballot included photos of the candidates' faces. Figure 3.3 plots the relationship between Republican appearance advantage (horizontal axis) and the difference between the Republican candidate's treatment vote share and control voter share (vertical axis). The slope of the regression line is 0.20, implying that the Republican candidate with the largest appearance advantage—Bob Corker in the Tennessee Senate race—is predicted to benefit by 20 percentage points over the most appearance-disadvantaged Republican—Steve Royal in the North Carolina State Treasurer race—when voters see candidates' faces on the ballot. Put in terms of a difference we are more likely to observe, a one standard deviation improvement in Republican appearance advantage, coupled with showing photographs to voters, would yield an expected 4-point boost in vote share for the Republican candidate. The estimated slope is significant (95% CI: [0.05, 0.35]) in a bivariate regression (see Column 1, Table 3.2) and robust to a number of alternative specifications.

As we noted earlier, a possible concern is that candidates' race or gender is the basis of the appearance advantage. To rule out this possibility, we estimate the bivariate model only for races in which the two candidates shared the same race and gender (Column 2, Table 3.2). If anything, the estimated effect of appearance advantage is larger in this subset of elections. A statistically and substantively significant effect also remains after controlling for candidate gender, race, and incumbency in our full sample (Column 3, Table 3.2).[14] The

[14]Research has found differential effects of candidate appearance where one or both candidates are female (Chiao, Bowman, and Gill 2008; Poutvaara, Jordahl, and Berggren 2009). Unfortunately, we lack a sufficient number of races to shed further light on this topic (half of the races are male-male and the other half are mostly female-male races).

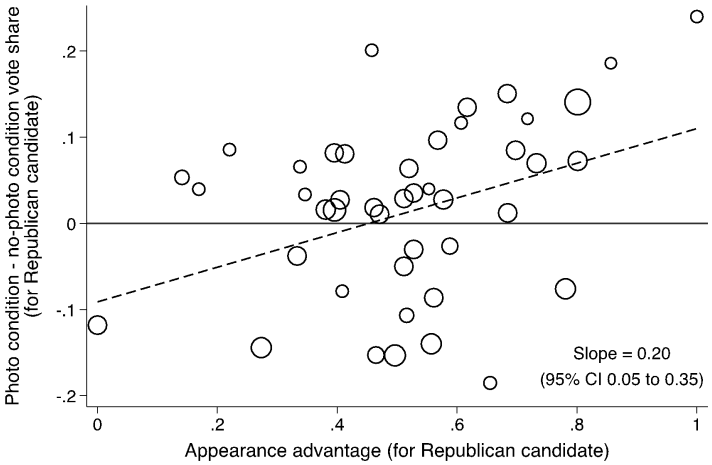


Figure 3.3: Appearance-Advantaged Candidates in Statewide General Elections Benefit in Photo Condition. Note: Each point represents a state-level general election from 2012. Observations weighted by the total number of respondents reporting a vote choice in the race.

finding also holds in U.S. Senate and gubernatorial races (Column 4, Table 3.2), which tend to be more prominent, as well as lower-ticket races (Column 5, Table 3.2). In sum, the effect of the photos is robust across a variety of specifications, and holds in lower-salience primary races and in higher-salience general election races.

While generally more professional than primary candidates, particularly unprofessional looking candidates could still drive these findings. As in the primary, however, we observe a similar effect when we exclude races with the lowest-rated candidates (Table 3.2, Column 7). These results therefore imply that candidates who look the part drive the main findings, not candidates who play the clown.

Table 3.2: Appearance-Advantaged Candidates in Statewide General Elections Benefit in Photo Condition (Dependent variable: Photo condition minus no-photo condition vote share).

	(1)	(2)	(3)	(4)	(5)	(6)		(7)
	All	Matched on race and gender	All with controls	Senate & Governor	Other	Min. Rating <= 25th percentile	Min. Rating > 25th percentile	
Appearance Advantage	0.20***	0.23*	0.15*	0.24*	0.19**	0.16		0.23*
(for Republican)	(0.07)	(0.13)	(0.08)	(0.14)	(0.08)	(0.10)		(0.13)
Incumbent			0.01					
			(0.02)					
Female Republican			0.05					
			(0.04)					
Female Democrat			0.00					
			(0.03)					
White Republican			-0.02					
			(0.10)					
White Democrat			0.02					
			(0.09)					
Constant	-0.09**	-0.11	-0.07	-0.13	-0.07	-0.09*		-0.09
	(0.04)	(0.07)	(0.13)	(0.08)	(0.04)	(0.05)		(0.07)
Candidates (N)	44	26	44	18	26	19		25
R2	0.15	0.12	0.21	0.16	0.18	0.14		0.12

Note: This table shows candidate-level regressions (each column showing a separate model) with standard errors in parentheses. The number of participants in this analysis is 1,933. Dependent variable: photo condition minus no-photo condition vote share (coded so that higher values indicate greater Republican vote share). We do not cluster the standard errors at the state level because, with only 18 clusters, clustering is unreliable (Angrist and Pischke 2009, ch. 8), but we do cluster by participant and contest in the individual-level analysis shown in Tables 3 and 4. * p<0.1, ** p<0.05, *** p<0.01

3.5 External Validity

Across a variety of elections, appearance-advantaged candidates tend to benefit when their photographs appear on the ballot, while appearance-disadvantaged candidates tend to lose support. Since the studies randomly assigned participants to the photo or no-photo condition, they demonstrate that candidate appearance can exert a causal effect on voters' choices, and are therefore inconsistent with concerns about endogeneity and the characterization of previous observational findings of appearance effects as largely or entirely spurious. Do these findings generalize to the real world? We now assess external validity.

Do Ballot Photographs Trigger Memories?

One alternative interpretation that would undermine external validity is that the photos do not directly influence votes, but indirectly do so by triggering memories of the candidates. If candidates who exert greater effort in their campaigns also look better, then the photos of those candidates could trigger positive memories about appearance-advantaged candidates and so could produce the experimental findings reported above even if voters did not actually judge candidates on their appearance. Attractive candidates tend to receive more media attention (Waismel-Manor and Tsfati 2011), which could exacerbate this potential bias. To assess this possibility, we asked participants in Study 2 factual questions specific to their states' races after the treatment. Viewing the photos failed to increase recall of relevant facts: participants assigned to the photo ballot performed 0.7% worse on these questions than participants assigned to the control ballot ($p = 0.33$)—the opposite of what we would expect if the photographs cued memories of other information. As we noted above, we also find the appearance effect among nonviable candidates and downballot candidates. Since these candidates usually lack conspicuous campaigns, these findings help further rule out this alternative.[15]

Overestimating or Underestimating the Appearance Effect? Of course, ballots in the U.S. do not show photos of the candidates (though they do in some countries, such as Brazil). These experiments, therefore, do not necessarily tell us that appearance matters in real American elections, only that it might matter. Are there reasons to think we are substantially overestimating or underestimating the "true" appearance effect? There are arguments for both positions. One obvious reason we may be overestimating the effect is priming. When the ballot shows voters the pictures, they may more often cast their vote for the appearance-advantaged candidate simply because this makes the candidate's looks more perceptually salient (i.e. at the top of voters' minds) than it would be the case when voters normally go to the polls.[16] A second reason for overestimating is that respondents

[15]Another interpretation of the finding is that candidate age—as discerned from the pictures—influences treated participants to change their votes. Previous studies, however, have found that controlling for age, using various functional forms, leaves the appearance-vote relationship unchanged (Lawson et al. 2010, 581; Todorov et al. 2005).

[16]Several classroom and lab studies have conducted experiments on appearance effects (e.g.,Johns and

misjudge their likelihood of voting, especially in the primary, and the photo ballots may disproportionately influence nonvoters.

On the other hand, we could be underestimating the effect because of noncompliance. By showing the photos to the treated group, we intended all treated participants to know how the candidates looked and all control participants not to know, which would constitute perfect compliance. In practice, however, some (even many) voters in the control group doubtless knew what the candidates looked like because the campaigns had already "treated" them. This is an instance of one-way noncompliance. Just as in a medical trial where some in the placebo condition take the real drug instead of the placebo, such noncompliance would lead us to underestimate the magnitude of the treatment effect. The presence of this kind of noncompliance should lead the experimentally-induced appearance effect to diminish as Election Day approaches because, as a result of campaign activity and media coverage, the awareness and hence the influence of the candidates' looks would have grown among the control group. Moreover, since general election campaigns are much more pervasive than primary campaigns, this compliance problem seems more problematic in Study 2.

Is the upward bias from priming and nonvoters greater than the downward bias from noncompliance? Because we cannot measure these directly, we cannot definitively say. Nevertheless, we can indirectly observe the effects (or lack thereof). Take noncompliance. As the campaign heats up, we would expect increasing numbers of control participants to know what candidates look like because the campaign is "treating" them. Consequently, candidate appearance should increasingly predict vote choice in the control group. Since we conducted Study 2 over 17 days, we can test this prediction. When we estimate the effect of appearance separately in the first half of the control group to take the survey and in the second half, we find support for this predicted pattern. As we show in OA section 3.1, the effect of appearance in the control group increases significantly over time. Noncompliance in the control group therefore likely biases our estimate of an appearance-effect downwards. To some extent, we can correct for this downward bias by estimating the photo-ballot's effect earlier in the campaign, when noncompliance is less of a problem (i.e., the control group is not already voting based on appearance). When we re-estimate the effect shown in column 1 of Table 3.2 in the first half of the study, the estimate rises considerably, from 0.20 to 0.32 (see OA section 3.1 and 3.2, which discuss statistical significance).

We may also be able to indirectly assess the upward bias from priming appearance. If we are finding an appearance effect because people cannot resist voting for attractive candidates when they can see their photos, appearance should predict votes consistently in the treated group over time. In fact, however, whereas the appearance-vote relationship increases over time in the control group, it appears to decrease by about half in the treatment group, though this drop is not always statistically significant (see OA section 3.2). This decrease may occur because the campaigns inform voters about other aspects of the candidates, and so voters rely less on appearance as a low-information heuristic as Election Day approaches.

Shephard 2007; Rosenberg and McCafferty 1987; Spezio et al. 2008). Our experiment builds on these by examining whether candidate appearance can influence real-world voters' decisions in actual elections.

This suggests that we are finding more than just an irresistible response to the photographs; thus, such a response may not be leading us to overestimate substantially the appearance effect in the real world. In light of these shifts in the magnitude of the treatment effect over time, we examined whether the importance of other variables, such as party or incumbency, also changed over time, but they do not.[17]

We can also indirectly assess the extent of upward bias from including nonvoters. Although we cannot measure which respondents truly did intend to vote, we know that the most politically knowledgeable voters vote at higher rates (Delli Carpini and Keeter 1996, 226-27). As we show in the next section, the photo ballot induced just as much appearance voting among high-knowledge individuals as it did among low-knowledge individuals in the primary study. In the primary study, therefore, we are not overestimating the effect by counting too many low-knowledge respondents as voters. In the general election study, the pattern is different: we primarily observe the effect among low-knowledge respondents. But the implications are less clear because low-knowledge individuals are more likely to vote in general elections.

Assessing whether the experimental effects we observe are externally valid—that is whether we are overestimating or underestimating the real-world effect of appearance on voters—is difficult. The findings in this section, however, should assuage concerns that these estimates represent a large overstatement. Noncompliance (participants already knowing what the candidates look like) probably suppresses the appearance effect estimates and "facial priming" does not seem to entirely overwhelm participants.

As we noted at the beginning of the paper, our research design cannot tell us precisely how much candidate appearance really matters to election outcomes, but that it could matter. Because we conducted these experiments with actual voters and actual candidates just prior to an election, we can be more confident that that the results from our mock ballots reflect voters' actual use of candidate appearance in real world choices. In other words, these experiments provide evidence for the existence of the effect in real world elections, but the magnitude of that effect in actual elections is less clear. Substantive Significance: Can Appearance Change Who Wins? Although our focus is on the existence of the effect, not the magnitude, what would be the consequence for outcomes if the estimated effect manifested itself in actual U.S. elections? To assess this, we consider how election outcomes would have changed in the absence of the appearance effect. We do this by removing the predicted effect of appearance from candidates' actual vote share.[18] In the primary elections (Study 1), the rankings of candidates change in four of the 14 districts (29%). In the 11 races with more than two candidates, the top two finishers change in two elections (districts 6 and 33 have different second-place finishers). In the general elections (Study 2), the winner changes in

[17]We also tested for over time patterns in Study 1. Since we conducted Study 1 over fewer days and since primary campaigns usually pale in comparison to general election campaigns, we might not expect to see such patterns, which is what we find.

[18]We estimate this by multiplying candidate appearance advantage by the appearance effect reported in Column 1 of Table 3.2, adding the constant to the outcome, and then subtracting that result from the candidates actual vote share in the 2012 election.

six of 44 races (14%). While these simple extrapolations from the experimental estimates may overstate the real world impact of candidate appearance on election outcomes, even if we conservatively assume that we are overestimating the appearance effect by twofold in the general election, we find that the winner would change in four races. Thus, to the extent that advertising in close races can even approach the salience of photographs on a ballot, we would expect a candidate's appearance to influence electoral outcomes.

These experimental effects are similar in size to Atkinson et al.'s observational effects in 1992-2006 Senate races (and these results preclude alternative explanations). They conclude, however, that the appearance effect is too small to change outcomes of any Senate races during that period, which seems inconsistent with our finding that some outcomes would change in the elections we examine. A key difference between our analysis and Atkinson et al. is that we simulate different counterfactuals. They ask what would have happened if the challenger were an average looking candidate, a counterfactual that ignores the impact of incumbent appearance. In our analysis, we simulate what would happen if both candidates had the same appearance.

