Reading Between (the Party) Lines

by

Sophie Beiying Chou

Submitted to the MIT Media Lab,
School of Architecture and Planning
in partial fulfillment of the requirements for the degree of

MS in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2016

Author . . . .
MIT Media Lab
May 5, 2016

Certified by . . . .
Deb Roy
Associate Professor
Thesis Supervisor

Accepted by . . . .
WHO IS THE CHAIR(WO)MAN?
Chairman, Department Committee on Graduate Theses

Reading Between (the Party) Lines

by

Sophie Beiying Chou

Submitted to the MIT Media Lab
on May 5, 2016, in partial fulfillment of the
requirements for the degree of
MS in Media Arts and Sciences

Abstract

TO-DO

Thesis Supervisor: Deb Roy
Title: Associate Professor

The following people served as readers for this thesis:

Sepandar Kamvar. . . . .

Associate Professor of Media Arts and Sciences

MIT Media Lab

Iyad Rahwan . . . .

Associate Professor of Media Arts and Sciences

MIT Media Lab

Acknowledgments

[FILL IT WITH GRATITUDE]

Contents

Chapter 1

Introduction

Does anyone trust the news anymore? Not according to the latest Gallup Poll, which showed that only 4 in 10 Americans believe that mass media does a good job of reporting the news "fully, fairly, and accurately." It's a major decline since the poll was first taken in 1999, back when more than half (55%) of Americans believed the news was trustworthy [11].

And the trend has been steadily downward: in short, the majority of Americans have had little to no trust in mass media news coverage since 2007: a discouraging view for a tumultuous time in journalism.

But beyond frustrated readers and reporters, why does distrust in the news matter? For one, media bias—or at the very least, the *belief of* a biased media bias—may have a significant impact on the practice of democracy. A 2006 study from Georgetown University shows that those with more negative attitudes towards the news tend to be more highly influenced by their partisan prior beliefs and less by contemporary issues and messages when voting [6]. This implies that distrust of media plays a large role in the polarization of American politics.

In light of the upcoming 2016 elections, this thesis explores perceptions of media trust and fairness in coverage of the presidential candidates. Claims of media bias and favoritism are especially high-stakes in election years, where trust has been shown to plummet [11]. And in this election cycle, cries of bias have been especially loud: Analysis at the New York Times showed that the news media gave Republican candidate Donald Trump a $1.9 billion advantage in free publicity, an amount 190 times as much as paid advertising [2].

In this thesis, we examine some of the factors that contribute to the perception of media bias. In particular, how does the *content* of a story (reading level and vocabulary) affect the reader versus the *context* (publication and author)?

We break down the larger question of media bias in two dimensions: trust and fairness in reporting, and examine the role of language in influencing the reader. Although studies have been conducted to both examine the psychological effect of wording on believability and the impact of media brands and bias, separating and comparing these two factors remains largely unexamined [15, 4]

Two studies are performed: a preliminary and main experiment to collect reader's perceptions of news stories through crowdsourcing. We manipulate the source of the story to examine effects of media brands on the reader, and also compare trust and fairness rankings between high and low reading level stories.

Although the general consensus of mistrust is clear, perception of media bias is a complex phenomenon to dissect, as it combines social and psychological effects with the traits of the story itself. In this thesis, we hope to shed new light on what motivates readers' trust and distrust of news media.

Chapter 2

In Media We ... Trust?

In an election year prefaced by deep cynicism towards American institutions (a 2015 survey showed that just 19% of the population trusts the federal government), attitudes towards the news media fare no better. Almost two-thirds of Americans think that the national news media is a negative influence on the country: but how does this attitude influence Americans? [3]

In this section, we outline three main potential sources of media bias—the reader, the source, and the language—and explore the impact of each.

2.1 The Role of the Reader in Perceptions of Bias

It comes as no surprise that our own political stances have a significant effect in our perceptions of bias in the media.

In even seemingly neutral stories, partisans tend to view reporting as biased against their own views. This phenomenon—deemed the "hostile media effect"—was first studied at Stanford University by Robert P. Vallone, Lee Ross, and Mark R. Lepper in 1985 [12]. Although "true" neutrality of a story is nearly impossible to quantify due to the subjective nature of the concept, Vallone et. al were able to successfully demonstrate that partisans of *both* sides (pro-Israeli and pro-Arab) viewed the same news segments as hostile towards their beliefs and favorable to the other side.

