A Tale of Two Twitterspheres:

Political Microblogging During and After the 2016 Primary and Presidential Debates

Ron Berman

Shiri Melumad

Colman Humphrey

Robert Meyer*

* Authors are listed alphabetically; all contributed equally to the research. Ron Berman and Shiri Melumad are Assistant Professors of Marketing at the Wharton School, University of Pennsylvania, Colman Humphrey is Data Scientist at Via Transportation, and Robert Meyer is the Frederick Ecker/Metlife Professor of Marketing at the Wharton School. The Authors are indebted to Shawn Zamechek of the IT Staff at the Wharton School for assistance in data collection, and Yoon Duk Kim of Wharton for assistance in data analysis, We thank Jonah Berger for comments and suggestions on an earlier draft of this manuscript.

Abstract

In this research we study the process by which social media posts are created and shared during live political debates. Using data from over 9.5 million Tweets posted during and shortly after four key debates leading up to the 2016 Presidential election, we test a series of hypotheses about how Tweeting evolves over time during such events. Among our findings are that as debates progressed the content of the "Twittersphere" became increasingly decoupled from the live event, and that the drivers of the success of Tweets during the debates differed from those observed after. During the debates users acted akin to narrators, posting shorter Tweets that commented on unfolding events, with linguistic emotionality playing a limited role in sharing. But when the debates were over users acted more like interpreters, with successful posts being more elaborate and visually and emotionally rich accounts of the event. Evidence for the generalizability of the findings is provided by an analysis of Barack Obama's last State-of-the-Union Address, where similar dynamics are observed.

“Hombres”

The most shared Tweet during the 2016 Third Presidential Debate

Introduction

Since their launch in the early 2000s microblogging sites such as Twitter have played an increasingly important role in political marketing. Whereas marketing efforts by candidates once focused on television spots, billboards, and stump speeches, today social media serves as a central communication medium for many election campaigns (Stromer-Galley 2014). In 2016, for example, users posted over one billion Tweets about the U.S. Presidential election (Levy 2016), something that led *New Yorker* writer Nathan Heller (2016) to call the campaign the country’s first “Twitter Election”—one where the battle for the “Twittersphere” loomed almost as large as the election itself. Increasingly, the most successful candidates are not necessarily those with the most reasoned positions on issues, but rather those whose ideas emerge most clearly from the cacophony of 140-character sound bites about an election created and shared by voters, the news media, and, of course, by the candidates themselves.

What motivates people to Tweet about political events, and what factors drive the popularity of postings? In this research we contribute to an understanding of these questions by analyzing Tweeting patterns during and immediately after a series of pivotal political events that preceded Donald Trump’s eventual election as president: The Republican party (GOP) primary debates as well as the third debate between Trump and the Democratic candidate Hillary Clinton. We study Tweeting in the context of debates for both substantive and theoretical reasons. First, debates form an important focus of study because public perception of candidate performance during these events can have a major influence on election outcomes, and Twitter is increasingly looked to as a bellwether of that perception (e.g., McGregor and Molyneux 2018; McInney and

Warner 2013; Zheng and Shahin 2018). As such, an understanding of the factors that drive the popularity of Tweets can potentially offer tactical guidance for how to gain visibility on Twitter during these events.

Second, developing a richer understanding of political microblogging bears implications for marketers that extend beyond debates and elections. There is growing evidence that microblogging about political topics often serves to activate the political identities of both consumers and generators of the content (e.g., Conover, et al. 2011; Sylwester and Purver 2015), and these identities play an important role in influencing a wide range of consumer consumption behaviors (e.g., Kim, et al. 2018; Kashmiri and Mahajan 2017; Winterich, et al. 2016), such as the types of products they purchase (Ordabayeva and Fernandes 2018) and the types of financial risk they are willing to undergo (Han et al. 2019). As such, understanding what factors tend to motivate consumers to post Tweets during (and after) debates may enable marketers and campaign managers alike to more effectively leverage and interpret this medium in the context of live political (and possibly non-political) events. Indeed, we show that our findings generalize to various political domains, ranging from the 2016 debates to Barack Obama's final State-of-the-Union address.

Using a battery of automated natural-language processing tools we emerge with two central findings. The first is that if there is a "recipe" for designing widely shareable Tweets, it is one that critically depends on timing. During debates—when the competition for Tweets is most intense and viewer attention most divided between the live event and Twitter feeds—we find that Twitter acts much like a play-by-play news medium with users acting as *narrators*. Here, both the Tweets that are created and those that are most shared tend to be shorter, descriptive accounts that reference the discussions unfolding onscreen. The play-by-play, however, is a noisy one; as the debates progress the content of the Twittersphere becomes increasingly decoupled from the live event, such that, at times, it is more a topical free-for-all that competes for attention from the event rather than complementing it. Our second major finding is that these drivers change dramatically once the debates are over. In the immediate wake of the debates, when feeds no longer compete with the debate itself for attention, Twitter acts more like an *interpretative* medium, one where the Tweets that are created and are most often shared are those that discuss events in greater detail, use words suggestive of a focus on overall success or achievement, and include emotional language and visually rich content. There is also a surprising shift in the substance of what is discussed. The 2016 debates were each marked by a series of particularly contentious exchanges involving Donald Trump, and it was these exchanges---not the positions on policy widely Tweeted about during the debate---that were the greater focus once the debates were over.

Our research offers two sets of contributions to the field of political marketing. First, our work contributes to a growing literature that analyzes the effectiveness of different media channels and communication styles during elections (e.g., Gordon and Hartmann 2013; Hoegg and Lewis 2011; Jung and Critcher 2018; Phillips et al. 2008). Here we offer one of the first detailed longitudinal analyses of how the features of microblogs evolve over time during and after live political debates, an analysis that could offer candidates, bloggers, and firms insights into how to gain greater visibility during such live events. Second, while set in the context of presidential debates, we see the work as contributing to the broader study of what causes consumers to create and share ideas about politics over social media---something that, as noted, can have downstream effects on consumption behavior as well outside politics (e.g., Han, et al. 2019; Ordabayeva and Fernandes 2018).

We organize our paper in three sections. We first review prior work on the factors that drive microblogging activity in political and other contexts. We then report the focus of our analysis: a longitudinal study of Tweeting behavior dynamics over the course of the debates, as well as the factors that drove the staying power of popular retweets during—versus after—the debate. We conclude with a discussion of the implications of the research for understanding the role of new media in the electoral process, as well as the more general topic of the drivers of long-term contagion.