To investigate the disparity between our results and Atkinson et al., we reanalyzed their data using a specification similar to that used for our experiments (see OA section 5.4 for details). In contrast with their result, we find that the appearance effect (estimated in the following subsets of Senate data) is large enough to change the winner in 9% of the Senate races between 1992 and 2006. The winner changes in 21% of races that the Cook Report deemed competitive (tossups) at least one year before the election. In races with an incumbent, the sample they focus on, 6% of races would have a different outcome, while the winner would change hands in 11% of open seat races.

Candidate Appearance as a Low-Information Heuristic

The candidate-level analyses demonstrate that appearance-advantaged candidates experience greater success when we attach their photographs to the ballot. A natural follow-up question is, "Why?" Research in psychology finds that people rely on appearance most heavily when evaluating others they know little about; that is, they use appearance as a low-information heuristic (Bar, Neta, and Linz 2006; Hassin and Trope 2000; Zebrowitz 1997). Research in political science has only begun to investigate whether the same holds for voting based on appearance, but at least one study suggests that it does. Lenz and Lawson (2012) find that voters rely most on appearance when they watch a relatively high amount of television but know little about politics (see also Brusattin 2012; Riggio and Riggio 2010). To see if this pattern holds in the present studies, we conduct individual-level analyses on the experimental data. These analyses also provide important robustness checks for the candidate-level findings reported above.

Individual-Level Analysis for State-Level General Election Races

We begin with the state-level general election contests (Study 2). If voters use appearance as a low-information heuristic, we might expect the appearance effect to vanish among those with better cues. In particular, politically knowledgeable voters may know enough about the candidates to not fall back on appearance. We might also expect it to diminish among strong partisans (individuals identifying as "strong" Democrats or Republicans). Since the ballot showed party labels and the contests pitted one Democrat against one Republican, strong partisans can rely on party rather than appearance as a cue. Of course, decades of research consistently show that voters rely heavily on partisanship in their voting decisions (Campbell et al. 1960; Schaffner and Streb 2002).

To test these predictions, we estimate these models at the individual level to increase the precision of the estimates. Our dependent variable is whether participants Vote Republican in a given race (coded Republican 1, Democrat 0). As in the race-level analysis, we measure Appearance Advantage in terms of the Republican candidate (Republican Appearance minus Democratic Appearance) rescaled to a 0-1 range. We use a linear probability model (ordinary least squares) but the results are the same with probit estimation (see OA section 4.1).[19] We cluster the standard errors at the election race and participant level (see OA section 4.1 for alternative specifications that yield similar findings). We measure general political Knowledge with a four-item scale and classify as highly knowledgeable participants who answered three or more items correctly (see OA section 4.2 for wording).

Table 3.3 shows the results. In column 1, we regress Republican vote choice on an indicator for the photo ballot (Treatment), Appearance advantage (for Republican), and the interaction between the treatment and appearance advantage (Treatment x Appearance Advantage [for Republican]). This interaction is the coefficient of interest—it tests whether candidate appearance predicts vote choices better in the treatment group than in the control group. Consistent with the candidate-level findings above, the interaction is positive and statistically significant. Its size, 0.20, implies that participants in the photo condition are 20-percentage points more likely to vote for the most appearance-advantaged candidates compared to the least advantaged one (95% CI: [0.06, 0.34]).

The next four columns of Table 3.3 test the low-information heuristic predictions. As expected, candidate appearance has a significantly higher effect for low-knowledge voters (Column 3) than high-knowledge voters, for whom the observed coefficient is near zero (Column 4). Similarly, candidate appearance affects weak partisans and independents (Column 5) more than strong partisans (Column 6), though this difference (0.22 versus 0.18) is not statistically significant. Because high-knowledge voters may also be strong partisans, Table 3.4 examines the impact of candidate appearance on vote choice for subsets of participants based on both variables. It shows that knowledge, not partisanship, appears most protective. Among low-knowledge voters, we find that candidate appearance has a large and significant effect on weak partisans and independents (Column 1) and strong partisans (Column 2).

[19] We use linear probability models because they are consistent under weak assumptions and the estimates are simpler to interpret, especially with interaction terms (Ai and Norton 2003).

Among high-knowledge voters, we find no appearance effect among either group (Columns 3 and 4). Taken together, these findings support the low-information heuristic interpretation of appearance effects. By implication, if voters were better informed about politics, they would not rely on candidate appearance.[20]

[20]In OA sections 4.3-4.5, we find evidence that strong partisanship can diminish the appearance effect, especially among high-knowledge individuals. We find this in downballot races (no senatorial and gubernatorial races) and when we substitute local for general knowledge.

Table 3.3: Voters Favor Appearance-Advantaged Candidates at Higher Rates on the Photo Ballot (Dependent variable: Vote Republican indicator variable).

	(1)	(2)	(3)	(4)	(5)	(6)
	All participants	Matched on race and gender	Low knowledge	High knowledge	Weak/ indep.	Strong partisan
Treatment	-0.09*	-0.11	-0.20***	0.01	-0.14**	-0.05
	(0.05)	(0.07)	(0.07)	(0.06)	(0.06)	(0.07)
Appearance advantage	-0.03	0.03	-0.06	-0.01	-0.05	-0.07
(for Republican)	(0.06)	(0.14)	(0.08)	(0.09)	(0.09)	(0.08)
Treatment x Appearance advantage	0.20***	0.24**	0.39***	0.01	0.22***	0.18*
(for Republican)	(0.07)	(0.11)	(0.11)	(0.08)	(0.08)	(0.10)
Votes (N)	4,816	2,918	2,324	2,492	2,826	1,626
R2	0.00	0.00	0.01	0.00	0.00	0.00

Note: This table shows individual-level regressions (each column showing a separate model). The dependent variable is coded Republican vote 1 and Democratic vote 0. Constant not shown. Standard errors clustered at the individual and race-level in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 3.4: Low-Information Voters are Most Susceptible to Candidate Appearance (Dependent variable: Vote Republican indicator variable).

	(1)	(2)	(3)	(4)
	Low knowledge & Non-strong partisan	Low knowledge & Strong partisan	High knowledge & Non-strong partisan	High knowledge & Strong partisan
Treatment	-0.18**	-0.24**	-0.10	0.13
	(0.07)	(0.10)	(0.09)	(0.10)
Appearance advantage	-0.10	-0.05	0.00	-0.08
(for Republican)	(0.11)	(0.13)	(0.13)	(0.08)
Treatment x Appearance advantage	0.34***	0.47***	0.09	-0.11
(for Republican)	(0.11)	(0.18)	(0.13)	(0.09)
Votes (N)	1,475	694	1,351	932
R2	0.01	0.03	0.00	0.01

Note: This table shows individual-level regressions (each column showing a separate model). The dependent variable is coded Republican vote 1 and Democratic vote 0. Constant not shown. Standard errors clustered at the individual and race-level in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Individual-Level Analysis for U.S. House Primaries in California

In comparing general and primary elections, there are reasons to expect that partisan and high-knowledge voters would be more likely to rely on candidates' appearance in the primary context. While party labels clearly provide important cues in most general election races, they are significantly less valuable in primary contests, which usually feature multiple candidates from the same party. (Ten of the 14 races included in Study 1 did so.) Furthermore, voters know much less about congressional candidates than gubernatorial and senatorial candidates like those included in our general election study (Krasno 1997). In fact, knowledge is so low in congressional primaries that even politically knowledgeable voters appear largely ignorant of candidates' policy positions, except for what they can glean from the candidate's party (Ahler, Citrin, and Lenz Forthcoming). Consequently, all voters in primaries, including politically knowledgeable strong partisans, may be more likely to rely on candidate appearance in their voting decisions because they know so little about the candidates and cannot rely on party labels as a guide.

Consistent with this reasoning, voters in Study 1 appear about equally susceptible to candidate appearance, regardless of their partisanship and knowledge: both high-knowledge and low-knowledge voters who saw the ballot with photos voted for appearance-advantaged candidates more frequently. Similarly, we find that strong partisans (those who place themselves at 1 or 7 on the 7-point party ID scale) vote for appearance-advantaged candidates at a similar rate to independents and weak partisans when they see photos of the candidates. We present these findings in the OA (see section 4.4). Across both studies, therefore, appearance matters for voters' choices. But voters also seem to rely on appearance less when they know more about politics. The apparent moderating effects of knowledge and (less consistently) partisanship in the higher-salience general election contests, combined with a more wide-ranging effect of candidate appearance in House primaries, supports the view that candidate appearance acts as a low-information heuristic that voters discard when they have are more informed or can use more reliable cues. If voters knew more about the candidates, we might not find that candidate appearance—a piece of information that may carry little signal—influences vote choices. As it is, however, we find that looks sometimes may matter enough to affect electoral outcomes.

3.6 Conclusion

This study took earlier research finding a sizeable correlation between candidates' appearances and their electoral fortunes as a starting-point. Given that this result understandably evokes familiar normative concerns about citizen competence and pessimism about democratic accountability, scholars rightly raised questions about the authenticity of the appearance-choice connection, arguing that omitted variables may have generated a relationship that in reality is spurious. Given that candidate appearance correlates with so many other variables—candidate spending, incumbency, and incumbent vulnerability—it

seemed plausible that candidate appearance was not a direct cause of voters' behavior.

To overcome the indeterminacy about the meaning of these observational studies, we designed two experimental studies. These randomized experiments supported the notion that candidate appearance can affect voters' stated choices. When we exposed voters (in a treatment group) to photos of the candidates not long before Election Day, they reported intending to vote for appearance-advantaged candidates at higher rates and appearance-disadvantaged candidates at lower rates more often than those (in a control group) shown a ballot without photographs of the candidates. Since candidate effort (or other omitted variables) could not differentially influence voters in the photo conditions compared to voters in the control conditions, we no longer have to worry about the plausible alternative interpretations. By introducing exogeneity into a morass of endogenous relationships, we determine that candidate appearance does seem to have a direct, causal influence on voters.

The effect of candidate appearance on voting in these experiments is robust. It holds up in primary and general election races, among incumbents and challengers, among viable and nonviable candidates, among Democratic and Republican candidates, among up-ballot races (senator and governor) and down-ballot races (e.g., attorney general), and in contests between candidates matched on race and gender. It also holds up in candidate-level analyses and in individual-level analyses. These effects show up even though voters are casting ballots for real-world candidates in their districts not long before Election Day and even though the ballots provide other information, such as candidate party, incumbency status, and occupation. While our experiments cannot precisely estimate the real world effect, they demonstrate that candidate appearance can influence voters. Even under the conservative assumption that the true effect is half the experimental estimate, candidate appearance changes election outcomes. We also presented evidence that the effect is unlikely to be due to the photo condition triggering memories of candidates or from an irresistible tendency of participants to vote based on candidate looks when the ballot shows photos (priming). Finally, we show the findings with a Mechanical Turk sample and a demographically representative sample of registered voters in California (via SSI).

Some scholars argue that voters generally possess the cues necessary to reach "as-though informed" decisions (Lupia 1994; Lupia and McCubbins 1998; Popkin 1991). Our results present evidence that this is not always the case: candidate appearance, an arguably uninformative cue, can influence voters' choices and electoral outcomes. Since merely showing photographs of candidates' faces produces changes in voting intentions, our findings raise questions about the quality of some voter heuristics and, more broadly, about citizens' ability to hold politicians democratically accountable.

This, of course, all assumes that appearance is not an informative heuristic, that is, that appearance is uncorrelated with traits voters intentionally desire in the candidates. So we should ask, do looks provide valid information about candidates' abilities? Although some studies find small correlations between attractiveness and IQ scores (e.g., Zebrowitz et al. 2002), most researchers conclude that the inferences about competence people draw from faces fail to correspond with reality (Alley 1988; Cohen 1973; Hassin and Trope 2000; Kalick

et al. 1998), though none investigate candidates for political office.[21]

Researchers in other fields have found that people routinely make costly decisions based on facial inferences even when those inferences are shown to be uninformative, such as when lending money online (Ravina 2012), eliminating competitors on a television game show (Belot, Bhaskar, and Van De Ven 2012), playing incentivized trust games (Wilson and Eckel 2006), and incentivized public goods games (Andreoni and Petrie 2008).[22] Research on the beauty wage premium in the labor market reaches similar conclusions (e.g., Fletcher 2009). Summarizing years of his own and others' research on the beauty premium, Daniel Hamermesh (2011) concludes that it results from taste-based discrimination, which he calls lookism, writing, "We have met the enemy and he is us" (Hamermesh 2011, 122). So, voters may be doing the same here: voting on candidate appearance even though candidate appearance is not informative about competence or other relevant traits. Consistent with this interpretation, voters say they place little weight on appearance in their voting decisions (see OA section 5.1). Of course, research on candidate appearance is still in its early stages and our design cannot demonstrate why candidate appearance matters, only that it does.

The notion that candidates' looks should not matter pervades popular conversation about politics. From Bill Clinton's 2012 stump speech charging that Mitt Romney predicated his campaign on "looking like a president" (Nelson 2012) to the 2013 kerfuffle over President Obama's comment that California has the "best looking attorney general," opinion leaders contend that we shouldn't judge our politicians based on how they look. According to our findings, however, this normative ideal fails to describe voters' behavior, especially among those with little other information to go by. More optimistically, our findings also point to a remedy. In the general election, political knowledge—and, to a lesser degree, partisan attachment—protected voters from this superficial tendency to use looks as a cue (though in primaries, where information is generally scarce, they did not). Opinion leaders who truly believe that appearance should not matter can potentially alleviate this tendency through campaigns that inform voters about candidate characteristics more relevant to effective governance. Appearance will likely cease to matter only when most voters possess more substantive guides to their choices.

[21] Unpublished work by one of the authors finds that competent looking incumbents are no more effective in Congress, nor are they evaluated as being more effective by peers in the North Carolina legislature (see OA section 6).

[22] Other examples include students evaluating their professors teaching (Hamermesh and Parker 2005) and economists electing officers to the American Economic Association (Hamermesh 2006).

