

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

© Massachusetts Institute of Technology 2016. All rights reserved.

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

1	Introduction	7
2	In Media We ... Trust?	9
2.1	The Role of the Reader	10
2.2	The Role of Media Brands	10
2.3	The Role of Language	10
2.3.1	Language and Politics	10
2.3.2	The Seductive Allure [... of Simple] Language	10
3	Blank : used to be electome stuff	11
4	Preliminary Study	13
4.1	Data Selection	13
4.1.1	Publication Selection	14
4.1.2	Topic Selection	14
4.1.3	Flesch Kincaid Cutoffs	15
4.1.4	Redaction of Stories	15
4.2	CrowdFlower	15
4.3	Demographic Survey	16
4.4	Political Affiliation Survey	16
4.5	Quality Assurance	16
5	Pre-Survey Analysis	17

5.1	Topic Analysis	17
5.2	Flesch-Kincaid Analysis	17
5.2.1	Comparisons to other Reading Level Tests	17
6	Exploratory Study	19
6.1	Reader Demographics	19
6.2	Media Favorability of Candidates	19
6.3	Media Trustworthiness	21
6.4	Overall Bias Reportings	22
6.5	Media Brand Effect	22
6.6	Reading Level Effect	22
6.7	Other Linguistic Cues	22
6.8	Limitations	22
7	Main Study	25
7.1	Experimental Design	25
7.2	Data Selection	26
7.3	Limitations	27
8	Future Work	29
9	Conclusion	31
10	Tables	33
11	Figures	35

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 [8].

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 [5]. 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 [8]. 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 [3].

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 [10, 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. This thesis hopes to shed new light on understanding what motivates readers' trust and distrust of news media.

Chapter 2

In Media We ... Trust?

Despite the news media ecosystem’s rapid evolution in the past decade, the question of fairness in reporting remains a valued one. Although counterarguments for subjective reporting exist (Glenn Greenwald, most famous for his coverage of whistleblower Edward Snowden’s leaks, said that “All journalism is a form of activism. Every journalistic choice necessarily embraces highly subjective assumptions—cultural, political or nationalistic—and serves the interests of one faction or another”), fair treatment of subjects and sources remain a central tenant to most publications [1].

But an attempt at fairness on the side the reporter is not always perceived in equal effect under the eyes of the reader. Presenting contradictory facts to a reader’s beliefs can even sometimes *strengthen* their oppositions to it, a concept known as “motivated skepticism” [9].

In this section, we explore the theories behind three main potential sources of media distrust: In addition to the characteristics of the reader, we examine the source of the story and its use of language.

2.1 The Role of the Reader

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

2.2 The Role of Media Brands

2.3 The Role of Language

2.3.1 Language and Politics

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

Chapter 3

Blank : used to be electome stuff

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4

Preliminary Study

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) [2].

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 programmatically 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 [6].

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 11 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

Exploratory Study

6.1 Motivation

What was the purpose of the study? What were the hypotheses?

6.2 Dataset

What was the dataset?

How did you get the dataset?

Why those sources?

6.3 Limitations

From our exploratory study, we were able to obtain a significant but weak effect between disclosing the source and the levels of trust marked by readers towards an

article.

We also observed trends that suggested an interaction between disclosing the source and the reading level of a story.

However, the study faced several limitations: first, we did not obtain enough samples to show a statistically significant result for interactions between source and reading level.

Furthermore, multiple levels of independent variables (ie: 5 levels for input source) made modeling complex and the results less clear.

The dataset was also unbalanced and sparse (ie, because of large numbers of input variables we did not have complete representation for each category, such as high, low, and mid-reading level stories for every outlet and topic). We tried to control for those factors by randomization, however it made more difficult to analyze specific correlations between source and trust.

To further explore the interaction between disclosing the source and the reading level of the story, we set up another crowdsourcing experiment on CrowdFlower, this time targeting this specific interaction, to see if there is a significant effect between the two, detailed in the following chapter.

Chapter 7

Main Study

Following the limitations and patterns from the preliminary study, we designed a follow-up study to verify interactions between news source and reading level.