2.2 The Role of Media Brands in Perceptions of Bias

2.3 The Role of Language [Policial Persuasion]

2.3.1 Language and Politics

2.3.2 The Seductive Allure [... of Simple] Language

2.4 The 2016 Elections

2.4.1 Criticism of Media Bias

Chapter 3

Data Collection

3.1 The Electome

The Electome is a large, collaborative, and ongoing effort in the Laboratory for Social Machines that seeks to analyze the "competition of ideas" in the upcoming 2016 elections. It does so by using techniques in natural language processing, machine learning, and network analysis to make sense of "big data" collected from two main sources: traditional media (online versions of news publications) and social media (Twitter) [13].

This thesis, which emerged from the Electome, examines a narrowed portion of the first dataset centered around specific topics and candidates. The following section will describe the methods used to gather this dataset as well as shared machine learning tools for article classification.

3.2 Story Collection

News articles from 14 different news publications were systematically collected every hour from RSS feeds beginning from January 2015. The outlets tracked are:

- CNN
- Fox News
- The Wall Street Journal
- ProPublica
- Politico
- McClatchy
- The Washington Post
- Buzzfeed (News only)
- National Public Radio (NPR)
- The Huffington Post
- The Associated Press
- Reuters
- The New York Times
- The Los Angeles Times

The above outlets were chosen to form a diverse subset of the current U.S. news ecosystem, including a combination of private and public, liberal and conservative, legacy and new media publications. Also included are wire services and a mix of media delivery formats for which the outlet is known (radio, television, print, or web).

Steps to collect the news stories were as follows:

1. For each news publication:
   (a) Use regular expressions to extract all RSS feed urls for a news site.
   (b) For each RSS feed:
      i. Parse feed using open source xml reader library, Feedparser.
      ii. For each link to a story in the feed:
         A. Parse html using Beautiful Soup 3 (an open source python library)
         B. Insert headline, authors, story text, publication date and retrieval date into an SQL database.

Data depulication (by story url and headline) is then performed to ensure only one copy of each article is in the database. This step is necessary as articles from wire services often appear across many outlets and effect aggregate text analsyis.

On average, 2,000 stories are collected per day across all outlets. However, volume follows a consistent pattern of fluctuation depending on weekday, ranging from approximately 1,000 to 3,000 stories.

[INSERT HERE GRAPH OF NEWS STORIES VOLUME BY WEEKDAY]

As of March 1st, 2016, there were 855,000 stories collected in the database and 43,000 journalists.

3.3 Election Classification[1]

All stories collected from the sources above are passed through a machine learning classifier to determine if they are primarily about the 2016 U.S. elections. This thesis examines only those articles classified as election related.

The election classifier consists of a binary Maximum Entropy (MaxEnt) text classifier using Bag-of-Word (BoW) features selected from the news articles [9]. The features are ranked according to the chi-squared test (where high scores indicate that the null hypothesis of independence should be rejected and thus the occurence of class and term are dependent) with a cutoff of 20,000. We use the open-source Python library scikit-learn for the implementation our MaxEnt classifier [10].

The classifier is trained on a balanced dataset of 1,000 manually labelled news articles and evaluated on a separate balanced test set of 300 articles. We achieved a precision of 90% and recall of 91% (F-score of 92%).

Between January 1, 2015 and March 1, 2016 there were 24,837 articles with over 90%

[1]This section features shared machine learning tools within the Electome, with acknowledgements to Prashanth Vijayaraghavan.

confidence level of being election related. The number of stories classified as such has increased over time as election day nears.