Chapter 4

Short Story or Campaign Narrative? The News Cycles of the 2016 U.S. Presidential Election

If the previous chapter examines what happens when voters lack relevant information about political candidates in congressional primaries and statewide elections, this chapter looks at where the media does point its spotlight in the 2016 presidential election. Unlike House elections and other down ballot races, voters should have no trouble acquiring information about presidential candidates if they want to, but different patterns of coverage may yield different results in practice for inattentive voters. In particular, I examine the timing of coverage of negative stories because prior research suggests that both repetition and recency are important to effective political communication (Gerber et al. 2011; Hill et al. 2013). Repetition and recency are particularly important if we consider the voters in the prior chapter who are paying so little attention to politics that they cast their ballot based on candidate appearance.

4.1 Paper Abstract

When major events occurred in the 2016 U.S. presidential campaign, which saw a brief spike in coverage and which became a more permanent feature of campaign coverage? In this paper I examine six major events in the presidential campaign to test the hypothesis that news outlets of all persuasions will cover major events as news, but only partisan outlets will continue to discuss negative stories about their opponents long after the event that made the topic news. Broadly, I find that all outlets do indeed pick up major stories temporarily, but that the more tradition news news organization in my study does not stick with a higher level of coverage of any topic after a seven day window following the related event. Partisan outlets, in contrast, either continue to cover negative stories about the opposing candidate at a higher rate or were already on the story before a related event caused everyone else to

temporarily pick up the story.

4.2 Introduction

When a major event happens that has the potential to define a presidential campaign, news outlets can either ignore the event, pick up the story temporarily, or make the topic a more permanent feature of their campaign coverage. Campaigns would like voters to forget negative stories about their own candidate and receive more negative stories about their opponent, but will news outlets oblige? I hypothesize that all news outlets will cover a topic as campaign news right after a major event occurs, but that only partisan outlets will both pick up negative stories about their opponent and continue to prominently feature that topic throughout the remainder of the campaign. To test this theory, I look at front page news coverage of Hillary Clinton and Donald Trump on dailykos.com, breitbart.com, and nytimes.com from just prior to the Republican Convention through to the day before Election Day. I analyze the impact of six different events on which topics are discussed in the news. In particular, I look at Hillary Clinton's health, Clinton's email server, Clinton's "basket of deplorables" phrase, Trump's tax returns, Trump's exchange with the Khan family, and Trump's remarks on sexually assaulting women.

To briefly outline the paper, next I discuss a conceptual framework for understanding which topics news outlets will cover. Then, I briefly summarize the events that are likely to catapult these stories into the news. After that, I discuss the data, how topics are measured, and how to use permutation tests to detect changes in coverage. Finally, I present results for each outlet in my study showing that broadly speaking all outlets cover these topics in the seven days following the major event. The more traditional news organization in my study does not make any of these topics a more permanent feature of the campaign, while the more partisan outlets behave in a more partisan manner.

4.3 Receive Accept Sample (RAS): A Conceptual Framework For Thinking About Changing the Topic

Each of the topics in this study featured a major event that could catapult the story into the national news and had the potential to dominate coverage of the 2016 presidential campaign. How should we think about which topics media outlets choose to write about and which messages voters adopt? One way to conceptualize the importance of shifts in the topic of coverage is Zaller's (1992) Receive Accept Sample (RAS) framework of opinion formation. The first step in opinion formation for a voter is receiving a message, or what Zaller calls a consideration. To believe a message a voter first has to receive it. If a voter receives a message, they can either accept or reject it. Finally, having recently received and accepted

a message, the idea is a more salient consideration when the voter samples considerations to offer an opinion. In this framework then, it is extremely important to understand which considerations or stories voters receive from the media, which they accept, and which remain a salient consideration when voting for president. Next, I briefly discuss how voters process political communications and then turn to how the topics discussed in presidential election coverage fit into this framework.

The literature on media persuasion suggests that political information may have very short term effects. For example, Gerber et al. (2011) find in a randomized field experiment that television campaign ads have no impact beyond the week after they are aired. In observational work, Hill et al. (2013) also find effects "decay very rapidly in the first few days, as would be expected if most voters were making memory-based evaluations that are not stored in long-term memory." They find, however, both smaller effects and a slower rate of decay for presidential races consistent with voters caring more about presidential elections, being less easily persuadable than in down ballot races, but also the potential for messages throughout the campaign to have an impact. They stress, however, that "sizeable long-term effects occur only when communication is repeated over a long period of time" (Hill et al. 2013, 543). Thus, coverage may matter most at the very end of the campaign, but both spikes in coverage earlier in the campaign and whether or not a topic is repeated throughout the campaign could impact a close election.

When considering which topics news outlets will cover, one extension to the RAS model would be to think of topics as having different distributions of ideological considerations such that not only would political parties or partisan news outlets like to influence whether the considerations voters receive are liberal or conservative, but they would also like to increase the salience of issues on which voters are more likely to have and receive considerations that favor their particular ideological or partisan persuasion. In this vein, Groseclose (2011) argues that while there is no such thing as liberal facts and conservative facts, there are facts that would lead a reasonable, but undecided person in a particular direction. So while a left wing outlet may have a particular spin on a story, in general Democrats would rather independents see more stories about Donald Trump discussing sexually assaulting women and fewer stories on Clinton denying sending classified material through her private email server. Thus, my research design in this paper does not attempt to measure whether a particular story is positive or negative towards a candidate, but instead assumes that the stories I have selected are generally bad news for the candidate involved.

To believe a message a voter first has to receive it. Left wing websites should then be more more likely to discuss negative stories about Trump that would lead voters to vote for Clinton. In contrast, right wing websites should prefer to cover negative stories about Clinton. No matter the ideological slant of a news outlet, however, I expect them to cover a major news story when it first breaks. I expect, however, that beyond the immediate aftermath of a story partisan news outlets will continue to cover negative stories about the opposing candidate, but not negative stories about their own preferred candidate. In contrast, I expect more traditional journalistic outlets will always chase the latest story. Thus, I hypothesize that:

- H1: All publications will cover a topic more in the 7 day following a major event related to that topic.
- H2: Partisan publications that support one candidate more than the other will continue to cover a negative topic about the opposing candidate more throughout the remainder of the campaign than they did prior to a major related event. Traditional journalistic news organizations will not.

4.4 The Top Stories of the 2016 Election

Each of these topics was selected because it was a major story of the 2016 campaign that (a) revolved around a particular event and (b) had the potential to become the primary story of the campaign following that event. Here is a list of the events in chronological order:

- Trump and the Khans (7/30/2016): Khizr Khan had given a speech at the DNC Convention attacking Trump, two days later Trump suggested that Khan's wife stood behind him during the speech because she was not allowed to speak.
- Clinton's Basket of Deplorables (9/9/2016): Clinton says 1/2 of Trump supporters are in a "basket of deplorables."
- Clinton's Health (9/11/2016): Clinton leaves 9/11 Memorial Ceremony early with help walking and walking pneumonia.
- Trump's taxes (10/1/2016): The New York Times reports on a leaked 1995 Trump tax return that suggests he may have paid no income tax for up to 18 years.
- Trump and sexual assault (10/7/2016): Access Hollywood/Billy Bush tape released. In the tape Trump says he can "grab 'em by the pussy" without waiting for permission.
- Clinton's emails (10/28/2016): FBI Director Comey announces he is reopening investigation into whether Clinton mishandled classified information after Clinton emails found on Anthony Weiner's laptop.

To briefly summarize these events, for Clinton the three major stories I look at are her health, her remarks labeling Trump supporters a basket of deplorables, and the reopening of the investigation into her emails by FBI Director Comey. Hillary Clinton needed help to leave the 9/11 memorial ceremony early due to pneumonia, raising questions about whether she was healthy enough to be president.[1] Second, when Clinton gave a speech at an LGBT fundraiser, she described half of Trump supporters as a "basket of deplorables." She later apologized for saying half of his supporters fell in this category, but stuck by the remark which

[1] https://www.nytimes.com/2016/09/12/us/politics/hillary-clinton-campaign-pneumonia.html

became a meme in the campaign.[2] Finally, Clinton's emails were in the news throughout the campaign because of her unauthorized use of a private email server while Secretary of State, but they were thrust back into the spotlight when FBI Director James Comey announced that he was reopening the investigation into Clinton shortly before Election Day because additional Clinton emails had been found on the computer of Anthony Weiner, husband of Clinton aide Huma Abedin.[3] The FBI reclosed the investigation a week later, announcing that they had found nothing new two days before Election Day.

For Trump, the three major stories I look at are his disparaging remarks about the Khans, his failure to disclose his tax returns, and his discussion of sexual assault with Billy Bush. Khizr Kahn, the father of an American soldier killed in combat in Iraq, gave a speech at the Democratic National Convention in which he said to Trump "You have sacrificed nothing and no one." In an interview two days later, Trump said that he had "made a lot of sacrifices" and suggested that Khan's wife Ghazala Khan stood behind him during the speech because she was not allowed to speak.[4] Second, I look at Trump's refusal to release his tax returns, which was a story throughout the campaign, but it became particularly salient when the New York Times released a story suggesting that based off of a $916 million dollar loss declared in a leaked copy of his 1995 returns he may have paid no income tax whatsoever for nearly two decades.[5] Finally, allegations of sexual harassment and sexual assault against Donald Trump trickled in throughout the campaign, but the release of the Access Hollywood/Billy Bush tape by the Washington Post was an explosive news story. In addition to lewdly discussing trying to sleep with a married woman while he himself was married to his current wife, Trump describes forcing himself on women: "You know, I'm automatically attracted to beautiful — I just start kissing them. It's like a magnet. Just kiss. I don't even wait. And when you're a star, they let you do it. You can do anything... Grab 'em by the pussy. You can do anything."[6]

Each of these topics had the potential to become a permanent feature of coverage that defined the 2016 presidential campaign. Which events caused a temporary spike in coverage and which led a more permanent shift in the media's narrative of the 2016 campaign? In the next section I describe which websites I analyze, how I measure these topics, and how I estimate whether an event changed coverage.

4.5 Data and Methods

In this paper, the dataset used includes webpages on dailykos.com, breitbart.com, and nytimes.com. In total, I have 11,308 articles that mention either "Clinton" or "Trump" at

[2] https://www.nytimes.com/2016/09/12/us/politics/hillary-clinton-campaign-pneumonia.html

[3] https://www.nytimes.com/2016/10/29/us/politics/fbi-hillary-clinton-email.html

[4] http://abcnews.go.com/Politics/donald-trump-father-fallen-soldier-ive-made-lot/story?id=41015051

[5] https://www.nytimes.com/2016/10/02/us/politics/donald-trump-taxes.html

[6] https://www.nytimes.com/2016/10/08/us/donald-trump-tape-transcript.html

least three times. This includes 2,160 pages from Daily Kos, 2,563 pages from the New York Times, and 6,585 pages from Breitbart. Daily Kos is a left wing site that features both paid writers and user generated blog posts. Breitbart is a far right news organization that was previously headed by Trump's campaign manager Steve Bannon. All of these pages were linked to from the homepage of their respective site and the time the link first appeared on the homepage is used to timestamp a story. Within the coverage of either Clinton or Trump, topics are defined by search terms that occur alongside a candidate. As Table 4.1 shows, each topic was defined by having the relevant candidate appear three or more times as well as at least one occurrence of a relevant word or phrase that defined the topic. I chose to define topics in this manner so that I would know exactly what features characterize a topic. Because I scraped all visible text and use counts of word terms for all visible text in paragraphs that include one of the candidates' names, the topic definitions do not guarantee that a page is on that particular topic because the topic could simply be featured on the page as part of the text to a link to or summary of another story. Thus, I measure a combination of direct coverage and indirect exposure to a topic on these websites.

Table 4.1: Topic Definitions

Topic	Candidate Count	Topic Count
Trump and the Khans	Trump >= 3	khan >= 1
Clinton's Basket of Deplorables	Clinton >= 3	basket of deplorables >= 1
Clinton's Health	Clinton >= 3	clinton health >= 1
Trump's taxes	Trump >= 3	trump taxes >= 1
Trump and sexual assault	Trump >= 3	sexual assault >= 1
Clinton's Emails	Clinton >= 3	clinton emails >= 1

Fisher Permutation Test

While a topic may be present more after an event, it is difficult to know how much topics vary naturally. For example, Trump's taxes were discussed throughout the campaign. Without knowing how much the amount of coverage of this topic varied over the course of the campaign it is difficult to know how much of an uptick in coverage is a significant spike that can be attribute to one event like the leak of documents from Donald Trump's 1995 tax returns to the New York Times. One approach that helps address this problem is to use a permutation test. Under the sharp null, the true effect of the event on the presence of the topic is zero for all articles. By permuting, or randomly drawing an event date and time of day (hereafter datetime), we can calculate the distribution of estimates under the sharp null and then compare the actual estimate to that distribution to learn the probability of observing an outcome as extreme or more extreme than the outcome that was observed given that the sharp null hypothesis is true.

In an experiment with a binary treatment, where permutation inference is most commonly used, we would calculate the null distribution by rerandomizing treatment assignment for each observation because this is the level at which treatment assignment occurs (Imbens and Rubin 2015). For this analysis, treatment assignment occurs based on the datetime that an event occurs. To estimate the actual effect of an event on coverage, I regress whether the topic is present on an indicator for whether an article occurred in the 7 day window following the event and an indicator for whether the story occurred over 7 days after the event. To obtain the null distribution I then randomly draw an event datetime $n = 1,000$ times, assign these indicators based on the datetime drawn, and estimate the same regression in order to obtain a null distribution for each parameter. I then compare the actual result to the distribution of results with randomly permuted datetimes.[7] Each figure in the results section displays the null distribution in blue with the estimated effect in the actual data marked with a vertical red line.