Debates, Political Tweeting, and Virality

Televised debates have long played a central role in electoral processes both in the United States and abroad (e.g., McInney and Warner 2013; White 1961). But whereas voters were once left to draw conclusions about candidate solely by their on-stage performance, social media sites such as Twitter have increasingly created two co-evolving contests: that being waged among the candidates on stage, and that being waged among microbloggers in the "Twittersphere" (Trilling 2015). It is in the Twittersphere where viewers, pundits, celebrities—and even the candidates—contribute to what Shamma et al. (2009, 2010) term the "community annotation" of events: a forum wherein voters voice opinions about comments made by candidates in real time, react to the opinions expressed by others, and, possibly, compete to have the most widely heard opinions. As such, the Twittersphere has been seen as offering a mirror of collective public sentiment toward the candidates (Zheng and Shahin 2018).

What motivates people to create and share Tweets during debates, and how might these forces vary over time as a debate progresses? As shown in Table 1, in recent years a large literature on political microblogging has emerged in computer and political science that has studied different aspects of Twitter use during election periods. One of the major interests in this work, for example, has been to explore whether one can predict the outcome of elections from the positivity of sentiment expressed toward candidates in Tweets (e.g., McKelvey et al. 2014; Mejova, et al. 2013; Ramteke, et al. 2016). In contrast, empirical studies of what drives the creation and sharing of Tweeting over time during live debates have been more limited. One exception is an early study of Tweeting during a 2008 presidential debate between Barack Obama and John McCain by Shamma, et al. (2009), who, like here, studied how Tweeting varied over time during the event. One finding was that the content of Tweets often strayed from the major political issues at the time, however the low volume of Tweeting observed during the debate (where there was almost no retweeting) precluded the authors from drawing inferences about what drove the popularity of Tweets, or how that may have varied over time.

[Insert Table 1 about here]

Although studies of Tweeting during debates *per se* may be limited, the broader problem of why users create and share online content has been widely studied in marketing and other fields (e.g., Berger and Milkman 2011; Guadagno et al. 2013; Hoang et al. 2013; Harber et al. 2005; Heath 1996). One dominant finding, for example, is that users are more likely to share content that they find emotionally arousing, such as humorous articles (e.g., Berger and Milkman 2012) or emotionally rich videos (Guadagno et al. 2013). Similarly, empirical studies of retweet rates in other political contexts (see Table 1) have similarly found that Tweets with proportionally more emotional words—either positive or negative—are more likely to go viral (e.g., Ferrera and Yang 2015; Hoang et al. 2013; Kanavos et al. 2014; Stieglitz and Dang-Xuan 2013). As such, one might conjecture that Tweets during a debate that conveyed greater emotion—such as anger, happiness, or humor—are not only more likely to be created, but will also see higher retweet rates than more dispassionate descriptive reporting or opinion.

But live-debate Tweeting has a number of unique features that may limit the generalizability of prior findings about online sharing to the present context. Perhaps the most important is a difference in why users contribute: rather than generating or sharing content of general interest, during debates users are commenting on an unfolding live event, offering personal views of exchanges between candidates that seem exceptionally notable. Some grounding for this motivational difference can be found in past survey research on how and why people Tweet during debates by Bramlett et al. (2017), Houston et al. (2013), and Maruyama et al. (2014), who found that, for some, Tweeting enhances feelings of participation in the event, heightening engagement with the live proceedings. As such, during debates Tweeters may see their role as narrators, keeping one eye on the proceedings, and the other on their Twitter feeds, ready to create or retweet content when a motivating event arises.

If this characterization of live Tweeting is correct, it holds several implications both for what Tweets will look like during events and the kind of Tweets will be most shared. For example, if speed of creation and relevance are what matter most during debates, Tweets that offer succinct and timely observations about candidate remarks would tend to prevail over more abstract and interpretive compositions that are less tied to the events. During debates this would be evidenced by a prevalence of shorter posts that relay quotes of what a candidate had just said, comments on the substantive issues that are being discussed, and a comparative dearth of graphic content that would be time-consuming to produce and view. Likewise, if descriptive narration is the primary goal of Tweeting, the success of a Tweet would depend less on factors that have been found to be important in driving virality in other settings, such as the use of words that evoke emotional arousal (e.g., Berger and Milkman 2012), or, in the present setting, that offer evaluation or analysis.

We can summarize this idea in terms of the following hypothesis:

H1: During (vs. after) a political debate Tweeting will be dominated by brief descriptive accounts of the exchanges among candidates, with sharing being less dependent on the ability of a Tweet to communicate user's subjective interpretations or emotionality.

If live-debate Tweeting indeed takes the form of narration, one might predict that there should also be a close temporal alignment between the topical content of Tweets at any point in time and the contemporaneous debate transcript. Working against this conjecture, however, is that as debates progress, newly-created content will face growing competition from an ever-expanding stock of past Tweets and past debate events, making it increasingly difficult for new content to gain visibility. Hence, rather than serving as a true "community annotation" of the unfolding debate (Shamma et al. 2010), as debates progress the Twittersphere may more resemble a topical free-for-all, coming back in alignment with the debate transcript only occasionally when a candidate makes an attention-grabbing remark. We summarize this idea through the following hypothesis:

H2: The temporal alignment between Tweets and the debate transcript will decay as the debate proceeds, such that Tweets will be increasingly decoupled from the issues being discussed by candidates.

Post-debate Tweeting

Above we hypothesized that during debates the pressure to comment on events in real-time would foster the creation and sharing of shorter, largely descriptive postings (H1). After the debate, however, the context of Tweeting fundamentally changes; rather than needing to comment on a fast-moving event, users now have the opportunity to reflect on the event in hindsight, and can possibly afford to be more thoughtful in the content that is created. Hence, users may see their roles as shifting from that of narrators to that of interpreters, with both the Tweets that are created and those that are most shared being more linguistically and visually elaborate. Consistent with this, we expect to see a shift from concrete description to abstract interpretation, with the most popular Tweets being those that offer emotional expressions of who "won" or "lost." Similarly, retweets after the debate may selectively focus on what users saw as the highlights of the debate, thereby bringing enduring importance to exchanges that may have been only briefly mentioned during the debate itself. As such, if Tweets during the debate reflect how viewers experience it, Tweets posted afterward reflect—and potentially shape—how viewers recall it. We can summarize these ideas through the following hypothesis:

H3: Content that is created after the debate (vs. during) will display greater linguistic and visual elaboration, and sharing will be increasingly influenced by the ability of a Tweet to convey interpretation and emotion.