[INSERT % ELECTION/ % TOTAL STORIES CHART HERE]

3.4 Topic Classification[2]

The final step of article processing within the Electome pipeline for this experimental dataset is the application of a 22-topic classifier. The following 22 topics were curated within the Laboratory for Social Machines as central issues of discussion within the election:

- Income Inequality
- Environment/Energy
- Jobs/Employment
- Guns
- Racial Issues
- Foreign Policy/National Security
- LGBT Issues
- Ethics
- Education
- Financial Regulation
- Budget/Taxation
- Veterans
- Campaign Finance
- Surveillance/Privacy
- Drugs
- Justice
- Abortion
- Immigration
- Trade
- Health Care
- Economy
- Other

3,000 articles classified as election related by the methods detailed in section 3.3 were manually labelled to form our training dataset. Articles were labelled as belonging to one or more topics. We then used a two-step model to create the classifier, due to the challenges of having a large number of classes and relatively small number of labeled stories. First, thousands of election related articles were inputted into a domain adaptive semi-supervised (stories were not all labeled) topic classification system. Then, a denoising autoencoder (DA) was used to learn salient features in an unsupervised fashion [14]. These features were used to train a topic classifier using the labelled dataset.

[2]This section features shared machine learning tools within the Electome, with acknowledgements to Prashanth Vijayaraghavan.

The classifier was evaluated on an independent dataset of 400 manually annotated articles. We achieved a precision of 91% and a recall of 94% (weighted F-score of 92%).

3.5 Flesch-Kincaid Readability Tests

In this study, we focus primarily on the Flesch-Kincaid (F-K) tests for estimating text readability. Originally developed for the U.S. Navy in 1975 for assessing the difficulty of technical manuals, the F-K reading level corresponds roughly to U.S. grade level and the reading ease score is inversely proportional to the grade level on a scale from 0 to approximately 120 [5].

We chose the F-K tests over other comparable ones due to its popularity in educational assessment and other applications, including in legislation. For example, it is required by law in Florida that life insurance policies have a Flesch reading ease of 45 or greater (less than 12th grade in reading level) [7]. The F-K tests are also bundled in many common word processing services, including Microsoft Office Word. As a comparison, basic article analysis is also computed using the Gunning fog index (see Section 5.2.1).

The formula for Flesch reading ease is as follows:

$$206.835 - 1.015\left(\frac{\text{total words}}{\text{total sentences}}\right) - 84.6\left(\frac{\text{total syllables}}{\text{total words}}\right)$$

And for reading grade level:

$$0.39\left(\frac{\text{total words}}{\text{total sentences}}\right) + 11.8\left(\frac{\text{total syllables}}{\text{total words}}\right) - 15.59$$

The two formulas are not directly comparable due to the difference in weighting factors. For ease of metaphor, we use the grade level tests in our analysis. Syllable length is highly weighted in this formula, so it is possible to generate a story of very high reading level that consists of a single word in a single sentence (the longest English word, *pneumonoultramicroscopicsilicovolcanoconiosi*, a type of lung disease, has a reading grade level of 197.2), which is a limitation of the method, since texts with polysyllabic words are not always necessarily more difficult to read.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4

Experimental Design

4.1 Data Selection

For this study, we chose to analyze stories collected between January 1, 2016 (the start of the election year) and March 1, 2016 (Super Tuesday). Since a large number of states hold primary elections and caucuses on Super Tuesday, it is seen as an early indicator of candidate electability. All stories had been filtered through both the election (see section 3.3) and topic (see section 3.4) classifiers.

Based on the results of Super Tuesday, we selected four candidates for this study by delegate count: Hillary Clinton (1,279), Bernie Sanders (1,027), Donald Trump (743), and Ted Cruz (517) [1].

News articles were then separated into single-candidate stories (i.e. articles featuring primarily one candidate in the headline) to be able to measure more clearly the perceived bias per candidate. This was done programatically using regular expressions to determine if a headline contained one candidate and one candidate only. A dictionary of related names was created to make sure that stories were correctly categorized (i.e. "Hillary", "Clinton", and "Hillary Clinton" were to be categorized as pertaining to “Hillary Clinton” but not if preceded by “Bill”).

4.1.1 Publication Selection

For the purposes of this study, stories were examined from five outlets:

- CNN
- Fox News
- The New York Times
- The Wall Street Journal
- The Associated Press

The choices consist of two pairs of outlets in both print and television across the liberal-conservative divide, plus a wire service. Of the 14 outlets above, both Fox News and the Wall Street Journal have an audience that leans conservative compared to the overall population (27% mostly conservative viewers versus 17% in the overall population for Fox News and 22% mostly conservative viewers versus 17% in the overall population) measured by a 2014 Pew survey [8].