4.6 Results

Broadly, I find that all publications covered these topics significantly more in the seven days following the breaking news event. In the full sample, Clinton's emails, Clinton's health, Clinton's basket of deplorables remark, the Trump/Khan exchange, and Trump's taxes were all covered at significantly higher rates in the seven day window following the relevant event ($\alpha = 0.05$ or lower). While Donald Trump and sexual assault were covered more in the seven day window after the Access Hollywood tape leaked, it was the only topic that was not covered at a significantly higher rate in the seven day window. This topic, however, was significantly more likely to be covered after the seven day window. One possible explanation for this result is that the topic was defined based on the discussing both Donald Trump and sexual assault. Defining the topic as discussing Trump and Billy Bush or Access Hollywood may have detected a larger immediate spike in coverage while discussing the event as sexual assault as the topic was measured may have taken longer to occur. Trump/Khan was the only other topic that was present significantly more after the seven day window than before the event ($\alpha = 0.1$). As we will see next when I break out my results by publication, this result is driven by Daily Kos' coverage. Thus, the media broadly picked up every major story and covered the related topic immediately following the event, but four of the six topics did not became a more permanent feature of campaign coverage. Next, I break out my results

[7]Datetimes are randomly drawn from the range between the start of the Republican Convention to the day of Comey's announcement that he was reopening the investigation into Hillary Clinton. Articles from a few days before the Republican Convention to the day before the election are included in the dataset. Including coverage prior to the start date and always having a few days of coverage after the seven day window ensures that the regression model will always be identifiable. Similarly, where my coverage of a site starts after the Republican Convention, I start the permutation two days after the first article in the dataset. Again, articles that occur before the event and after the seven day window are required to identify the model. Finally, I exclude from the permutation datetimes where the randomly drawn seven day window would overlap with the actual seven day window.

Figure 4.1: Overall Coverage of Clinton Emails

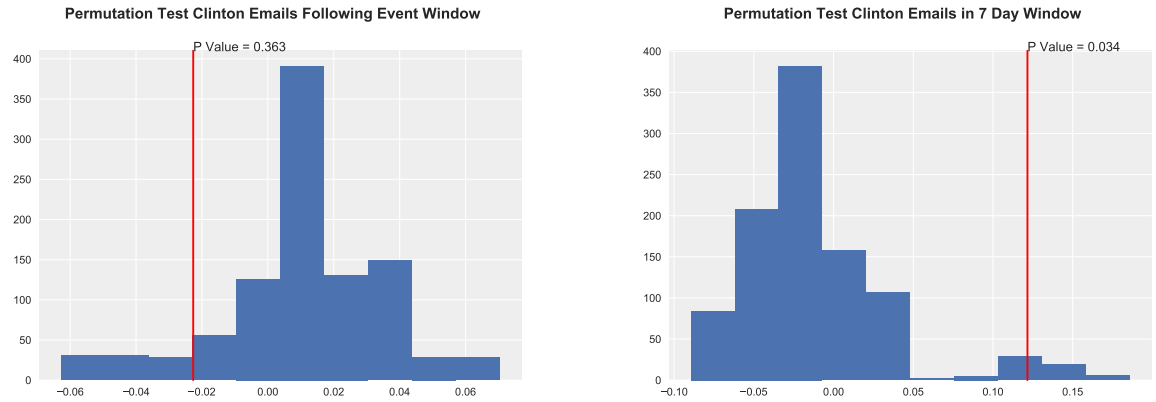
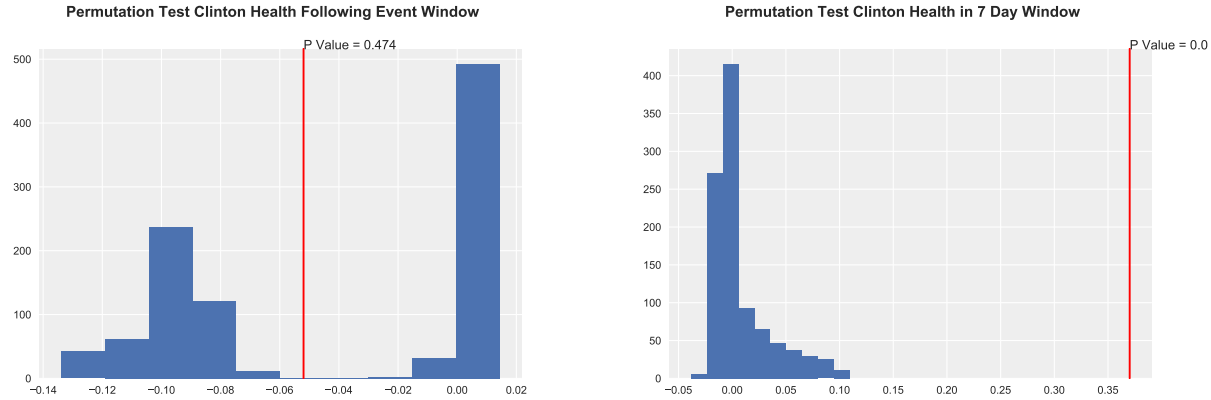


Figure 4.2: Overall Coverage of Clinton Health



by publication to test the hypothesis that partisan outlets will behave differently than more traditional media.

Figure 4.3: Overall Coverage of Clinton Basket of Deplorables

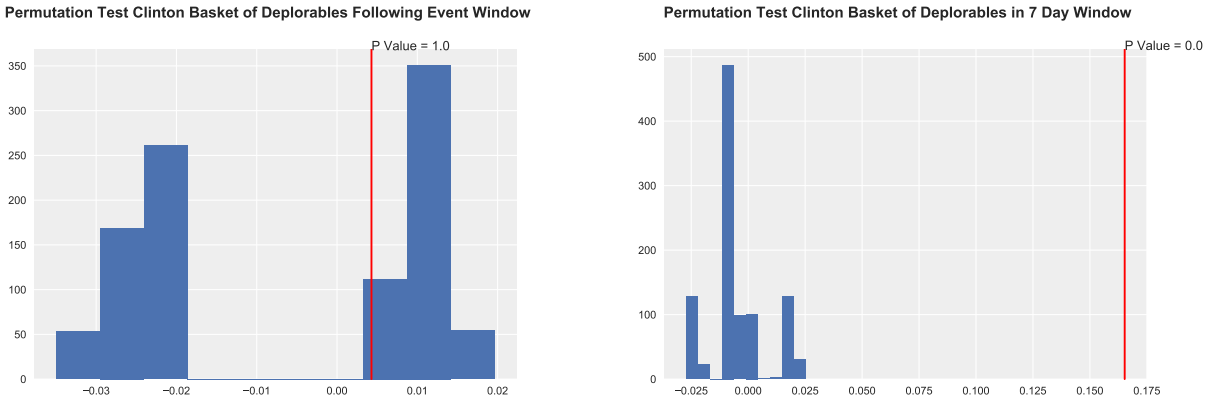


Figure 4.4: Overall Coverage of Trump/Khan

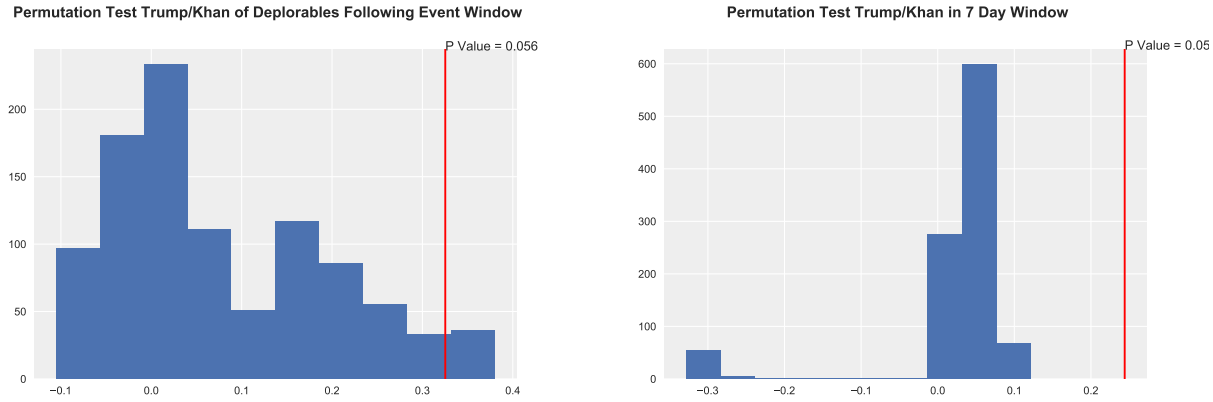


Figure 4.5: Overall Coverage of Trump's Taxes

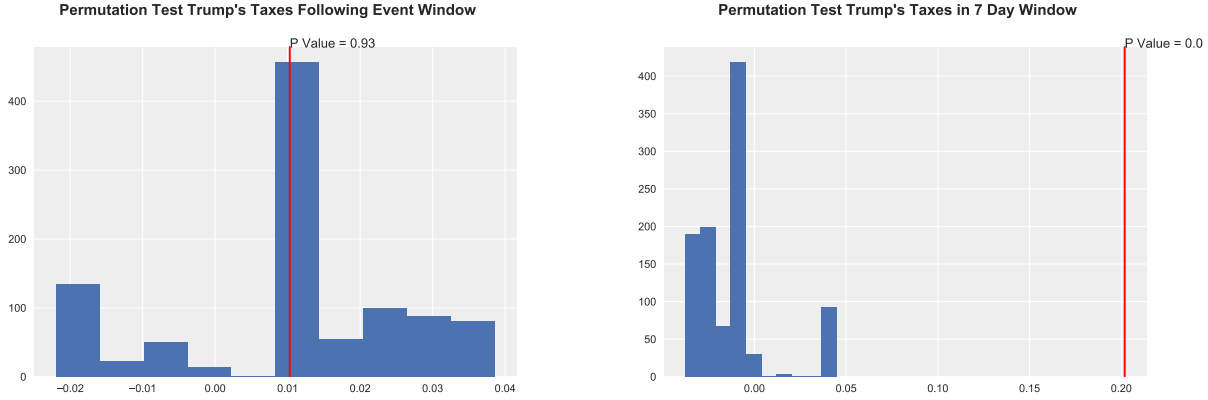
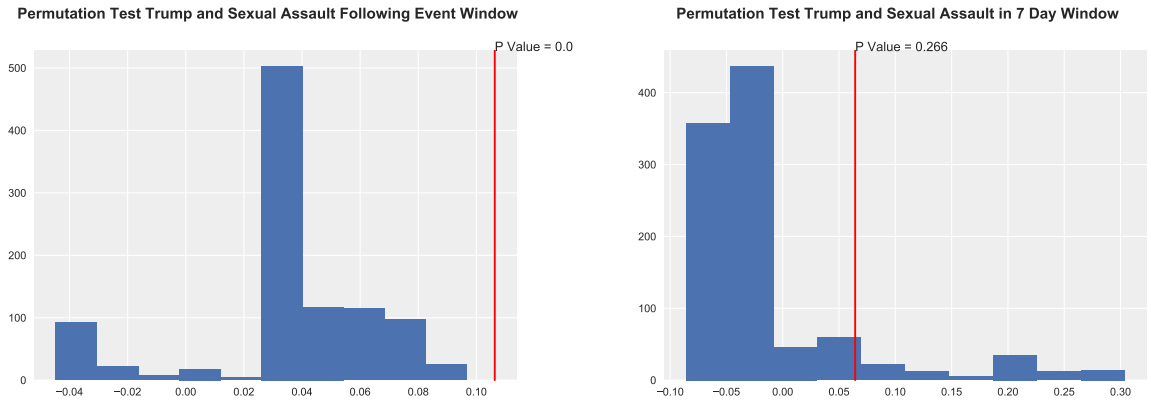


Figure 4.6: Overall Coverage of Trump and Sexual Assault



Daily Kos

I hypothesized that partisan websites would cover negative stories about their candidate as news, but would not continue to cover these topics such that we should see significantly more coverage of negative Clinton stories in the 7 day window following the story, but not after. For Daily Kos, this is the pattern of coverage I find for Hillary Clinton. They cover Clinton's emails, Clinton's health, and Clinton's basket of deplorables remarks at significantly higher rates in the seven day window following each respective event ($\alpha = 0.1$ or lower). After the seven day window, however, Daily Kos does not spend significantly more time on these negative Clinton topics than they did prior to the event.

For the negative stories about Trump, I expected Daily Kos to both cover the stories more in the seven day window after the major event and to continue covering these negative stories about Trump more than before the event. I find that for both the Trump/Khan exchange and Trump's remarks describing sexual assault, Daily Kos covers these stories significantly more after the seven day window than before the event. They also cover Trump/sexual assault and Trump's taxes significantly more in the seven day window. They do not cover Trump's taxes more after the seven day window, and while we see a substantively large increase in coverage of the Khan exchange during the seven day window, the effect is not significant. This latter result is likely because Khan becomes a near permanent feature of Daily Kos that almost always shows up on the webpage in the same place as Trump even if it is not the focus of a post. Overall then, Daily Kos picks up every Clinton story briefly, but does not make the topics more prominent features of their coverage over the remainder of the campaign. In contrast, both the Khan exchange and Trump's remarks describing sexual assaulting women become more permanent features of how Daily Kos covers the campaign.