What were your main hypotheses? H1 H2

7.1 Experimental Design

For the second study, our experiment was revised to have a 4 x 2 mixed-factorial design. In this study, reading level of articles and candidates featured in the articles were treated as within-subject variables, and the source of the story between-subjects.

	Source: None	Source: AP	Source: Fox	Source: CNN
High Reading Level	Clinton, Cruz, Sanders, Trump	Clinton, Cruz, Sanders, Trump	Clinton, Cruz, Sanders, Trump	Clinton, Cruz, Sanders, Trump
Low Reading Level	Clinton, Cruz, Sanders, Trump	Clinton, Cruz, Sanders, Trump	Clinton, Cruz, Sanders, Trump	Clinton, Cruz, Sanders, Trump

Table 7.1: Main Study Design

This time, we reduced the number of stories to $N=8$, and also changed reading level from a 3-level to 2-level variable (low, high) for clarity.

Most significantly, since we observed some significant effect from disclosing source to the reader in Study 1, we added a manipulation in this experiment to further study the effect of revealing the source:

Following Baum’s research in showing the effects of media brands and reader bias by manipulating reported brands, all eight stories in Study 2 were in fact written by the Associated Press, however, we manipulated the source shown to the reader [4]. In group A, readers were shown the headline and text of the story with no other context. In group B, readers were additionally shown that the story was from the Associated Press (true label). In groups C and D, readers were shown that the story was from CNN and Fox News, respectively.

This setup was created to eliminate some of the confounding effects from using stories from different sources (writing style, focus of content, slant, etc.), while directly observing the effect of revealing a specific source to the reader. The Associated Press was chosen as the source of the stories as it is the highest circulation newswire service in the United States, and has 14,000 members that use its content [7]. Notably, both CNN and Fox News publish content in full or part from the Associated Press, although the specific stories chosen had not been published in full by either to avoid bias.

We removed the favorability question from Study 1 (as the 3-point scale did not yield significant results), instead asking the reader more directly about media bias by ranking the fairness of the story on a 5-point Likert scale. The trustworthiness question from Study 1 was kept, also on the same 5-point Likert scale.

7.2 Data Selection

Eight stories were chosen for this study: two (high and low reading level) per candidate.

How do your trustworthiness findings line up with the findings from Pew surveys and prior work? What hypotheses did you verify from prior work?