On the other hand, the New York Times and CNN both have audiences that lean mostly liberal (25% liberal versus 22% in all respondents for CNN and 25% for the New York Times). The Associated Press, which was not included in the survey, has members in outlets across the political divide and was chosen as an experimental control.

[MIGHT INCLUDE THOSE DISTRIBUTIONS HERE]

4.1.2 Topic Selection

The top four topics by volume (Immigration, Abortion, Campaign Finance, Foreign Policy/National Security) were chosen for the survey to ensure a significant number of stories for each candidate for each topic. For overall topic distributions, see 9 in the Appendix.

4.1.3 Flesch Kincaid Cutoffs

survey size = 5 stories + 1 Gold 120 stories total = 24 surveys (24 rows) x 2 = 48 rows

4.1.4 Redaction of Stories

4.2 CrowdFlower

Media Perceptions Study 1 (Massachusetts Institute Of Technology)

Instructions ▴

Overview

This job is part of a MIT scientific research project. Your decision to complete this job is voluntary. There is no way for us to identify you. The only information we will have, in addition to your responses, is the time at which you completed the survey. The results of the research may be presented at scientific meetings or published in scientific journals. Clicking on the 'SUBMIT' button on the bottom of this page indicates that you are at least 18 years of age and agree to complete this job voluntarily.

Steps

For each news story below, read the story completely and then select the choices that best describe your reaction to the story. At the end of the survey, you will be asked a few questions about your demographics and political affiliation. [The letters “XXX” are used to hide and show certain words by the researchers. It was not originally in the story and try not let it affect your answers.]

Rules & Tips

There are no "correct" answers. Take your time to read the story and fill in the blanks. It is strictly forbidden to attempt to Google or attempt to read the news articles elsewhere during this job.

4.3 Demographic Survey

4.4 Political Affiliation Survey

4.5 Quality Assurance

Chapter 5

Pre-Survey Analysis

5.1 Topic Analysis

5.2 Flesch-Kincaid Analysis

5.2.1 Comparisons to other Reading Level Tests

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 6

Study One

Basic stats here: we ran over X days over data points

6.1 Reader Demographics

6.2 Media Favorability of Candidates

Each reader was asked to score the five stories according to how favorable each one was to the featured candidate (by headline).

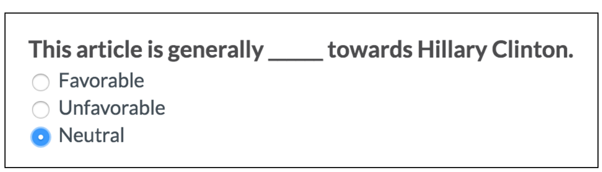

Figure 6-1: Example of favorability scoring question

Scores were collected on a three-point scale, Favorable (1), Unfavorable (-1), or Neutral (0).

Overall, media coverage of Trump was viewed as most negatively biased, with over half of stories (51.1%) viewed as unfavorable towards the candidate.

Of the stories shown, both Sanders and Clinton were viewed as having more positive than negative coverage, at 38.9% of the 180 annotations being positive. Sanders also had the least negative coverage, with only 18.3% stories shown being viewed as negatively biased against the candidate. Republican candidate Cruz was also seen to have more negative (33.3%) than positive (28.9%) stories about him, although the majority were seen as neutral (37.8%).