Figure 4.7: Daily Kos Coverage of Clinton Emails

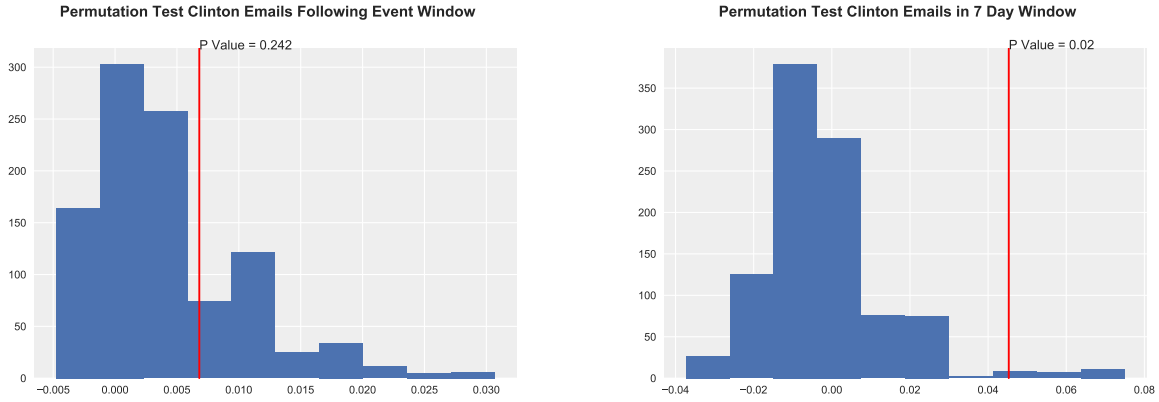


Figure 4.8: Daily Kos Coverage of Clinton Health

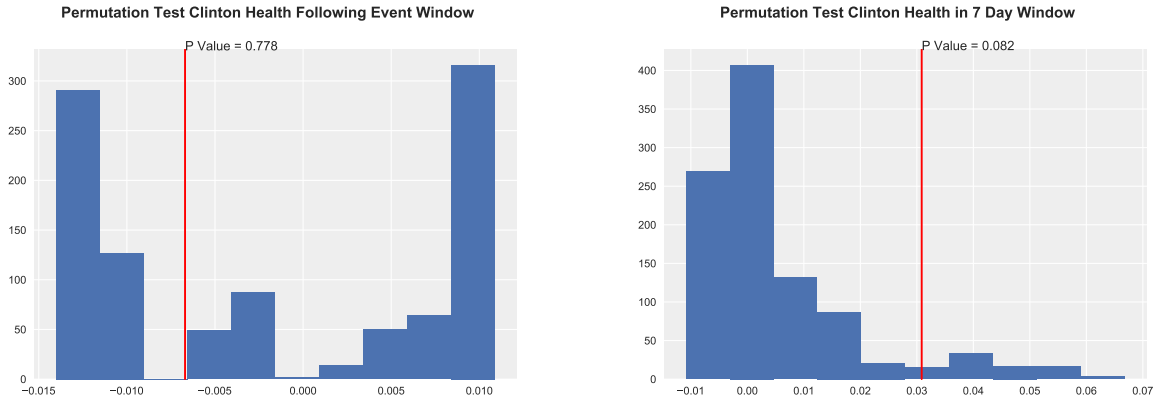


Figure 4.9: Daily Kos Coverage of Clinton Basket of Deplorables

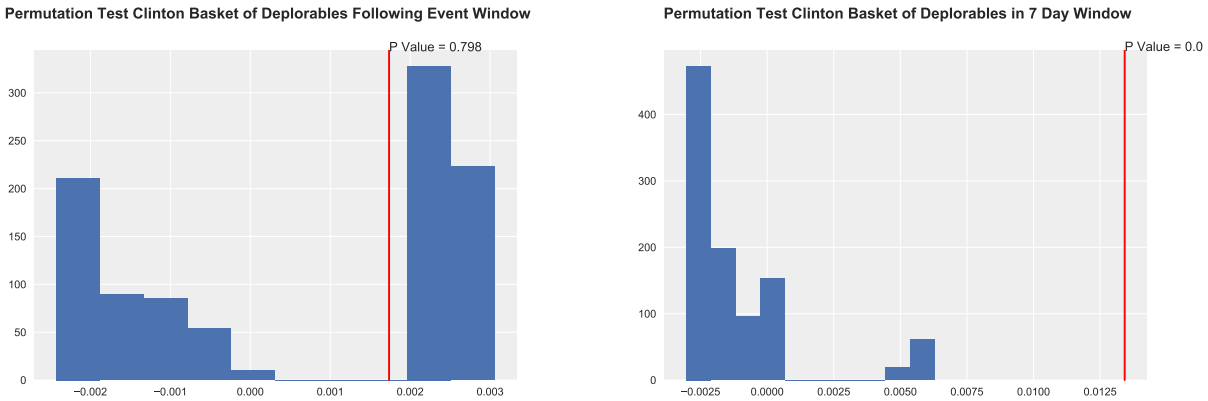


Figure 4.10: Daily Kos Coverage of Trump/Khan

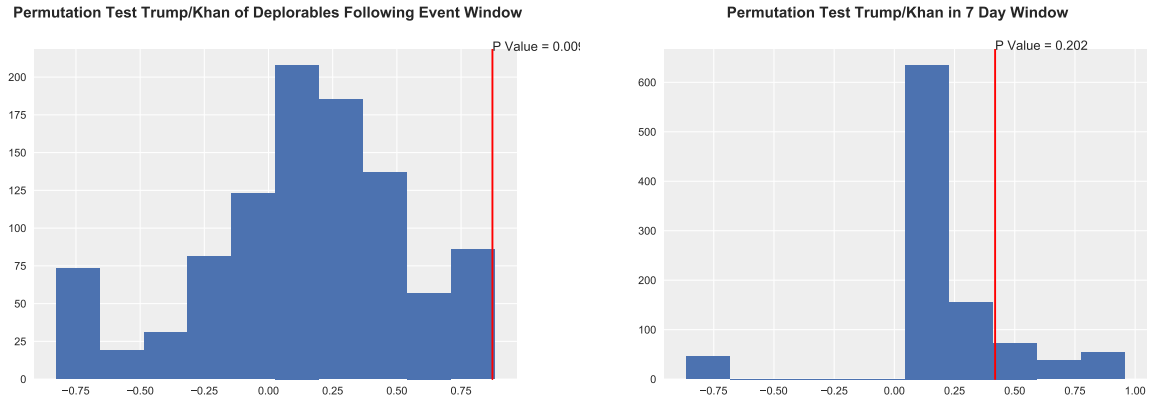


Figure 4.11: Daily Kos Coverage of Trump's Taxes

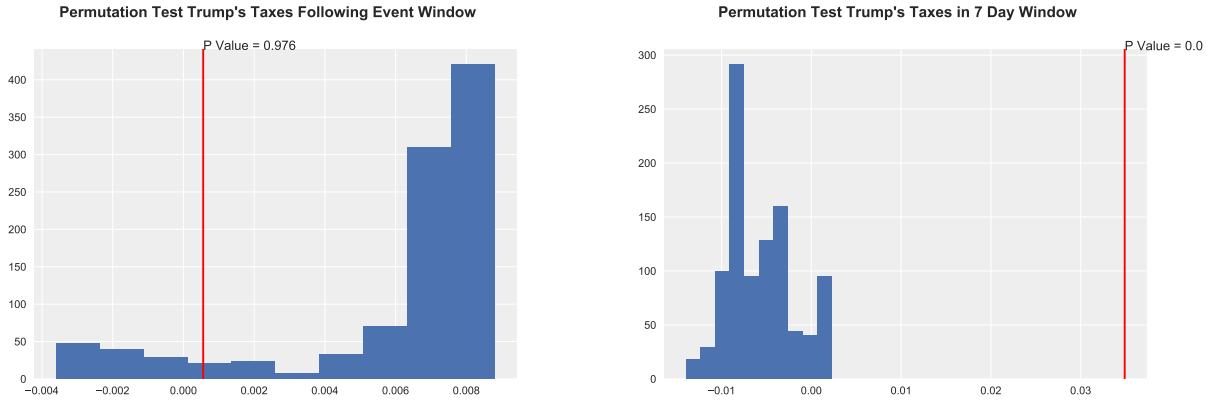
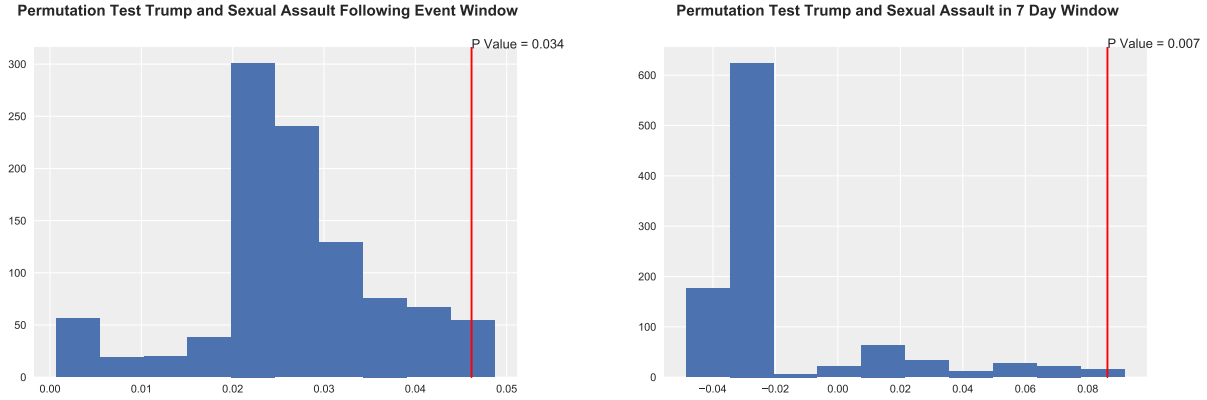


Figure 4.12: Daily Kos Coverage of Trump and Sexual Assault



Breitbart

Unfortunately, I did not begin scraping Breitbart until 2016-8-28, after Trump's exchange with the Khans, so I do not include that event in my analysis of Breitbart. Breitbart covers every other topic except Trump's remarks on sexual assault at a significantly higher rate during the seven day window following a major related event. This is broadly consistent with my hypothesis that partisan news outlets will briefly pick up every major campaign story, even negative ones about their candidate. Surprisingly, Breitbart covers Trump and sexual assault at a significantly higher rate after the seven day window than before. On closer inspection, however, Breitbart is both discussing accusations of sexual assault against Bill Clinton and criticizing the media for treating Trump unfairly on the topic because, as Trump says, "they want to stop our movement."[8] When it comes to Trump, Breitbart is thus either following the predicted partisan pattern or counter-attacking on the topic.

In contrast, however, I do not find evidence that Breitbart continues to cover negative stories about Clinton at higher rates than before an event. Perhaps it is unsurprising that Breitbart does not use the phrase "basket of deplorables" to discuss Clinton's remarks because she was pejoratively referring to Breitbart. However, after spikes in coverage of Clinton's health and Clinton's emails, Breitbart also does not discuss these topics more than before the 9/11 ceremony or before the Comey announcement. These results, however, reflect that Breitbart was already covering these negative Clinton topics before they became major news stories due to an event. In the weeks prior to Clinton leaving the 9/11 event due to pneumonia, Breitbart had been running rumors that Clinton was having even more serious health problems. As one Breitbart headline put it after Clinton left the 9/11 ceremony with pneumonia, "Hillary Clinton 'Health Conspiracy' Turns Out to Be Real."[9] In the period before the 9/11 ceremony in my data, Clinton's health made up roughly 10% of coverage, while after the 7 day event window it made up roughly 2% of coverage.

Clinton's emails were a more long running topic across the political spectrum, but Breitbart focused less on Clinton's emails in the final few days of the campaign after the seven day window following Comey's October surprise so there is no significant difference. Indeed, Breitbart mentioned Clinton's emails on 6% of pages before Comey's announcement, 21% of pages in the seven day window following the announcement, and only 2% of pages in final few days of the election after the seven day window. In total then, Breitbart's coverage matches my first hypothesis that news outlets will pick up all negative stories briefly as campaign news, but does not match my second hypothesis that partisan news outlets will cover a negative topic about the opposing candidate more after an event window than before. Yet, the results are nonetheless consistent with partisan coverage in which Breitbart was already on message before the negative Clinton events occurred.

[8] http://www.breitbart.com/2016-presidential-race/2016/10/15/trump-media-clinton-campaign-want-to-stop-our-movement-with-blizzard-of-sexual-assault-accusations/

[9] http://www.breitbart.com/big-government/2016/09/14/hillary-clinton-health-conspiracy-turns-real/