Finally, one of the major—and perhaps unsurprising—findings of prior empirical studies of drivers of retweet rates is that the prominence of the Tweeter matters: Tweets posted by users with larger followings reach more users, and are hence retweeted more (e.g., Suh et al. 2010). A unique feature of live events that may temper this effect, however, is that during debates both broadcasters and Twitter encourage users to Tweet to a common community hashtag, such as "#debate". This encouragement democratizes the forum because, in principle, Tweets posted by users with fewer followers now have similar exposure to the Twitter feed as those from celebrities with millions of followers. Hence, when the number of people Tweeting to a common hashtag is particularly high, having a large following may be less influential than how quickly the Tweet is posted, for example. But after the debate is over—when the pace of Tweeting slows and the emphasis is less on narration and more on interpretation---the prominence of the Tweeter may again assume greater importance as a driver. We can summarize this in terms of the following hypothesis:

H4: After a debate, the size of a Tweeter's following will play an enhanced role in driving retweet rates compared to that observed during a debate.

To test these hypotheses, below we report the results of empirical analyses of the temporal dynamics of Twitter activity during versus after a series of major political events---debates preceding the 2016 Presidential elections.

Empirical Analysis

Institutional Background: Debates and the 2016 Presidential Election.

The 2016 election campaign was a historic one in many respects. After suffering defeats in two successive presidential elections, the Republican Party (GOP) felt that they could reclaim the White House given a strong candidate, and by the summer of 2015 no less than seventeen had announced their candidacy (Coppins 2016). Notable among them was businessman, television personality, and self-proclaimed billionaire Donald Trump, who hoped to ride a recent wave of populism to the presidency by promising to "make America great again." The Democratic candidacy, in contrast, was less contested: Hillary Clinton was anticipated to be the party's eventual nominee, and while she was formidably challenged by Bernie Sanders, by early 2016 she was well on her way to the nomination.

Because of the contested Republican nomination, the party held a series of debates that received extensive media attention. The first debate, held on August 6th, 2015, was the most watched non-presidential debate in history, with over 24 million getting their first chance to see Donald Trump debate the other candidates. After the August event, eleven more debates led up to the GOP convention, with the pool of participants gradually dwindling. By March only four candidates remained—Ted Cruz, John Kasich, Marco Rubio, and Donald Trump, and on March third the final debate was held, which was widely seen as the last chance to derail Trump's path to the nomination. Given its importance, it was almost as widely watched as the inaugural GOP debate, drawing over 17 million television viewers (Flores 2016).

Following their nomination to compete in the upcoming presidential election, Hillary Clinton and Donald Trump agreed to three debates prior to the election. These debates drew over 220 million viewers combined, with the first and third debates being the most-ever watched in U.S. history. They also set records for Tweeting, with each drawing over 17 million Twitter interactions during the debates, and millions more before and after (Twitter 2016).

The data

Our data were comprised of all unprotected Tweets posted to Twitter's primary debate hashtags (e.g., #GOPDebate for Republican primaries, #Debate, #DebateNight for the Presidential) from sixty minutes before to ninety minutes after four debates in 2015 and 2016: the inaugural Republican party debate of August 6th, 2015, the eleventh GOP debate on February 25th, 2016, the twelfth GOP debate on March 6th, 2016, and the third Presidential debate on October 19th, 2016.[1] Reflecting the different audience sizes for the debates, the August dataset was comprised of 1,653,158 Tweets, February 955,549, March 896,680, and the October Presidential debate 6,005,044 Tweets. In addition to the Tweets, we also retrieved from public news sources the complete transcripts of each of the debates (see Web Appendix 2 for transcript source).

Each Tweet record contains the text of the Tweet, a timestamp, an indicator for being an original Tweet or a retweet, and the cumulative number of retweets received until the end of the data record. Each record also included information about the Tweeter, specifically, the user's

[1] We thank GNIP (Twitter's social media aggregator) for the data.

Twitter handle (preferred username), the number of accounts that followed the user ("followers"), and the number of accounts the user followed ("friends"). We also added two measures of how often each user Tweeted about the debates within our data: the number of original Tweets created immediately before, during, and after the debates, and the number of retweets posted before, during, and after.

Pre-processing: measuring Tweet and transcript content.

Our central interest was in testing a series of hypotheses about the dynamics of Tweeting during and after debates. We hypothesized, for example, that the drivers of Tweet popularity would differ during-versus-after the debate, with shorter, more narrative Tweets dominating the Twittersphere during the debate (H1), but more elaborative and interpretative Tweets dominating after (H3). To map content to measures that correspond to these hypothesized dimensions, we subjected each Tweet and spoken line of transcript[2] to two sets of automated text analyses, one focusing on the linguistic and surface feature of the text, and the other its substantive content.

Linguistic and surface features. Each Tweet and transcript line were first culled of extraneous characters (e.g., hashtags), and then subjected to analysis by three different automated natural-language processing tools:

LIWC (Linguistic Inquiry and Word Count; Pennebaker, et al. 2015). LIWC is a widely used text analysis tool that provides 90 measures of the grammatical and stylistic content of text (marketing applications include Berger and Milkman 2012; Ludwig et al. 2013; Melumad, Inman and Pham 2018; applications to the analysis of Twitter data include Eichstaedt et al. 2015). While a large number of linguistic predictors could have been studied,

[2] A spoken line was defined as an excerpt by one speaker. Lines thus varied in length, from brief interruptions to more extended passages.

we focused on three sets of linguistic measures that would allow us to test for the predicted change from descriptive to interpretive language within the context of debates (H1 and H3). These were: 1) the use of positive versus negative emotional words in the Tweet; 2) the use of words referencing achievement (e.g., “won, “better”), power (e.g., “superior”, “bully”) and reward (e.g., “prize”, “reward”); and 3) composite indices of analytical and authentic writing styles. These latter indices are measured on a 1-100 scale, with a more analytical style capturing a more logical and less narrative writing style, and a more authentic style capturing a more casual and disclosing style (Pennebaker, et al. 2015). We predicted that, if Tweets posted after the debate were indeed more interpretive than they were descriptive (H3), then Tweets posted *after* the debate should include relatively more words related to achievement, power and reward, greater emotionality, greater use of analytic or logical writing styles, and less use of authentic styles.