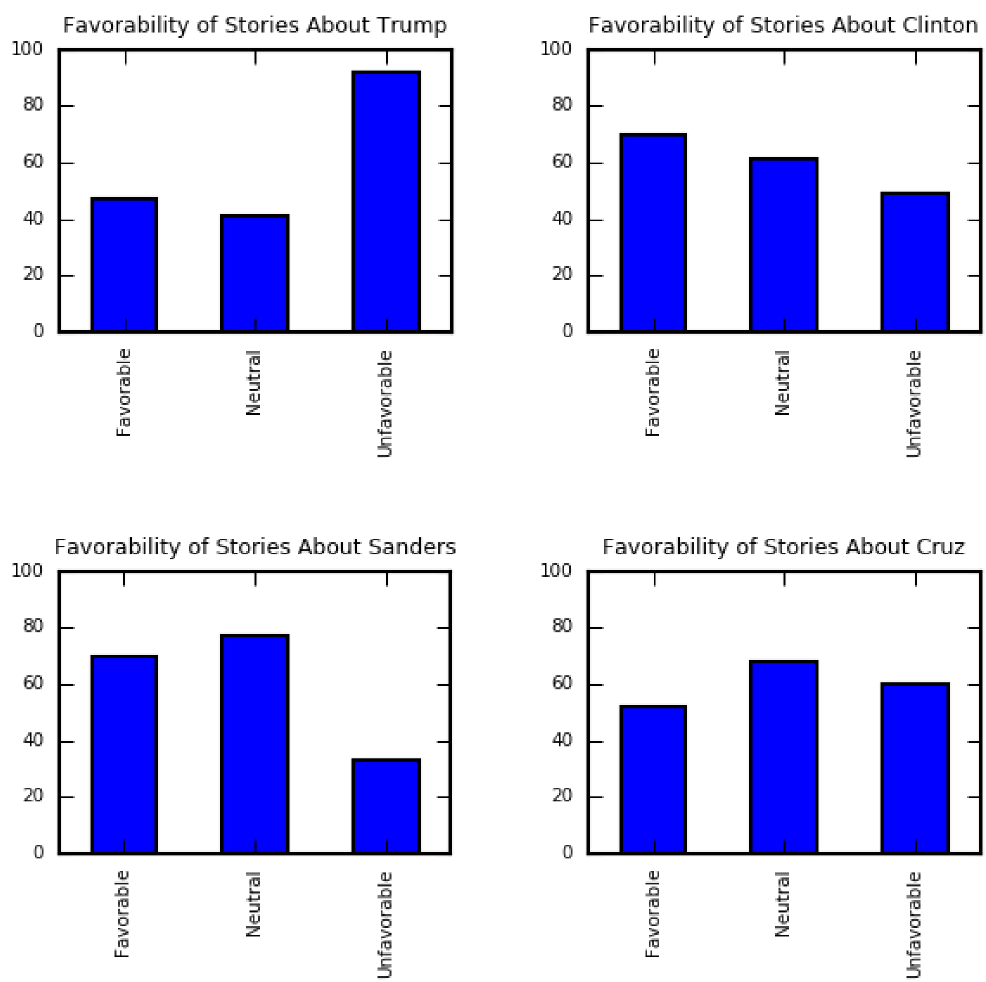

Figure 6-2: Media Favorability of Candidates

These trends persist when we filter responses by stories that were considered trustworthy or at least neutral (score > 0).