Figure 4.13: Breitbart Coverage of Clinton Emails

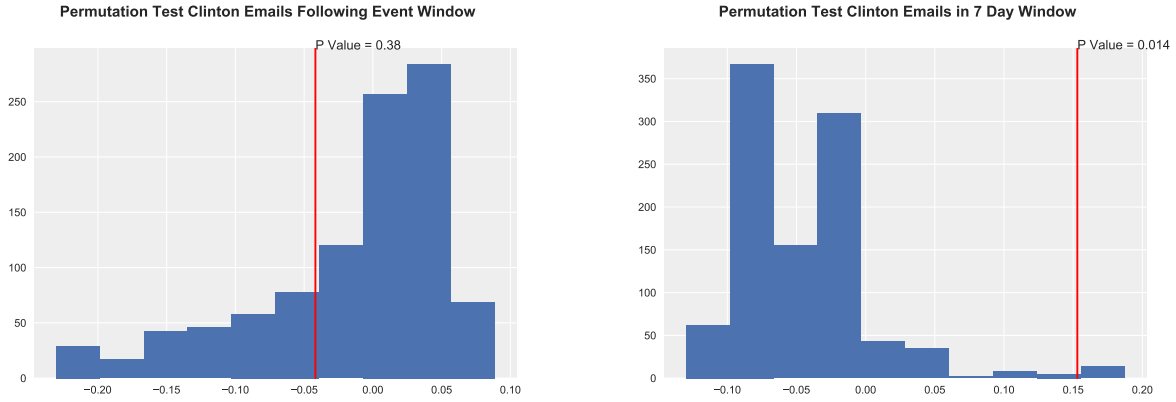


Figure 4.14: Breitbart Coverage of Clinton Health

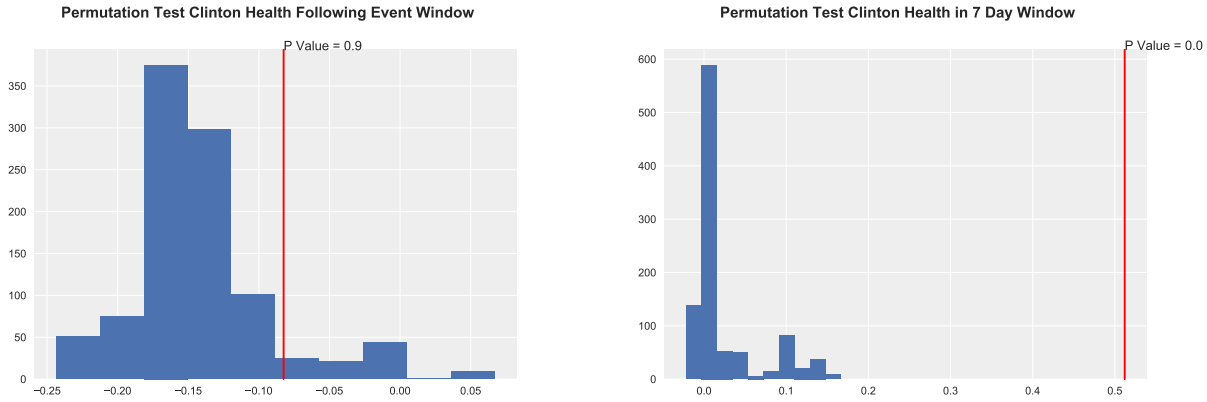


Figure 4.15: Breitbart Coverage of Clinton Basket of Deplorables

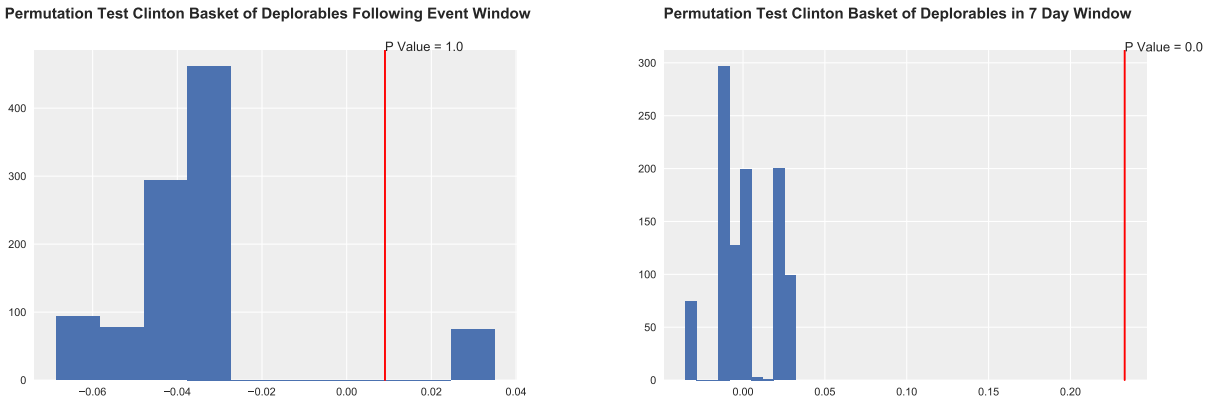


Figure 4.16: Breitbart Coverage of Trump's Taxes

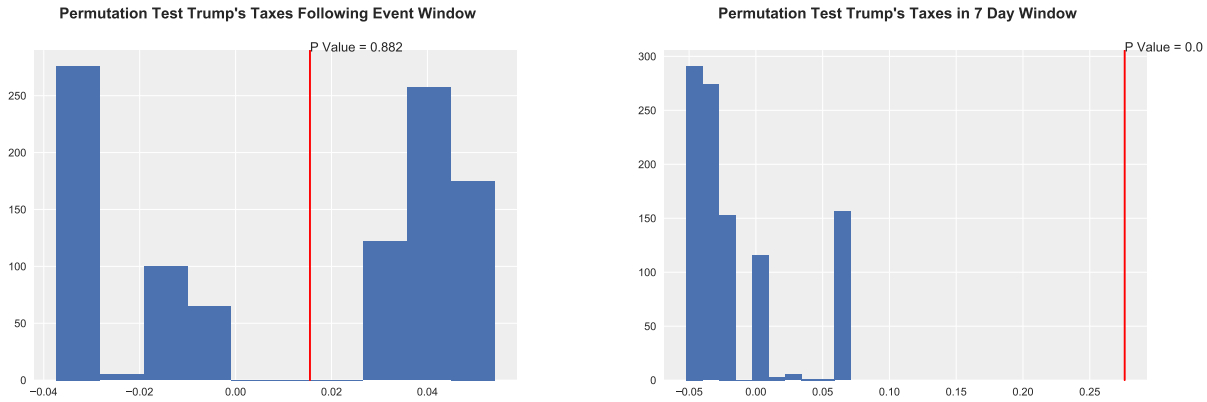
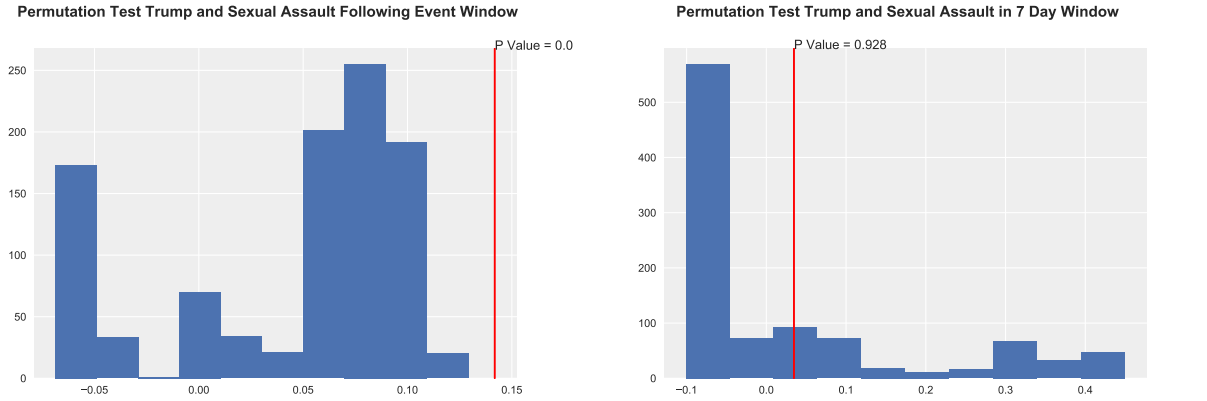


Figure 4.17: Breitbart Coverage of Trump and Sexual Assault



The New York Times

For every topic analyzed, the New York Times covered the topic significantly more in the seven days following the event ($\alpha = 0.1$ or lower). From roughly a 5.5% increase for Clinton's basket of deplorables remark to a 25% increase for discussing Kahn along with Trump, the New York Times covered every one of these topics more just following the major related event. In contrast, none of these topics occurred at significantly higher rates after the seven day window than before the event. In other words, the New York Times covered every one of these stories as the event occurred, but did not make these topics a permanent feature of the campaign any more than they were prior to the major event that caused them to temporarily cover the topic more. Of these topics, the only one that appeared on more than 2% percent of pages before the event occurred was Hillary Clinton's emails, which occured on 4% of pages on average prior to the announcement. Thus, the other topics would not have

had to become overwhelmingly prevalent for me to find that the New York Times covered them significantly more after the 7 day event window than before the event.

One notable aspect of the New York Times' coverage was their coverage of Trump's taxes. In the seven day window following their own story revealing that Trump may have paid no income taxes for as many as 18 years, the topic appeared on 10% of pages. Yet, the topic as I defined it only appeared on .2% of pages before their story and only 1.4% of pages after the seven day window.[10] Why would the New York Times let their own big break fade out of the news? To understand this result, we should keep in mind that journalists "must advance stories either through indexing or through finding developments that push the narrative" (Bennett 2003, 135). By indexing Bennett means reporting on and cataloguing elite opinion. Thus, absent a new development on Trump's taxes or a relentless focus by the Clinton campaign, the New York Times was not going to run a story every day entitled "Donald Trump Still Has Not Released His Tax Returns, May Have Paid No Taxes for Nearly Two Decades" even though they broke the story. Furthermore, another major event did happen just six days later: the Washington Post released the leaked Access Hollywood tape on which on which Trump chatted with Billy Bush about sexually assaulting women. Perhaps it is understandable that the New York Times moved on to covering the next major event under the circumstances, but the consequence was that their big break about Trump's unreleased taxes fell out of the headlines.

The two stories that the New York Times covered both very intensely for a brief period of time and over a longer period of time were Trump's remarks discussing sexual assault and Clinton's unauthorized use of a private email server while Secretary of State. Before and after the seven day window following Comey's announcement that he was reopening the investigation into Clinton, mentions of Clinton's emails appeared on 4-5% of pages. Even during the 7 day window following Comey's announcement, the story appeared on only 10% of pages. Thus, the New York Times coverage of Clinton's emails did significantly increase in the seven day window following the Comey announcement, but also included a slow drip of coverage throughout the campaign. For Trump and sexual assault, the topic was almost never covered as I measure it before the release of the Access Hollywood tape. In the seven day window after the tape was leaked the topic made up 14% of coverage, but thereafter just 5% of coverage mentioned the topic.

On net, if we take the share of pages that only mentioned a negative topic about Trump and subtract the share of pages that only mentioned a negative topic about Clinton the daily average is -.02. Thus, the New York Times coverage was ever so slightly more likely to include a negative Clinton topic than a negative Trump topic. This analysis obviously does not include a sentiment analysis of how they cover these topics. It also only looks at a subset of topics where I expected there to be a large spike in coverage following a major event and does not include other negative topics for Trump, like allegations of fraud at

[10]Because I only look at one particular term, it is possible the substantive topic appears on more pages using different terminology. However, within my definition of the topic I am nonetheless able to detect a large change in the window suggesting that they use the same term and I am measuring the actual overtime variation of coverage of Trump's taxes.

Figure 4.18: New York Times Coverage of Clinton Emails

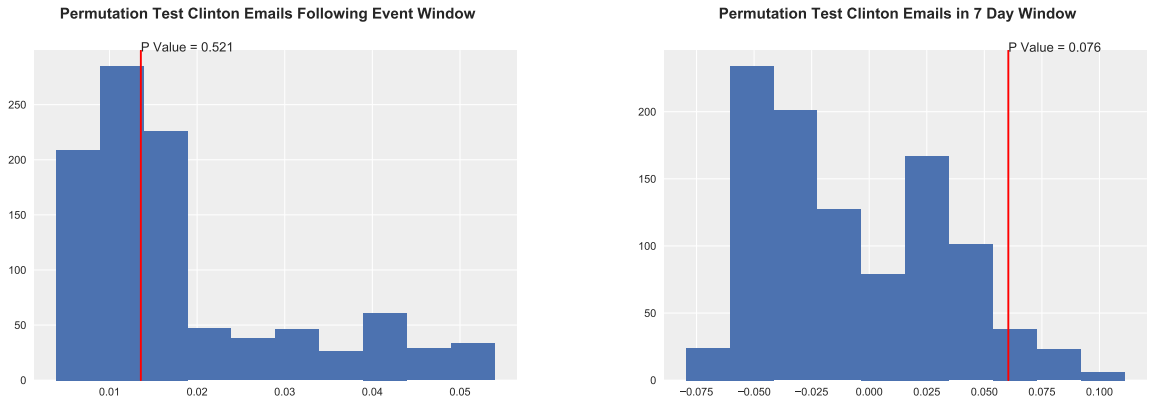
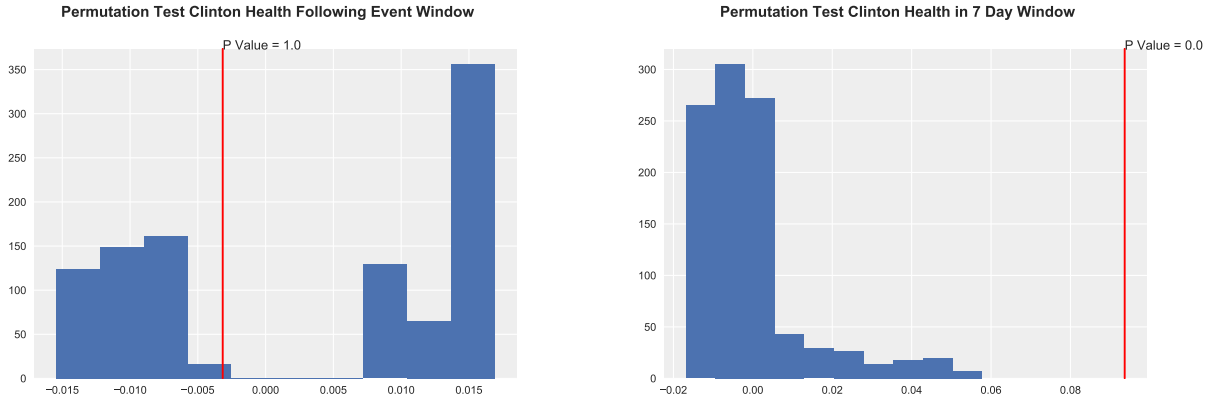


Figure 4.19: New York Times Coverage of Clinton Health



Trump University or Russian involvement in the election. Nevertheless, this is a striking result when we consider that this also includes the New York Times' editorial coverage.