Speciteller. To provide an additional measure of linguistic elaboration we subjected each Tweet to analysis by Speciteller, a machine-learning algorithm developed by Li and Nenkova (2015) that measures the degree of linguistic elaboration in text. Speciteller generates predictions of the probability that human judges would perceive a given text as being specific (vs. gist-like), where higher specificity scores result from longer texts that make greater use of references and elaborative language (Li and Nenkova 2015).[3] We hypothesized that Tweets during the debate would display lower specificity scores than those posted after (H1, H3).

[3] For example, Speciteller would assign the short Tweet, “Did he really say ‘Bad Hombres’?” a specificity score of .02, while the longer, “Trump: Your husband disagrees with you..Oh nooooooo! Not a wife who disagrees with her husband?!?! CALL THE NATIONAL GUARD!” would be assigned a score of .91.

Visual content. We analyzed the link contained within a given Tweet text to identify the presence of visual content. We separately recorded whether a Tweet contained a video, photo, or animated .gif.

Topical Content. In addition to linguistic features, we also measured the topical content of the Tweets. If, as hypothesized, users act more like narrators during the events and interpreters afterward, we should see users making frequent references to the concrete policy topics being discussed by the candidates (e.g., Obamacare) during the debates, but for these references to decrease afterward when the focus of Tweeting shifts to more abstract and holistic assessments.

A set of policy-related keywords were derived using both manual and automated methods. First, the original debate transcripts were read by a research assistant and one of the authors, who developed an initial list of policy topics and themes that were referenced at least twice during the debates (e.g., Obamacare, abortion, need to unify). This list was then augmented by words suggested by an automated keyword analysis of Tweets (e.g., Beliga 2016), that compared the frequency with which words were used in the transcripts to their relative frequency in Google's Trillion Word Corpus (available at https://github.com/first20hours/google-10000-english; Accessed Sep 28 2017).[4] Judgment was then used to collapse the master set of keywords to smaller clusters of twelve topical categories that reflected either substantive policy domains (e.g, Asia, defense spending) or abstract themes (e.g., promises of reform, political correctness) discussed by candidates during

[4] A word that emerges as unusually popular may be indicative of the topic of discussion in a given text body, hence it might be a potentially useful keyword. For example, although the word "news" was the 163rd most popular in our data, it is also the 59th most popular word in Google's corpus, and hence is not "unusually popular". In contrast, the word "immigration" was the 145th most popular word in the August debate, but its rank in Google's corpus is 4381, making it unusually popular.

the debates. The complete set of topical categories and illustrative keywords is reported in Web Appendix 3.

In addition to these policy topics, we also measured whether a Tweet made reference to either the physical appearance of the candidate (e.g., “hair”, “tie”, “suit”) or one of eight contentious exchanges involving Donald Trump that arose during the debates (e.g. calling Hillary Clinton a “Nasty Woman”; see Web Appendix 3 for descriptions of the exchanges and sample keywords). Because we identified these latter events using post-debate news coverage—which may have been informed, at least in part, by Tweeting activity--our interest was not in the level of Tweet volume associated with each, which we expected to be high. Our interest, rather, was in the time course of this volume: whether interest in these events peaked when they first occurred during the debate when users were narrating the proceedings, or after the debate when they were interpreting it in hindsight.

Were the Tweeters robots? As a final step in pre-processing we explored whether the Tweets might be contaminated by posts from automated accounts (bots), something known to have been prevalent in the 2016 campaigns season (Kollanyi et al. 2016). To explore this possibility, we subjected the usernames in our dataset to analysis by a machine-learning algorithm for robotic detection developed by Varol et al. (2017), which uses 1150 features of Tweeting behavior by a user to compute a probability of the account belonging to a non-human user (see Varol et al. 2017 for a discussion of validation). The results of this analysis, reported in Web Appendix 4, reject an influential effect of automated Tweeting. While some non-human accounts may well have been generating content during the debates, bot likelihoods were not statistically associated either to retweet counts or the physical features of Tweets. For example, the correlation between the probability that a user was a bot and Tweet’s popularity was -.008, and that between bot probability and Donald Trump (the most controversial candidate) was .003---neither approaching statistical significance despite extremely large samples.

Results

Overview

We divide our analysis into three sections. We begin with a brief overview of how the volume of Tweets and retweets evolved over time during the debates, and then report a set of findings related to our second hypothesis (H2), the relationship between the content of Tweets and the time course of the debate transcript. We then focus on tests of H1, H3, and H4, how the nature of Tweeting differed during versus after the debates. In this analysis we first discuss changes in the nature of the Tweets that were being created, and then the drivers of retweet rates.

General features of Tweeting and retweeting

In Figure 1 we plot the total number of Tweets and retweets that were generated per minute before, during, and after the two most widely watched debates in our study: the August 2015 GOP debate and the October 2016 presidential debate. The figure illustrates the extremely high rates of Tweeting that marked the events. At its peak during the October presidential debate, for example, combined Tweets and retweets were being generated at a rate of over 70,000 per minute, and even an hour after it was over Tweets continued to be generated at almost 20,000 per minute.

[Insert Figure 1 about here]

More importantly, the figure shows that as the debates wore on there was a decided change in the *kind* of Tweets that dominated the Twittersphere. Whereas early in the debates Twitter feeds were equally populated by new posts and retweets, as the debates progressed the

Twittersphere was increasingly dominated by retweets of older content, a trend that accelerated after the debate was over. After the conclusion of the presidential debate, for example, 84% of content took the form of retweets of Tweets that had first been posted during the debate. Hence, rather than offering a fresh retrospective commentary on what had happened during the debate, the post-debate Twittersphere more resembled a highlight reel, with users preferring to retweet posts that, perhaps, they saw as best capturing the key moments in, and comments about, the debate. The features that marked these popular post-debate Tweets will be discussed below when we report our analysis of drivers of retweet rates during versus after the debate.

Were users narrating during the debate?