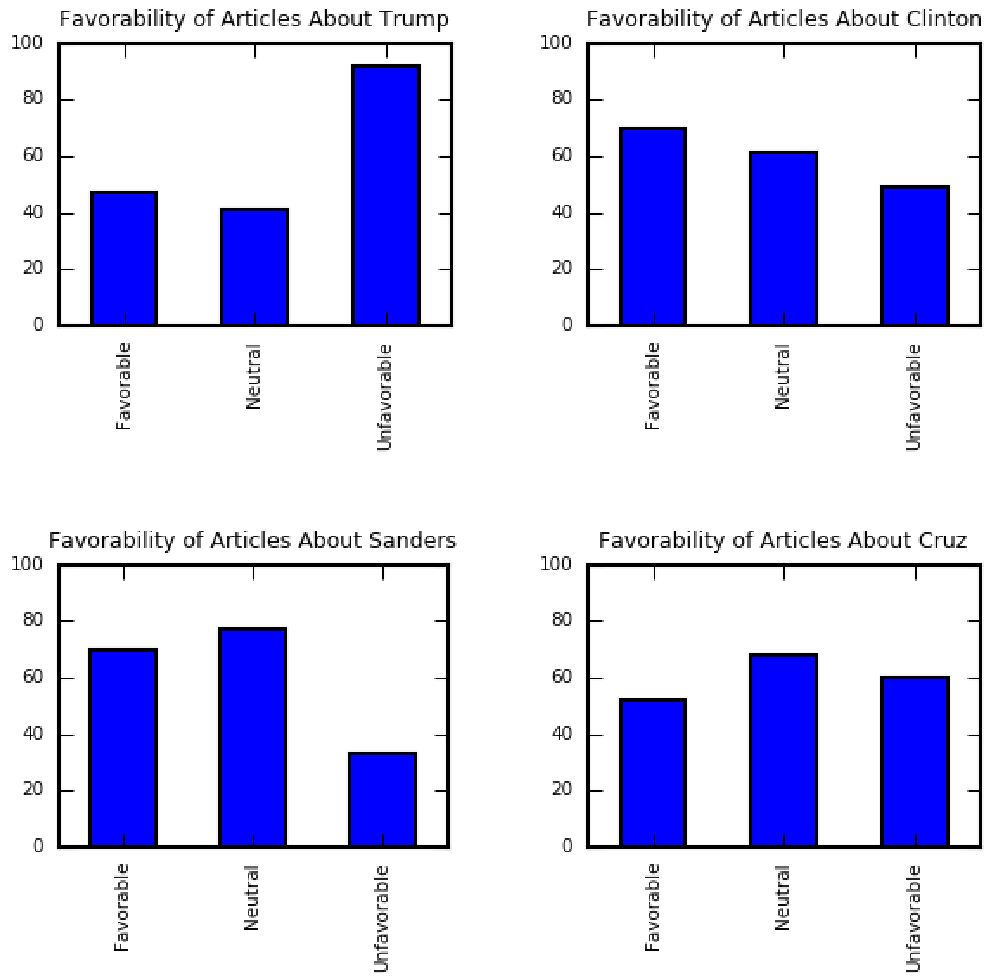

Figure 6-3: Media Favorability of Candidates, Trustworthy Articles

In the following section, we examine more patterns of media trustworthiness.

6.3 Media Trustworthiness

Each reader was also asked to score the five stories according to how trustworthy they found each to be.

Scores were collected on a five-point (Likert) scale: Strongly Agree (2), Agree (1), Neutral (0), Disagree (-1), Strongly Disagree (-2). Overall, readers seldom slected


I find this article trustworthy.
- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

Figure 6-4: Example of trustworthiness scoring question

"Strongly Disagree", and the option consisted of less than 2% of all choices.

In the analysis below, we collapse the results into three categories: Agree (> 0), Neutral (0), and Disagree (<0).

Despite reportings on national distrust of news, the majority of stories were marked as trustworthy for all candidates.