Figure 4.20: New York Times Coverage of Clinton Basket of Deplorables

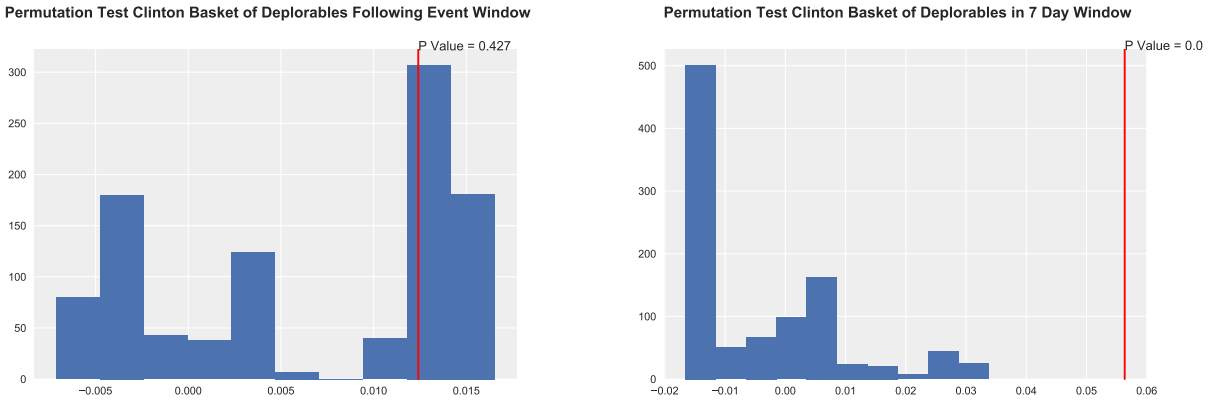


Figure 4.21: New York Times Coverage of Trump/Khan

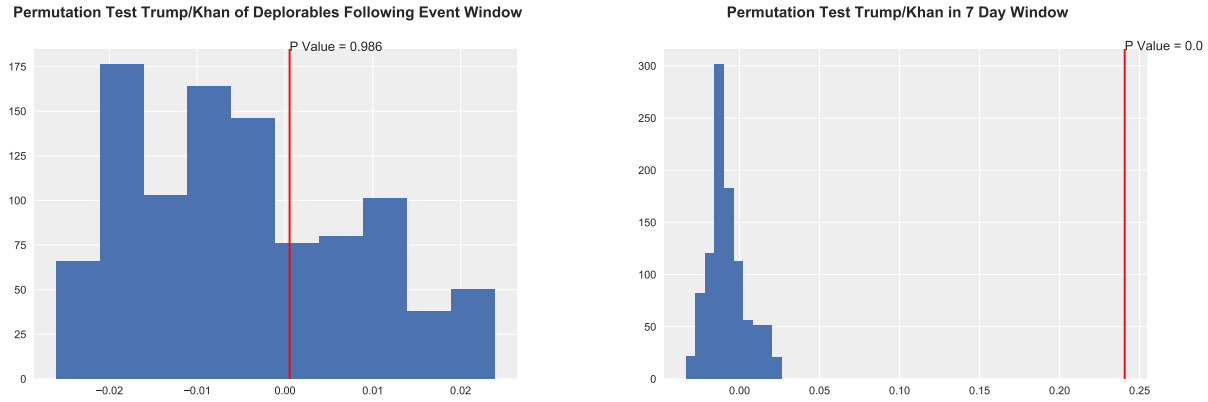


Figure 4.22: New York Times Coverage of Trump's Taxes

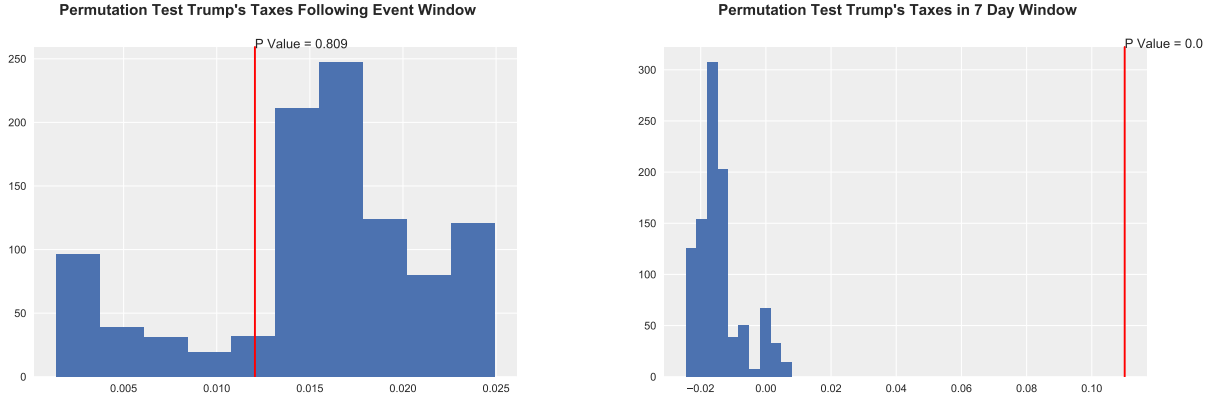
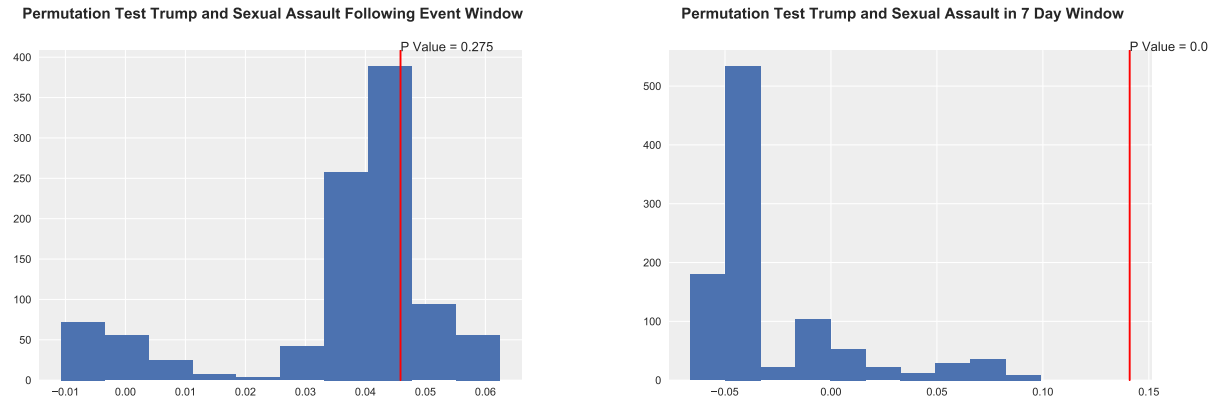


Figure 4.23: New York Times Coverage of Trump and Sexual Assault



4.7 Conclusion

I find that news outlets across the political spectrum pick up major negative news stories, even when they are a partisan outlet and the story is about their preferred candidate. In the New York Times, however, none of the six stories that I analyze become a more prominent part of campaign coverage throughout the rest of the campaign. In contrast, more partisan outlets like Daily Kos did make negative stories about Trump a permanent part of their campaign coverage following a major event. Breitbart, in contrast, was already featuring many of the negative Clinton topics like I analyzed. For example, Breitbart was speculating about Clinton's health before Clinton left the 9/11 ceremony, needing help walking due to walking pneumonia. Thus, I find mixed evidence that the events I analyze caused permanent upticks in coverage for a partisan outlet's opponent. Nevertheless, the behavior I observe is consistent with partisan patterns of coverage for partisan outlets.

One way someone might characterize the New York Times coverage would be to call it unbiased because they covered the negative topics about Clinton and the negative topics about Trump included in this study at relatively equal rates, in a relatively similar manner over time. As I have noted, this paper does not measure the sentiment of coverage and operates on the assumption that these topics are generally bad for the candidate involved, so one response from anyone dissatisfied with the coverage might be to point to the sentiment of that coverage. More fundamentally, however, to measure ideological bias a researcher would need to define ideological truth. The term bias implies that there is a position that we know to be true: saying the New York Times has a liberal bias implies that the truth is to the right of the New York Times (Groseclose 2011). In contrast, saying that the New York Times has a liberal slant only implies difference relative to a reference point. If we take the views of the New York Times' Editorial Board as a reference point, the balance of topics covered by the newspaper has an oddly right wing slant.

In total, my findings point towards the importance of events — whether candidate re-

marks, convention speeches, investigative journalism, or FBI investigations — to drive coverage. They also suggest that more traditional media organizations like the New York Times are heavily focused on the latest story to break. The short duration of media coverage of a topic in a presidential race underscores the importance of coverage close to Election Day. This is particularly true in down ballot races where voters are more persuadable, but effects decay faster (Hill et al. 2013). For presidential races, where past research has found communication effects decay at a slower rate, it also underscores the importance for candidates of keeping negative stories about their opponent in the news day after day. To the extent that traditional news organizations require new events or new quotes from politicians to advance a political story, Donald Trump's twitter account is an excellent source of data for future study.

Chapter 5

Conclusion

In the course of writing my dissertation, I read hundreds of local newspaper articles about congressional candidates. What struck me most was how hard it would be to learn about politics from reading your local newspaper. The news is incredibly fragmented, typically failing to provide a link between different stories or explain the broader political context (Bennet 2016). In my quantitative analysis of congressional coverage I found that journalists will report on a policy position of a candidate in about 1 in 3 articles, but in reading this coverage I found that these reports were rarely accompanied by a detailed explanation of what was at stake. Instead, I found that newspapers provide incumbents of all stripes with overwhelmingly neutral coverage that rarely offers criticism of any kind. Because I broadly define and measure criticism, what this finding ultimately means is that newspaper readers see very little information that contradicts their incumbent representative. Some of the most informative pieces that actually described the impact of a policy were letters to the editor. Even here, however, newspapers typically favored a long list of short letters and typically did not explain the broader political context of the debate.

Incumbent representatives in competitive districts do receive slightly more criticism, in part because the newspaper can get a quote from their challenger. While candidates in competitive district see more criticism, in my random sample of in-depth coverage read by research assistants over 60% of criticism did not come from an opponent, the opponent's campaign, or the opposing party. Thus, it seems like there is plenty of spare column space for challengers willing to criticize their opponent. Journalists advance narratives in part by covering the opinions of political elites, so the lack of criticism in part reflects a lack of criticism from challengers.

One way to interpret the lack of criticism—even from challengers—would be to conclude that there is simply not much to be said because incumbents are for the most part faithfully representing their constituents and newspapers are helping voters meet "The needs of democracy... by scrutinizing the records of those incumbents whose achievements are in doubt and reelecting the rest with minimal fuss" (Zaller 2003, 124). This ignores, however, that the general lack of incumbent criticism may lead to fewer high quality challengers who can successfully supply journalists with these critical story lines. I showed that incumbents

that vote against a majority of their constituents on landmark legislation receive the same overwhelmingly neutral tone of coverage as more faithful representatives. I also found that even incumbents referred to the House Ethics Committee over alleged corruption do not receive significantly more criticism, but only a whiff of a scandal. Thus, I conclude that newspapers are not acting as effective third party watchdogs and House elections are being conducted in a dim light.

If Chapter 2 showed that even diligent newspaper readers will rarely read criticism of their incumbent representative, Chapter 3 examined what can happen when voters do not acquire political information. Participants in these experiments had the same information as voters in real world elections because we used actual voters and actual candidates just prior to real world elections. Thus, our participants knew roughly what they would know on Election Day when they cast their ballot. On our mock ballot, however, some voters saw the information they would see on the real ballot, while others saw that information plus candidate photographs. By comparing the results for participants assigned to the photo treatment to the control group we were able to provide experimental evidence that candidate appearance influences vote choice. Absent political knowledge, even strong partisans will sometimes cast a vote for a candidate just for looking the part. And in congressional primaries, voters have so little information that candidates get votes for looking the part even from strong partisans who are more generally knowledgeable about politics.

Because high information voters are not influenced by candidate appearance in the general election, Chapter 3 concludes optimistically about the potential to inform voters about candidate characteristics more relevant to effective governance. Local newspapers, however, do not provide vibrant political coverage and voters will always face a principal-agent problem in controlling their elected representatives. Given the embarrassing portrait in Chapter 3 of voters who cast their ballot for candidates just for looking the part, it is worth emphasizing that voters' behavior reflects the fact that an individual voter will almost never be pivotal in a large election. Given that the probability of changing the outcome of an election with a single vote is near zero, "The rational citizen will be more interested in information about how the election is likely to come out than in information that will help him cast a wise vote" (Zaller 1999, 15). To the extent that citizens do pursue information about making a wise vote choice, "The rational voter is engaged by political conflict and bored by political consensus" as policy is unlikely to change if there is elite consensus on an issue (Zaller 1999, 15). Thus, a media organization has an incentive to portray conflict in the political process and maintain the appearance of competitive elections. Indeed, I found that horse race coverage displaces policy coverage in coverage of House elections. Nevertheless, this does not for the most part lead to coverage of substantive debates, but to reporting on the campaigns themselves. While this may be rational for individual voters and newspapers may simply be responding to consumer demand, the collective consequence is that incumbent representatives are not facing intense scrutiny.

In Chapter 4, I examined a very different electoral context and media environment: the 2016 U.S. presidential election. Unlike House elections, where the local newspaper may be the only media organization covering the incumbent representative, every media organization in

one way or another covered the 2016 U.S. presidential campaign. Thus, if voters want to they should have no trouble acquiring information about the candidates. However, in practice repetition and recency are particularly important if we consider the voters in Chapter 3 who are paying so little attention to politics that they cast can cast a ballot based on candidate appearance. Thus, I examine the timing of coverage of negative stories about the two presidential candidates.

When a major event happens that has the potential to define a presidential campaign, news outlets can either ignore the event, pick up the story temporarily, or make the topic a more permanent feature of their campaign coverage. I show that both the more traditional New York Times and partisan websites pick up negative stories about the presidential candidates right after a related event occurs, but that only the partisan websites continue to cover negative topics about the opposing candidate at a higher rate for the remainder of the campaign. In the New York Times, none of the six topics that I analyze become a more prominent part of campaign coverage throughout the rest of the campaign. Thus, they briefly shine a spotlight on important events, but after a seven day window following the event they do not cover the topic significantly more than they did before the related event. In the context of the 2016 presidential campaign, focusing briefly on the latest story gave the mix of topics in the New York Times an oddly right wing slant relative to the opinions of its Editorial Board. As Donald Trump's presidency unfolds, however, the absence of a Democratic foil may mean that what is left are negative stories about Trump, even if most of the stories are briefly lived.