Prior work on live-event Tweeting had conjectured that the Twittersphere might act as a "community annotation" of the event---a live running commentary on the proceedings as they transpire (e.g., Shamma et. al 2009, 2010). We conjectured in our second hypothesis (H2), however, that this annotation is likely to be highly noisy, with the content of Tweets becoming increasingly decoupled from the live debate as it unfolded. Indeed, the growing prominence of retweets (vs. new original Tweets) shown in Figure 1 gives some initial credence to this idea: near the end of the debate the majority of Tweets that users would have been seeing would have been of the form of reposts of older material.

To explore this idea more rigorously, we subjected the linguistic content of the Tweets and the debate transcript over time to analysis by *doc2vec* (Le and Mikolov 2014), a machine-learning algorithm that measures the semantic similarity of two bodies of text---in our case Tweets and the debate transcript. Doc2vec uses a neural network with a bag of words model in a training stage to predict which words in a given body of text will likely appear together. These predictions are used to create a lower dimensional vector space that can represent bodies of texts. Once trained, two texts that are similar in content will be represented by two vectors with small cosine distance. For example, the texts "he made a strong argument" and "he made a powerful argument" will be considered similar (in terms of cosine distance) because the words "strong" and "powerful" will be predicted to appear after "he made a" and before "argument".[5]

To perform the analysis, we first noted the time that each Tweet was posted, and defined for each a "matching event window" which was comprised of all remarks made by candidates from three minutes before to up to one minute after the Tweet.[6] Remarks made shortly after a Tweet was posted were included both to allow for possible misalignment between the recorded times of Tweets and those of the remarks, and to capture Tweets that may have commented on the first part of a longer speech by a candidate, whose time would be recorded by the median.

We trained our model with the original Tweets from all debates using a bag of words model with window length 5, and a vector space with 200 dimensions for output. We then projected each original Tweet as well as each transcript part within each temporal event window to the resulting vector space. The result is that an original tweet will be considered more similar to the transcript text if they have smaller cosine distance between them.

In Figure 2 we plot the results of this analysis, showing how the contemporaneous similarity between Tweets and the transcript evolved over the course of the debates. The figure reveals two distinctive visual patterns. First, as hypothesized (H2), as time progressed there was indeed a growing decoupling of the content of Tweets from the contemporaneous remarks being made by the candidates during the debates. Over the first 90 minutes—when data were available for all debates—the linear time trend was negative and highly significant (b_{linear}= -.003; t = -12.12, p < .001). The effect, however, was also highly nonlinear; during the first 17 minutes of

[5] For a more complete discussion of doc2vec and its computational approach see Le and Mikolov (2014).
[6] The [-3, +1]-minute window emerged from an exploratory analysis that included wider windows (e.g., [-5, +1]). The smaller window was chosen for its better performance in computed (cosine) similarity to contemporaneous candidate remarks.

the debates—when all users were focused on the same small set of events and the stock of prior Tweets was small---there was a high congruence between the content of Tweets and transcript (mean cosine similarity = .40). Afterward there is a rapid decoupling, with only a comparatively slow (albeit significant) continued decrease in similarity (linear slope from 18 to 90 minutes: $b_{\text{linear}} = -.001$; $t = -3.24$, $p = .001$).

[Insert Figure 2 about here]

The second insight is that while Tweets indeed became increasingly decoupled from the actual transcript as the debates wore on, there was minute-to-minute variation in this association, perhaps suggestive of a waxing-and-waning of attention during the proceedings. To explore whether this variation indeed reflected changes in the topics the candidates were discussing, we modeled the cosine similarities between each Tweet and retweet and the transcript as a function of the topics being discussed by candidates at the time a Tweet or retweet was posted. Our model was a linear regression of the form

$$CS_{ij} = b_o + \sum_{k=1}^{m} b_k X_{kj} + \sum_{t=1}^{3} \gamma_t D_t + \zeta CS_{ij-1} \quad (1)$$

where CS_{ij} is the cosine similarity between the content of Tweet or retweet i and its corresponding 4-minute transcript block, X_{kj} is a vector of the mean characteristics of the transcripts in block j, D_t is a fixed effect for debate where D_t=1 if the Tweet was posted in debate t and 0 otherwise, CS_{ij-1} is a one-minute lag in mean cosine similarity, and b_k, γ_t, and ζ are parameters. The vector X_{kj} was composed of measures of the content of the transcript block that corresponded to a given Tweet as follows:

1. Whether the candidates were making reference to one or more of 14 different substantive policy topics;

2. Whether the transcript block included one of the seven contentious exchanges (see above);
3. The LIWC measure of positive and negative emotionality of the candidates' words; and
4. Indicators of the candidates who were speaking.

The lag of cosine similarity was included to control for non-independence of errors in the regression due to transcript topics spanning adjacent 4-minute time intervals.

Note that structural identification of (1) is facilitated by the fact that the variables that are being used to predict the cosine similarity of Tweets---the content of the comments being made by the candidates—are fully exogenous in our data. That is, because the comments being made by the candidates are unpredictable by users and are, of course, unaffected by Tweets, there is little risk the that estimates of the parameters of (1) would be threatened by concerns about endogeneity (for example, strategic behavior by users).

Least-squares estimates of equation (1) for original Tweets and retweets for different subsets of controls are reported in Table 2, and, to aid visualization, are plotted in Figure 3. Because of the extremely large sample sizes underlying the analysis (for example, our model of retweets included over 4.3 million observations), all model effects were highly statistically significant, hence it becomes more useful to focus on the variation in absolute effect sizes.

[Insert Table 2 and Figure 3 about here]

The results yield two primary insights. First, the data give strong support for the suggestion above, that during the debate, user attention indeed waxed and waned as a function of the topics that the candidates were contemporaneously discussing. When the candidates took up the topics of immigration, abortion, or Asia (for example, trade with China or North Korean disarmament)--or one of the contentious exchanges involving Donald Trump erupted--we see a contemporaneous surge in new Tweets and retweets about those same topics. In contrast, when the debate involved topics that intuitively may have held lower appeal to contributors---such as defense spending and education policy—the content of Tweets became less well-aligned, suggesting that the focus of Tweeting was more dispersed. While the substantive content of the debate seemed to influence viewers' attention to the proceedings, the emotionality of the words spoken by the candidates did not: the similarity between the content of Tweets and the transcript was largely unaffected by the degree of positive or negative emotionality in the transcript.

