



**THE GOP DEBATES BEGIN  
(and other late summer 2015 findings on the presidential election conversation)  
September 29, 2015**

A PEORIA Project Report  
Associate Professors Michael Cornfield and Lara M. Brown  
Jamie Chandler, Data Scientist  
The Graduate School of Political Management ([GSPM](#))  
The George Washington University

*In Partnership with [Signal Labs](#)*

## **INTRODUCTION**

### What is the **PEORIA** Project?

Welcome (back) to the **PEORIA** Project, the GSPM's harnessing of Signal Labs's realtime, cross media story-tracking platform to analyze the "public echoes" arising from the 2016 presidential campaigns. While others during the "invisible primary" will investigate each candidate's poll standing, dollars raised and spent, and endorsements won, we track and measure **words**—the chatter about the candidates and the echo of their campaign messages in both mainstream and social media.

**PEORIA** is an acronym for **Public Echoes Of Rhetoric In America**, chosen as an allusion to the old vaudeville and marketing [phrase](#) "will it play in Peoria?"

Our fundamental premise is that how candidates and their messages play on the trail with the media and the public both affect and reflect the voters' presidential preferences. When a candidate says and stages it right, it resonates positively with the public, creating an echo that benefits the campaign. Of course, the opposite can also occur with negative echoes. From positive to negative, people respond to crafted messages,

brands, catch-phrases, sound bites, slogans, and gaffes<sup>1</sup> as they surface in news and social media, affecting their choices down the road.

The **PEORIA** Project follows the candidates and their campaign messages, measuring the public echoes that surface in all types of media.

### What does this third PEORIA report examine?

Our first two reports analyzed the public echoes during the period from March 15 to July 19, 2015, focusing on the formal presidential candidacy announcements and the initial branding attempts of these campaigns.

The third report focuses on the conversations surrounding the first two GOP debates (August 6 and September 16), along with the pre-debate jockeying among the Democratic candidates. The time frame from this report runs from July 20 through September 20, covering 62-days. Although we looked at all of the candidates on both sides of the aisle, most of our tables and charts present only the data for the 11 GOP candidates who were in the CNN prime time debate (we include Vice President Joe Biden who has frequently been discussed as possible Democratic candidate).

## **METHODOLOGY**

Here we present information on the Signal Labs platform and our own metrics as applied to the data the platform contains.

### A) Signal Labs

#### **What does the data universe contain?**

*SOCIAL MEDIA: Every single tweet, publicly available Facebook user posts, every single mention in social/online video (YouTube, Vimeo, MediaBistro), 30+ million blogs. NEWS and MAINSTREAM MEDIA: news stories from more than 100,000 online outlets including licensed content, all LexisNexis News Content (print news, magazines/journals/newspapers,etc), all television closed caption content from 900 channels in every media market in the US.*

#### **What counts as a mention?**

*Any tweet, news story, blog, video, LexisNexis story or broadcast clip (closed captioning) that matches a query (a query is a combination of certain keywords or phrases). For this project and related ones, Signal has built a custom database with*

---

<sup>1</sup> Of course, the public responds to images as well. We presume that any image which has a significant effect on candidate reputation and voter choice becomes a topic of discussion and acquires its own caption or summary title, e.g. "Dukakis in the tank" and "Bush looking at his watch." We, thus, pick up memorable images through the words that are commonly used to describe them.

*real-time continuous queries of the presidential candidates' names. Multiple mentions within a content unit or "document" are not counted extra.*

### **How is share of voice calculated?**

*Share of voice is calculated by summing up the mentions (across all media types) in each candidate profile and taking the ratio of that candidate's total to the entire set of candidates.*

### **How are sentiment classifications (positive, negative, and neutral) determined?**

*Sentiment is determined using natural language processing technology (NLP). Signal's NLP algorithm assigns a positive, negative, or neutral score to every document to provide an overall sentiment rating. Frequency, intensity, and sentence structure are factored into the model. For example, "love" has a higher score than "like", but an overall negative prediction will still occur if negations such as "not" or "neither/nor" are present within the sentence. Adverbs also serve as multipliers, with phrases like "very good" scoring higher than "good." The backbone of this algorithm is a Recursive Neural Tensor Network, a type of deep learning algorithm that allows us to continually modify and fine-tune our model as time goes on. Unfortunately, sentiment detection is still not an exact science, and NLP fares poorly when sarcasm is present or the overall diction is ambiguous.*