Sanders has strongest trustworthiness, most favorable

6.4 Overall Bias Reportings

6.5 Media Brand Effect

6.6 Reading Level Effect

6.7 Other Linguistic Cues

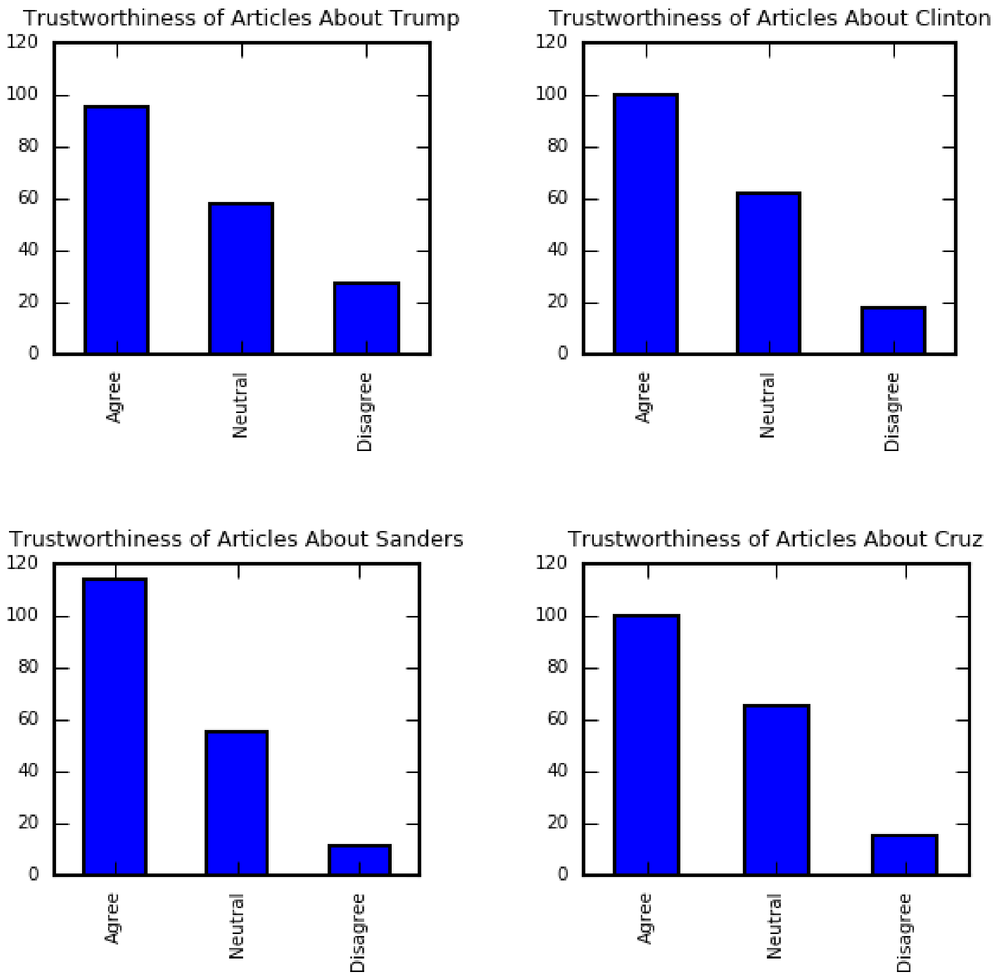

Figure 6-5: Media Trustworthiness of Candidate Coverage

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 7

Study Two

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 8

Tables

Table 8.1: Armadillos

Armadillos	are
our	friends

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 9

Figures

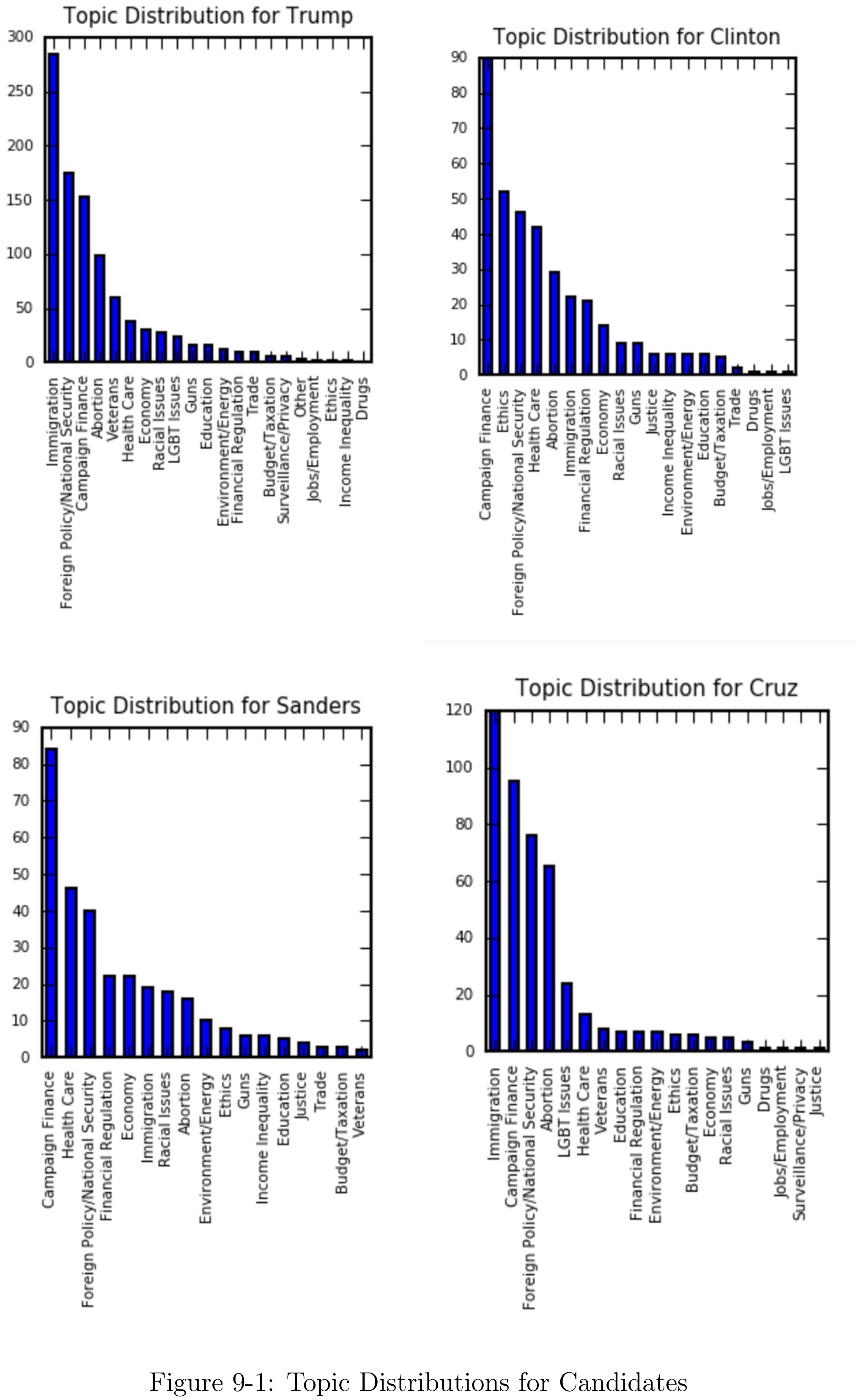

Figure 9-1: Topic Distributions for Candidates

Bibliography

[1] March 1 election results 2016 updates - the washington post. https://www.washingtonpost.com/2016-election-results/super-tuesday/. (Accessed on 04/06/2016).

[2] Measuring donald trumpâĂŹs mammoth advantage in free media - the new york times. http://www.nytimes.com/2016/03/16/upshot/measuring-donald-trumps-mammoth-advantage-in-free-media.html. (Accessed on 04/20/2016).

[3] Beyond distrust: How americans view their government. *Pew Research Center*, nov 2015. (Accessed on 04/07/2016).

[4] Matthew A Baum, Phil Gussin, et al. In the eye of the beholder: How information shortcuts shape individual perceptions of bias in the media. *Quarterly Journal of political science*, 3(1):1–31, 2008.

[5] J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, DTIC Document, 1975.

[6] Jonathan McDonald Ladd. Attitudes toward the news media and voting behavior. In *Annual Meeting of the Midwest Political Science Association*, 2005.