The second major insight is that the focus of discussion had different effects on the alignment of retweets compared to original Tweets. The systematically higher fit of retweet models (Table 1) suggests that variation in alignment was more closely related to the consensus appeal of topics than was the case for original Tweets. A possible explanation for this difference is that it could reflect a contrast in what motivates one to retweet versus Tweet: while original Tweets are a reflection of what an *individual* finds interesting, retweets reflect content that *other users* find interesting. A good example is what happened when candidates took up the topic of the environment, an issue of plausibly less widespread interest to viewers (Figure 3, Table 1). While the onset of the topic was successful in spawning a surge of contemporaneous original Tweets about the topic, these were not Tweets that users seemed to find particularly worth retweeting or sharing with others at the time.

The Drivers of Tweeting and Retweeting

Overview. Our analysis thus far reveals that, as the debates evolved, the Twittersphere served as a highly imperfect source of real-time annotation of the event, one that waxed and waned in synchronicity depending on the nature of the topics under discussion. Given this, if one wanted to craft a Tweet that would be widely retweeted in this noisy context, what would it look like? Our central hypothesis was that the answer will depend on *when* one is posting. We hypothesized that during (vs. after) the debate parsimony would prevail over richness; the most popular Tweets would be those that offered succinct play-by-play commentary on specific ongoing exchanges, with factors that might otherwise be important for virality in other contexts—such as use of emotional words or emotionally arousing video clips—mattering less (H1). When the debate was over, however, the most successful Tweets would offer less description and more evaluation (H3), with the name-familiarity of originator of the Tweet taking a larger role (H4).

We divide our reporting of the test of these hypotheses into two parts. We first explore whether these differences can be seen in the structure of Tweets that users *created* during versus after the debate. We then investigate how these features affected virality, as measured by the number of retweets a given Tweet earned.

Features of original Tweets. We identified 2,823,970 unique original Tweets that were created during the debates, and 429,821 that were created afterward. We then measured whether these two sets of Tweets differed along fifteen linguistic and surface dimensions that fell into one of four groupings:

1) *Surface features of a Tweet:* length, inclusion of photos or videos, and inclusion of quotes;
2) *Linguistic style:* Speciteller's specificity index, LIWC authentic and analytic style scores;
3)) *LIWC emotion and drive words:* percentage use of positive and negative emotional words, and words suggestive of achievement, power, and reward;

4) *Topical content*: Topical content was measured by whether a Tweet included reference to any one of twelve different substantive policy topics (e.g., immigration, the economy), whether a Tweet made reference to one of the contentious exchanges involving Donald Trump; whether a Tweet made reference to a more abstract topical theme such as reform; and whether a Tweet made reference to a candidate's external appearance or emotions (e.g., hair, dress, or expressions of anger).

The least-squares means of these measures before and after the debate and the corresponding F-tests of contrasts are presented in Table 3.[7] All reported contrasts control for fixed effects of debate. As before, because of the extremely large sample sizes, all differences exceed standard levels of significance (all p-values < .0001) hence it is more useful to focus on directionality of effects.

[Insert Table 3 about here]

Taken together, the Table supports a conclusion that Tweets created after the debate were markedly different from those created during in the manner hypothesized. As predicted by H1, during the debate (versus after) original Tweets tended to be shorter, were less likely to contain visuals, and showed evidence of being more focused on documenting the live event, specifically by being more likely to include a quote from the debate and make a reference to a policy-related topic or theme being raised by a candidate. We also see support for the predicted shift in linguistic style toward interpretation and evaluation after the debate (H3): Tweets created after the debates were written in a more analytic and less authentic style, were more likely to include words referencing achievement, power, and reward, and were more linguistically elaborate as reflected in higher specificity scores.

[7] The correlation matrix of predictors is reported in Web Appendix 5.

We also note that while after the debate there was, as predicted, fewer references to the policy topics being discussed by the candidates, there was an *increase* in references to what might be seen as the more sensationalist aspects of the proceedings: the contentious exchanges involving Donald Trump that erupted during the debates, and comments referencing the appearance and expressions of the candidates. Hence, if afterward Tweeters were referencing the moments of the debate that they saw as best capturing the proceedings, it were these more entertaining aspects—not the expressions of position on specific policy topics.

To provide a more detailed illustration of the evolution in the structure and content of Tweets during versus after the debate, in Figure 4 we plot over-time variation in four of the measures of content: word count, use of direct quotes, enriched content (video or photo), and the proportion of positive emotional words (LIWC Tone). In Figure 5 we plot the evolution of references to policy issues versus contentious exchanges. The two figures provide three useful clarifications of how the differences in the content of Tweets evolved during and after the debates. First, if the main driver of the tendency to create shorter Tweets observed during the debate was the intensity of competing traffic for Tweets, we might expect the transition to longer content to evolve gradually as the pool of new Tweets gradually subsided after the debates (Figure 1). The plot of word count and graphics over time (Figure 4) are consistent with this prediction; while both increased after the debate, the evolution was a gradual one.

[Insert Figures 4 and 5 about here]

In contrast, after the debate the shift in the tone and topic was more discrete, suggestive of a change in the focus of—and perhaps motivation for—Tweets. Consistent with H1, Figure 5 shows that, during the debates, the Twittersphere could be seen as offering something of a play-by-play of the ongoing event, with there being large swings in the frequency with which Tweets referenced particular substantive policy issues and controversial exchanges as they came and went in the debate. Likewise, as shown in the lower-left panel of Figure 4, this play-by-play often included quotes taken from the proceedings. But as soon as the debates concluded there is a discrete shift in these patterns: consistent with H3 content now shows a decrease in the use of quotes (Figure 4, lower left), a decrease in references to substantive policy issues that periodically dominated Tweets during the debate (Figure 5) but an increase in positive emotionality (Figure 4, upper right). In contrast, as noted above while references to policy issues decreased on average after the debate, references to the contentious exchanges increased slightly, suggesting that it was recollections of these controversial exchanges---rather than references to policy---that were seen as more Tweet-worthy after the debate.

Were there segments of Tweeters? A natural concern with the above analysis is that the change in the composition of Tweets that occurred after the debates accrued to a change in the pool of contributors. While the measurable characteristics of users posting during the debate were quite similar to those posting after--for example, users posting during the debate had an average of 387 followers, while those after 395---it is still possible that the change in Tweets reflected the arrival and departure of two different segments of Tweeters. Specifically, one segment who preferred to Tweet during the event and offer play-by-play commentary on the proceedings as it unfolded, and those who refrained from Tweeting during the event, preferring to offer retrospective reflections only after it was over.