*Over the course of the project, GWU has the opportunity to manually override and/or correct sentiment which helps train and improve the models' performances. In addition, the project only reports "**net sentiment**," positive less negative or vice versa as the case may be. This move assumes that erroneous classifications are randomly distributed, and that the directionality of sentiment is a fairer albeit thinner indicator than reporting percentages from all three categories.*

### **How are other indicators determined?**

*Popular Tweets: number of retweets that a tweet gets*

*Top Issues: a second/third level of filtering. Profile queries (the candidates' names) control what gets ingested into the platform, and issues are tags that categorize the data ingested. Top issues is thus a sorted list of the most frequent tags by candidate.*

## FINDINGS

- **(See Slide #5.) Between July 20 and September 20, Donald Trump has continued to dominate the presidential campaign conversation.** His substantial volume of mentions is largely the reason that the Republican presidential conversation has been on average about three times as “loud” as the Democratic presidential conversation (see Slide #4). Also contributing to the party imbalance by volume: there are more GOP candidates, and there were two GOP debates to none featuring the Democrats.
  - Trump’s dominance is all the more impressive when one considers his average daily mentions versus the other top ten candidates in the contest (see Slide #5).
- **(see Slide #6) Bernie Sanders surpassed Hillary Clinton in terms of number of mentions during the second Republican debate on CNN (on September 16).** His “live-Tweeting” clearly brought him attention, and as will be seen later (see Slide #24), Sanders most successful Tweet that evening garnered him nearly twice as many ReTweets as Hillary Clinton’s most successful Tweet (11.7K versus 6.3K).
- **(see Slide #7) Trump lost Republican Share of Voice (percentage of total GOP mentions) in the aftermath of the second Republican debate on CNN.**
- **(see Slide #11) Carly Fiorina gained the most proportionately over the entire period.** Prior to either debate, she was commanding only about 1.8% share of voice (near the bottom); in between the debates, her share increased to 3.9% (and put her in the middle of the pack); and after the second debate, her share increased to 12.6% (second only to Trump). No other candidate has (not even Ben Carson who moved from 1.8% to 4.8% to 6.5%) moved as far as fast.
- **(see Slide #11) Scott Walker experienced the fastest decline.** He began with having 5.4% share of voice prior to either debate; in between his share of voice fell to 3.8%; and after the second debate, his share of voice dropped to 2.9%.
- **(see Slide #8) Bernie Sanders overtook Clinton during the post-Labor Day period.** In addition, Joe Biden’s share of voice since Labor Day has also taken a few ticks upwards (from 5.5% to 6.7%).
- **(see Slide #9) Mainstream media have been more equitable in their distribution of mentions, and mentioned Bush and Walker more than social media. Social media have talked about Trump and Cruz more than mainstream media.**
- **(see Slide #10) Mainstream media have been more focused on Clinton and Biden. Social media have talked about Sanders nearly twice as much as**

have mainstream media. In social media, Sanders runs about even with Clinton.

- **(see Slide #12) Net sentiment over the time period has not only changed, but appears to have taken a turn towards the positive.** We suspect three reasons for this change, two of which have to do with the historical timing (September 11 and Constitution Day are commemorative occasions) and one which may have to do with the political circumstances (as “non-establishment” candidates succeed, the public may be more satisfied with the contest), though we need to investigate further to understand what has happened lately. We also plan to watch this trend closely to see if it continues.
- **(see Slide #13) While Trump has won the volume war, he remains in the middle of the pack when it comes to the battle for sentiment.** His net sentiment is slightly negative (-4.6%). The net sentiment rating for both Carson and Fiorina are much more positive (respectively, 19.6% and 17.8%). Further, the more “establishment” candidates have fared much worse. John Kasich and Marco Rubio are the only two elective officeholders whose net sentiment remains on the positive side of the ledger.
- **(see Slide #14) Sanders is the only Democrat whose net sentiment is positive.**

## CONCLUSIONS

There was three times as much talk about Republicans as Democrats in this time period. We do not see this stemming from bias; indeed the talk about Republicans, like the smaller volume of talk about Democrats, divides into positive and negative. Instead, the larger volume about the GOP results from the facts that there were three times as many Republican candidates, two Republican debates to none for the Democrats, and the presence of Donald Trump, a conversational magnet.

There may, however, be ideological consequences to this imbalance, if the agenda and tone set in this time period turns out to have staying power and leaves an imprint on the campaigns when voting commences. The Republican debates this summer focused on foreign threats and religious liberties with not as much about socioeconomic conditions and inequalities in particular. The Democrats may talk more about the latter when they stage their debates. We will see if the conversation forks, bends, or otherwise shape-shifts when they start on October 13.