7.3 Limitations

Just 8 stories I know

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 8

Future Work

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 9

Conclusion

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 10

Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 11

Figures

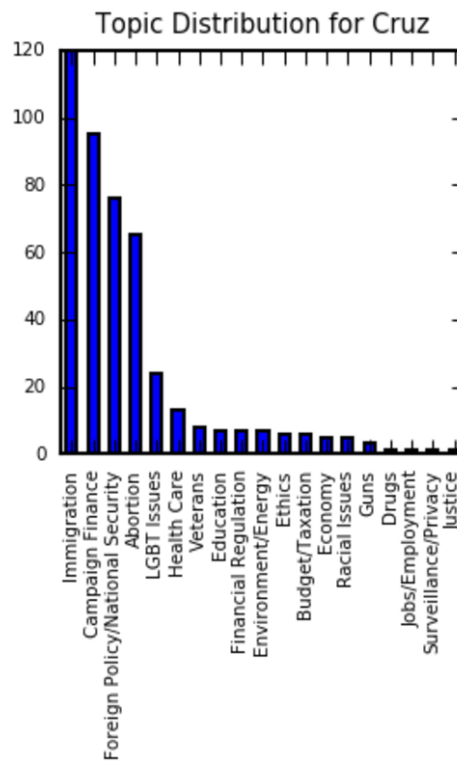
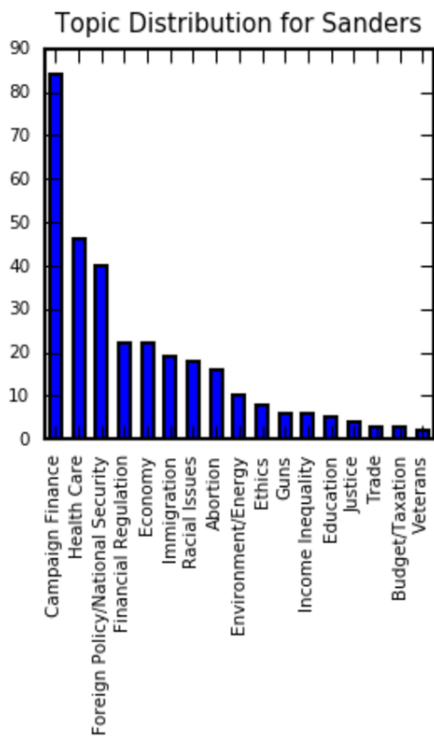
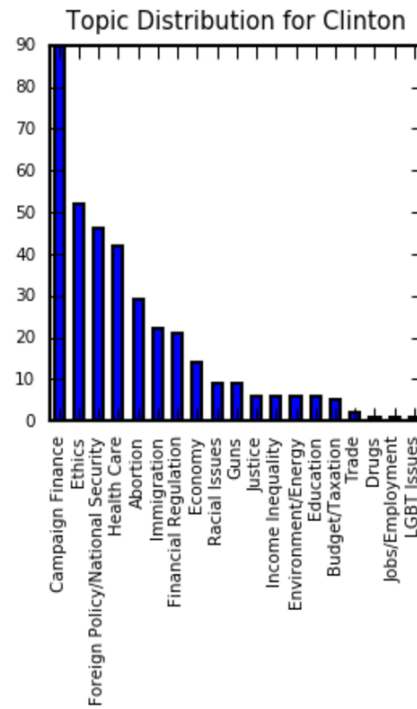
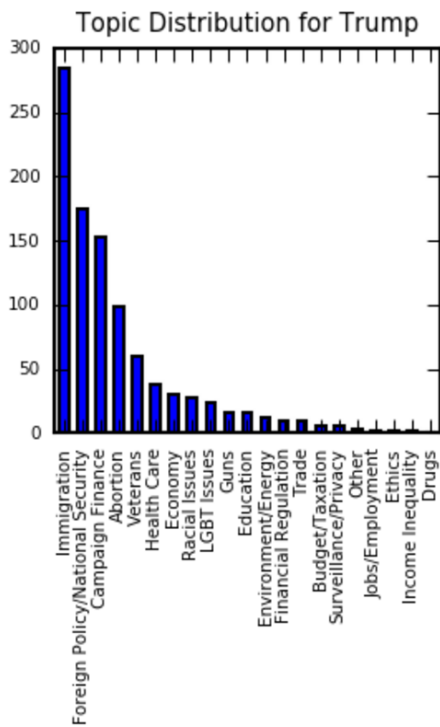


Figure 11-1: Topic Distributions for Candidates

Bibliography

- [1] Is glenn greenwald the future of news? – the new york times. <http://www.nytimes.com/2013/10/28/opinion/a-conversation-in-lieu-of-a-column.html?mtrref=www.google.com&gwh=450466A4B52C2124639CE220312812D5&gwt=pay&assetType=opinion>. (Accessed on 04/20/2016).
- [2] March 1 election results 2016 updates – the washington post. <https://www.washingtonpost.com/2016-election-results/super-tuesday/>. (Accessed on 04/06/2016).
- [3] Measuring donald trump’s 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).
- [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] Jonathan McDonald Ladd. Attitudes toward the news media and voting behavior. In *Annual Meeting of the Midwest Political Science Association*, 2005.
- [6] Amy Mitchell, Jeffrey Gottfried, Jocelyn Kiley, and Katerina Eva Matsa. Political polarization & media habits. *Pew Research Center*, Oct 2014. (Accessed on 04/02/2016).
- [7] Associated Press. Faqs. <http://www.ap.org/company/FAQs>, 2016. (Accessed on 04/21/2016).
- [8] 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).
- [9] Charles S Taber and Milton Lodge. Motivated skepticism in the evaluation of

- political beliefs. *American Journal of Political Science*, 50(3):755–769, 2006.
- [10] 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.