To test this alternative explanation, we identified 124,220 unique users who posted original Tweets both during *and* after the debates, and sought to examine whether the Tweets they created during versus after displayed the same contrasts as were observed for the population of users shown in Table 3. The findings, reported in Web Appendix 6, reject a notion that the differences in Tweets simply accrued to a change in the pool of contributors. Tweets created by the same users during versus after the debate showed the same temporal effects as was observed for the population, and in some cases the contrast was stronger; for example, they were even more likely to abandon discussion of policy issues to focus on the controversial exchanges that arose between the candidates.[8]

Predictors of retweeting. While the above analysis suggests that the kinds of original Tweets that were *created* differed during versus after the debate, it does not inform whether there was a parallel change in drivers of the popularity of these Tweets---that is, which ones were more likely to be retweeted by users. One of the challenges we face in testing whether different factors drive sharing during versus after the debate, however, is that Tweets generated after the debate were structurally different than those generated during (e.g., they tended to be longer; see Table 2), something that would confound tests of time effects on the correlates of retweeting. We were able to overcome this identification problem, however, by leveraging an important feature of our data: the fact that many of the 2.8 million Tweets posted during the debates continued to be retweeted after the debate. This allowed us to model the drivers of retweet counts computed for *the same set of Tweets* measured at two different points in time: during the debate when the Tweets were first posted, and afterward when these same Tweets were being considered in hindsight. To test the robustness of this latter analysis, we then separately modeled the drivers of retweeting for the smaller set of new Tweets that were created after the debate. If there was a homogeneous process driving retweet rates after the debate, we should see a similar signature of effects for both sets of models.

[8] Among users posting both during and after the debates, 13.43% of Tweets made an explicit reference to a policy topic, but afterward only 4.93% did. As shown in Table 2, the contrast for all users was 11.38% and 4.40%, respectively.

Because the count distribution of retweet frequencies was extremely skewed (90% of all original Tweets had no retweets) and over-dispersed, the retweets during versus after the debates data were analyzed via a series of negative binomial regressions. Formally, we assumed that the expected frequency with which Tweet *i* would be retweeted, $E(y_i)$ could be modeled as

$$E(y_i) = e^{\mu_i + \varepsilon_i}$$

where ε_i was a gamma-distributed error term with mean 1 and variance α^2 (Hilbe 2011), and μ_i was a linear combination of predictors, given by

$$\mu_i = b_o + \sum_{j=1}^{m} b_j X_{ij} + \sum_{t=1}^{3} \gamma_t D_{it} + \sum_{l=1}^{n} \zeta_l Z_{il} \qquad (2)$$

where X_{ij} was vector of Tweet and user characteristics, D_{it} a vector of debate-specific effects, and Z_{ij} a vector of candidate-specific effects.

For ease of exposition, we report the coefficients of a variant of equation (2) that included seventeen characteristics of the Tweets and Tweeters. These seventeen characteristics included the four groups of fifteen Tweet characteristics considered above (surface features, linguistic style, LIWC emotion and drive words, and topic), as well two characteristics of the Tweeter that had been found to affect retweet rates in other contexts (e.g., Suh, et al, 2010): the log of the number of users who follow the Tweeter, and Tweet activity, measured here by the total number of Tweets generated by a user during and after the debates. In addition to the Tweet and user characteristics, we included in the model fixed effects for debate and the one candidate that was present in all of the debates, Donald Trump. The Trump fixed effect was of particular interest as it ensured that our estimate of the effect of references to controversial exchanges on retweet rates would not be confounded with a general tendency for retweet rates to be to higher for Trump, who was the leading candidate in the GOP debates and the focus of the controversial exchanges.

In Web Appendix 7 we provide evidence for the robustness of the model estimates across a range of estimation methods (Poisson, log-linear; Table WA7-1) and with alternative treatments of candidate and debate fixed effects (Table WA7-2). We also report estimates for a variant of equation (2) that estimates effects for specific policy issues rather than an aggregate (Table WA7-3).

Results. In Table 4 we report the coefficients of expression (2) estimated for retweets of a common set of Tweets posted during the debate and after, as well differences in coefficients and pooled standard errors. Again because of our extremely large samples sizes all model effects and differences in parameters between stages reported in the table are highly statistically significant ($p < .0001$), so we focus on the absolute sizes and direction of the effects.

[Insert Table 4 about here]

First, during the debate the largest drivers of whether a Tweet was retweeted were perhaps unsurprising. While those created during the debates tended to be shorter and were less likely to contain visuals (Table 2), those that managed to stand out by including such content tended to see higher retweet rates. We might note that the coefficients for the percentage of positive and negative emotional words was quite small, suggesting that emotionality played a limited role in retweeting during the debate. A caveat to this result, however, is that the use of emotional words was negatively correlated with the insertion of visuals (r=-.08; see Web Appendix 5) suggesting that emotionality may indeed been playing a role, but more when conveyed through pictures than words. Finally, Tweets that referenced one of the contentious exchanges that arose involving Donald Trump---events we singled out in hindsight for their notoriety—also had, as expected, higher rates of retweeting. In contrast Tweets that strayed off topic from the debates by referencing the physical appearance of candidates and used informal language (e.g., swear words and netspeak) tended to realize lower retweet rates.

More important for our purposes, however, are the noticeable *changes* that occurred in the factors most predictive of retweeting during versus after the debate. Consistent with our hypotheses (H1, H3), after the debate textual and visual elaboration and interpretation were more important drivers of retweeting than during. This was evidenced by an increase in the importance of longer word counts and of visuals in driving retweeting after the debate, as well as an increase in the importance of Tweet features that suggest greater linguistic elaboration and interpretation. Specifically, after the debate Tweets written in a more analytic—but not authentic style—were more likely to be shared, as well as those that used words conveying achievement, power, reward, and positive emotion. In contrast, factors that would suggest description over interpretation—the relaying of quotes and references to specific polity topics---decreased in importance afterward. Likewise, as predicted by H4, there was also a significant increase in the importance of the prominence of the Tweeter in driving post-debate retweet rates. While postings by users with larger followings saw higher retweet rates on average, Tweets by more prominent users were more likely to be rediscovered and retweeted *afterward* than when they were first posted in the heat of the debates.