While my dissertation focused on the provision of all manner of political information through the mass media, the vast majority of voters only use one heuristic to make their vote choice and that piece of information is already on the ballot: is the candidate a Democrat or a Republican? Even these voters, however, benefit when third party institutions like the press effectively monitor politicians. In congressional primaries, even some strong partisans who are otherwise knowledgeable about politics can have so little information about the candidates that they unwittingly vote based on candidate appearance. In general elections, however, committed partisans are at the whims of inattentive swing voters. My dissertation shows that while these voters can learn about presidential candidates if they want to, even attentive news consumers would be hard pressed to learn about the policy positions and performance of each of their many elected representatives. For the most part, democracy is conducted in a dim light.

Bibliography

Ahler, Douglas J., Jack Citrin and Gabriel S. Lenz. 2016. "Do Open Primaries Improve Representation? An Experimental Test of California's 2012 TopTwo Primary." *Legislative Studies Quarterly* 41:237–268.

Ahler, Douglas J, Jack Citrin, Michael C Dougal and Gabriel S Lenz. 2016. "Face value? Experimental evidence that candidate appearance influences electoral choice." *Political Behavior* pp. 1–26.

Ai, Chunrong and Edward C. Norton. 2003. "Interaction terms in logit and probit models." *Economics Letters* 80:123–129.

Aidt, Toke S. 2000. "Economic voting and information." *Electoral Studies* 19(2):349–362.

Alley, Thomas R. 1988. *Social and applied aspects of perceiving faces*. Resources for ecological psychology Hillsdale, N.J.: Lawrence Erlbaum Associates.

Andreoni, James and Ragan Petrie. 2008. "Beauty, gender and stereotypes: Evidence from laboratory experiments." *Journal of Economic Psychology* 29:73–93.

Angrist, Joshua David and Jorn-Steffen Pischke. 2009. *Mostly harmless econometrics: an empiricist's companion*. Princeton: Princeton University Press.

Antonakis, John and Olaf Dalgas. 2009. "Predicting Elections: Child's Play!" *Science* 323:1183.

Arnold, R Douglas. 2004. *Congress, the Press, and Political Accountability*. Princeton University Press.

Atkinson, Matthew D., Ryan D. Enos and Seth J. Hill. 2009. "Candidate Faces and Election Outcomes." *Quarterly Journal of Political Science* 4:229–249.

Ballew, Charles C. II and Alexander Todorov. 2007. "Predicting political elections from rapid and unreflective face judgments." *Proceedings of the National Academy of Sciences of the United States of America* 104:17948–17953.

Banducci, Susan A., Jeffrey A. Karp, Michael Thrasher and Colin Rallings. 2008. "Ballot photographs as cues in low-information elections." *Political Psychology* 29:903–17.

Bar, Moshe, Maital Neta and Heather Linz. 2006. "Very first impressions." *Emotion* 6:269–278.

Belot, Michle, V. Bhaskar and Jeroen Van De Ven. 2012. "Beauty and the Sources of Discrimination." *Journal of Human Resources* 47:851–872.

Benjamin, Daniel J. and Jesse M. Shapiro. 2009. "Thin-slice Forecasts of Gubernatorial Elections." *The Review of Economics and Statistics* 91:523–36.

Bennett, W Lance. 2003. "The burglar alarm that just keeps ringing: A response to Zaller." *Political Communication* 20(2):131–138.

Bennett, W Lance. 2016. *News: The politics of illusion*. University Of Chicago Press.

Berggren, Niclas, Henrik Jordahl and Panu Poutvaara. 2010. "The Looks of a Winner: Beauty and Electoral Success." *Journal of Public Economics* 94:8–15.

Broockman, David E. 2016. "Approaches to Studying Policy Representation." *Legislative Studies Quarterly* 41(1):181–215.

Brusattin, Lorenzo. 2012. "Candidate visual appearance as a shortcut for both sophisticated and unsophisticated voters: evidence from a Spanish online study." *International Journal of Public Opinion Research* 24:1–20.

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

Chiao, Joan Y., Nicholas E. Bowman and Harleen Gill. 2008. "The Political Gender Gap: Gender Bias in Facial Inferences that Predict Voting Behavior." *PloS one* 3:e3666.

Clarke, Peter and Susan H Evans. 1983. *Covering campaigns: Journalism in congressional elections*. Stanford University Press.

Cohen, Rudolf. 1973. *Patterns of personality judgment*. New York, NY: Academic Press.

Converse, Philip E. 1964. The Nature of Belief Systems in Mass Publics. In *Ideology and Discontent*, ed. David E. Apter. New York: Free Press.

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

Explained, Writing. 2016. "AP Style Party Affiliation.".
URL: *http://writingexplained.org/ap-style/ap-style-party-affiliation*

Fiorina, Morris P. 1981. *Retrospective voting in American national elections*. New Haven: Yale University Press.

Fletcher, Jason M. 2009. "Beauty vs. brains: Early labor market outcomes of high school graduates." *Economics Letters* 105:321–325.

Gerber, Alan S, James G Gimpel, Donald P Green and Daron R Shaw. 2011. "How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment." *American Political Science Review* 105(01):135–150.

Grimmer, Justin and Brandon M Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political analysis* 21(3):267–297.

Groseclose, Tim. 2011. *Left Turn: How Liberal Media Bias Distorts the American Mind*. St. Martin's Press.

Hall, Crystal C., Amir Goren, Shelly Chaiken and Alexander Todorov. 2009. Shallow cues with deep effects: Trait judgments from faces and voting decisions. In *Political Psychology of Democratic Citizenship*, ed. Eugene Borgida, Christopher M. Federico and John L. Sullivan. Oxford University Press pp. 73–99.

Hamermesh, Daniel S. 2006. "Changing looks and changing discrimination: The beauty of economists." *Economics Letters* 93:405–412.

Hamermesh, Daniel S. 2011. *Beauty pays: why attractive people are more successful*. Princeton, N.J.: Princeton University Press.

Hamermesh, Daniel S. and Amy Parker. 2005. "Beauty in the classroom: Instructors pulchritude and putative pedagogical productivity." *Economics of Education Review* 24:369–376.

Hassin, Ran and Yaacov Trope. 2000. "Facing faces: studies on the cognitive aspects of physiognomy." *Journal of Personality and Social Psychology* 78:837–52.

Henderson, John. 2013. "Issue Distancing in Congressional Elections." *Working Paper* .

Hill, Seth. 2015. "Balancing Representation of Primary and General Electorates in Congressional Roll Call Votes." *Working Paper* .

Hill, Seth J, James Lo, Lynn Vavreck and John Zaller. 2013. "How quickly we forget: The duration of persuasion effects from mass communication." *Political Communication* 30(4):521–547.

Imbens, Guido W and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Johns, R. and M. Shephard. 2007. "Gender, Candidate Image and Electoral Preference." *British Journal of Politics and International Relations* 9:434–460.

Kahn, Kim Fridkin and Patrick J Kenney. 2002. "The slant of the news: How editorial endorsements influence campaign coverage and citizens' views of candidates." *American Political Science Review* 96(02):381–394.

Kalick, S. Michael, Leslie A. Zebrowitz, Judith H. Langlois and Robert M. Johnson. 1998. "Does human facial attractiveness honestly advertise health? Longitudinal data on an evolutionary question." *Psychological Science* 9:8–13.

Key, V. O. 1966. *The responsible electorate: rationality in presidential voting, 1936-1960*. Cambridge, MA: Harvard University Press.

King, Amy and Andrew Leigh. 2009. "Beautiful Politicians." *Kyklos* 62:579–93.

King, Gary. 1991. ""Truth" Is stranger than prediction, more questionable than causal inference." *American Journal of Political Science* 35:1047–53.

King, Gary, James Honaker, Anne Joseph and Kenneth Scheve. 2001. Analyzing incomplete political science data: An alternative algorithm for multiple imputation. In *American Political Science Association*. Vol. 95 Cambridge Univ Press pp. 49–69.

Klein, Markus and Ulrich Rosar. 2005. "Physische Attraktivitt und Wahlerfolg. Eine empirische Analyse am Beispiel der Wahlkreiskandidaten bei der Bundestagswahl 2002." *Politische Vierteljahresschrift* 46:263–287.

Krasno, Jonathan S. 1997. *Challengers, competition, and reelection: Comparing Senate and House elections*. New Haven, CT: Yale University Press.

Kuklinski, James H. and Paul J. Quirk. 2000. Reconsidering the rational public. In *Elements of reason: Cognition, choice, and the bounds of rationality*, ed. Arthur Lupia and Mathew Daniel McCubbins. New York: Cambridge University Press pp. 153–82.

Laskas, Jean Marie. 2012. "Bob Dole: Great American." *GQ* pp. 88–90.

Lawson, Chappell, Gabriel S. Lenz, Michael Myers and Andy Baker. 2010. "Looking like a winner: Candidate appearance and electoral success in new democracies." *World Politics* 62:561–593.

Lenz, Gabriel S. 2012. *Follow the Leader? How Voters Respond to Politicians' Policies and Performance*. University of Chicago Press.

Lenz, Gabriel S. and Chappell Lawson. 2011. "Looking the part: Television leads less informed citizens to vote based on candidates appearance." *American Journal of Political Science* 55:574–589.

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

Lupia, Arthur and Mathew D. McCubbins. 1998*a*. "The Democratic Dilemma: Can Citizens Learn What They Need to Know.".

Lupia, Arthur and Mathew D. McCubbins. 1998*b*. *The democratic dilemma: can citizens learn what they really need to know?* New York: Cambridge University Press.

Martin, D. S. 1978. "Person perception and real-life electoral behaviour." *Australian Journal of Psychology* 30:255–262.

Mattes, Kyle, Michael Spezio, Hackjin Kim, Alexander Todorov, Ralph Adolphs and R. Michael Alvarez. 2010. "Predicting election outcomes from positive and negative trait assessments of candidate images." *Political Psychology* 31:41–58.

Minozzi, William. 2011. "A jamming theory of politics." *The Journal of Politics* 73(02):301–315.

Nelson, Colleen McCain. 2012. "In Denver, Clinton Says Romney Plan Doesnt Add Up.".

Olivola, Christopher Y. and Alexander Todorov. 2010. "Elected in 100 milliseconds: Appearance-based trait inferences and voting." *Journal of nonverbal behavior* 34:83–110.

Poutvaara, Panu, Henrik Jordahl and Niclas Berggren. 2009. "Faces of politicians: Babyfacedness predicts inferred competence but not electoral success." *Journal of Experimental Social Psychology* 45:1132–1135.

Ravina, Enrichetta. 2012. "Love & loans: the effect of beauty and personal characteristics in credit markets." *Available at SSRN 1101647* .

Riggio, Heidi R. and Ronald E. Riggio. 2010. "Appearance-based trait inferences and voting: Evolutionary roots and implications for leadership." *Journal of nonverbal behavior* 34:119–125.

Rosar, Ulrich, Markus Klein and Tilo Beckers. 2008. "The Frog Pond Beauty Contest: Physical Attractiveness and Electoral Success of the Constituency Candidates at the North Rhine-Westphalia State Election 2005." *European Journal of Political Research* 47:64–79.

Rosenberg, Shawn W. and Patrick McCafferty. 1987. "The Image and the Vote: Manipulating Voters' Preferences." *Public Opinion Quarterly* 51:31–47.

Schaffner, Brian F. and Matthew J. Streb. 2002. "The partisan heuristic in low-information elections." *Public Opinion Quarterly* 66:559–581.

Snyder Jr, James M and David Strömberg. 2010. "Press Coverage and Political Accountability." *Journal of Political Economy* 118(2):355–408.

Spezio, Michael L., Antonio Rangel, R. Michael Alvarez, John P. O'Doherty, Kyle Mattes, Alexander Todorov, Hackjin Kim and Ralph Adolphs. 2008. "A neural basis for the effect of candidate appearance on election outcomes." *Social Cognitive and Affective Neuroscience* 3:344–352.

Spezio, Michael L., Laura Loesch, Frdric Gosselin, Kyle Mattes and R. Michael Alvarez. 2012. "Thin-Slice Decisions Do Not Need Faces to be Predictive of Election Outcomes." *Political Psychology* 33:331–341.

Tausanovitch, Chris and Christopher Warshaw. 2013. "Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities." *The Journal of Politics* 75(02):330–342.

Todorov, Alexander, Anesu N. Mandisodza, Amir Goren and Crystal C. Hall. 2005. "Inferences of Competence from Faces Predict Election Outcomes." *Science* 308:1623–1626.

van der Laan, Mark J, Eric C Polley and Alan E Hubbard. 2007. "Super learner." *Statistical Applications in Genetics and Molecular Biology* 6(1):1–21.

Waismel-Manor, Israel and Yariv Tsfati. 2011. "Why Do Better-Looking Members of Congress Receive More Television Coverage?" *Political Communication* 28:440–463.

Wilson, Rick K. and Catherine C. Eckel. 2006. "Judging a book by its cover: Beauty and expectations in the trust game." *Political Research Quarterly* 59:189–202.

Zaller, John. 1992. *The nature and origins of mass opinion*. Cambridge university press.

Zaller, John. 2003. "A new standard of news quality: Burglar alarms for the monitorial citizen." *Political Communication* 20(2):109–130.

Zebrowitz, Leslie A. 1997. *Reading Faces*. Boulder, CO: Westview Press.

Zebrowitz, Leslie A., Judith A. Hall, Nora A. Murphy and Gillian Rhodes. 2002. "Looking Smart and Looking Good." *Personality and Social Psychology Bulletin* 28:238–249.