[7] The Florida Legislature. The 2015 florida statutes title xxxvii. http://www.leg.state.fl.us/Statutes. (Accessed on 04/05/2016).

[8] Amy Mitchell, Jeffrey Gottfried, Jocelyn Kiley, and Katerina Eva Matsa. Political polarization & media habits. *Pew Research Center*, Oct 2014. (Accessed on 04/02/2016).

[9] Kamal Nigam, John Lafferty, and Andrew McCallum. Using maximum entropy for text classification. In *IJCAI-99 workshop on machine learning for information filtering*, volume 1, pages 61–67, 1999.

[10] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830, 2011.

[11] Rebecca Riffkin. Americans' trust in media remains at historical low. `http://www.gallup.com/poll/185927/americans-trust-media-remains-historical-low.aspx`. (Accessed on 04/20/2016).

[12] Robert P Vallone, Lee Ross, and Mark R Lepper. The hostile media phenomenon: biased perception and perceptions of media bias in coverage of the beirut massacre. *Journal of personality and social psychology*, 49(3):577, 1985.

[13] Prashanth Vijayaraghavan, Soroush Vosoughi, and Deb Roy. Automatic detection and categorization of election-related tweets. In *10th International AAAI Conference on Web and Social Media*, 2016.

[14] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pages 1096–1103. ACM, 2008.

[15] Deena Skolnick Weisberg, Frank C Keil, Joshua Goodstein, Elizabeth Rawson, and Jeremy R Gray. The seductive allure of neuroscience explanations. *Journal of cognitive neuroscience*, 20(3):470–477, 2008.