We might note that the changes in importance of factors that drove retweeting before versus after mirrored, to a large degree, the changes we reported above in kinds of Tweets that were created afterward. For example, both reveal an increase in interest in the contentious exchanges involving Donald Trump after the debate, something reflected both in new Tweets (Table 2) as well as an increase in the importance of such references in affecting retweets (Table 3). In contrast, not all changes in content after the debate were associated with greater popularity. Most notably, while after the debates there was an increase in Tweets that made references to appearance of candidates, Tweets that referenced appearance were less likely to be retweeted during the debate, and even to a lesser extent afterward. In contrast while we observed a decrease in the creation of Tweets that referenced more abstract policy themes during the debate (such as political correctness, needs to reform; Table 3), Tweets that had these references saw an increase in retweeting afterward (Table 4).

To test whether this same pattern of drivers of retweeting after the debate also applied to new Tweets created afterward, in Web Appendix 8 we report parameter estimates for a model of retweet rates for the smaller set of post-debate Tweets. Although the parameters of this model are not directly comparable to those of the model estimated on Tweets first posted during the debate (it is a different Tweet pool), the findings nevertheless conceptually replicate those reported above. Specifically, the most popular Tweets created after the debate were more likely to contain an embedded video and/or photo, make reference to abstract themes and the contentious exchanges over character, make greater use of drive words (achievement, power and reward), and were more likely to have been posted by bloggers with large followings.

Finally, to provide a clearer intuition into how the drivers of retweeting differed during the debate than after, it is instructive to visually contrast the two Tweets posted during the third Presidential debate that had the highest rate of retweeting during versus after:

"Hombres" (12,848 retweets during. 698 after);

"' Nobody Respects Women more than me'-- Donald Trump earlier Tonight. "'Such a Nasty Woman'—Just now" (1 retweet during, 30,779 after)

Both Tweets were posted by staff from Hillary Clinton's campaign near the end of the debate, and both reference a contentious remark made by Donald Trump. But note that they do so in very different ways, and different effectiveness. *"Hombres"* is linguistically sparse but easy to create and comprehend—a Tweet ideally suited for posting during the heat of the debate. In contrast, *"Nobody respects…"* is more elaborate and conveys evaluation in a subtle manner---a Tweet ideally suited for after the debate when the debate is being viewed in hindsight.

General Discussion

This research was motivated by a desire to better understand the drivers of social media use during live political events. We explored this issue by analyzing the drivers of Tweet creation and retweeting during a series of critical debates leading up to the 2016 presidential election. While Tweeting occurs during all phases of election campaigns, debates form a particularly important focus of study because of the large influence that public perceptions of performance can have on have on election outcomes (McKinney and Warner 2013)---and Twitter is increasingly looked to as major indicator of those perceptions (e.g., Dumenco 2016).

Using a battery of natural-language processing tools we emerged with two major findings about how the process of creating and sharing Tweets evolved during debates. The first was that Tweeting was marked by two distinct temporal regimes: a shift from *real-time narration* during the debates to *retrospective interpretation* afterward. During the debates users acted more as play-by-play narrators of the event (H1), by relaying quotes from the candidates and commenting on the policy issues being discussed onscreen in real time, while foregoing the inclusion of more emotional language or enriched content (video links and images). Tweets that had these features were also the ones more likely to be shared. In contrast, when the debates were over and the debates were being viewed in hindsight (H3), Tweets became linguistically more elaborate, and interpretation, emotional expression, and the originator of the Tweet (H4) now became larger drivers of success. After the debates there was also a change in the topical focus of Tweets.

Whereas during the debates there was an active focus on the policy issues that the candidates were discussing onscreen, afterward such references abated, with there being a greater focus on events that arguably reflected more on candidates' character than on their competence, specifically the contentious exchanges that erupted between Donald Trump and other candidates and moderators.

The second major finding was that as debates progressed there was an increasing decoupling between the content of Tweets from the on-state dialogue among the candidates. Hence, rather than offering a "community annotation" of debates (Shamma, et al. 2009; 2010), Twitter more often resembles a political free-for all, one where users Tweet not just about the contemporaneous issues being talked about by candidates, but also exchanges that occurred earlier in the debate, and, in some cases, political issues that lie outside the debate. One implication is that rather than serving to heighten viewers' focus on the views being expressed by candidates, Twitter may, more often, act as a distractor.

Possible Mechanisms

What were the mechanisms that drove the temporal changes we observed in Tweeting? We suggest that multiple forces were in play, some structural, some psychological. For example, a simple explanation for short, narration-like Tweets during the debates was that there was little opportunity to do otherwise. During debates Tweeters were likely multitaskers, watching the live debate out of one eye, following their Twitter feeds out of another, all the while thinking about what new content to create or share. Moreover, there was likely an implicit race to be timely---to be the first make a comment about a comment or exchange that would be shared by others. In such a pressured environment there would be little opportunity for reflection or elaboration, and Tweets that did so might well be overlooked in favor of content that was succinct and timely. After the debate, however, these constraints would have been relaxed, providing more opportunities for more elaborate content to be created and shared.

Psychological differences might also be at work. The observed tendency for Tweets after the debate to use more interpretative language might be seen as consistent with findings on construal-level theory (Trope and Lieberman 2010), which suggest that when events are viewed from a greater distance (e.g., temporal; physical), people tend to construe them more abstractly (vs. concretely). Within this framework, Tweeters who recall the debates may have thought less about its concrete specifics (e.g., the exact words that candidates used, details of appearance), and more about higher-level meanings or interpretations, such as the overall performance of the candidates. Prior supportive evidence for this effect in a political context has been provided by Bhatia and Walasak (2016), who found that, the closer a *New York Times* article was published to the election date, the more concrete language it used when describing the election. We also found some evidence for this in our own data: Tweets about the debates that included more past-focused (vs. present-focused) words were more likely to also include interpretative words suggestive of both achievement and reward (r(focuspast, achievement)=.07; $p<.001$; (r(focuspresent, achievement)=-.04; $p<.001$; (r(focuspast, reward)=.10 $p<.001$; (r(focuspast, reward)=.00 $p=.003$).[9]

Finally, another finding of ours that might be interpreted in a similar light is the change in topical content we observed during versus after the debate: the shift from a focus on policy topics (e.g., immigration) to more abstract themes (e.g., healing) and contentious exchanges. If users indeed saw themselves as reporters during the debates, the events that would loom largest are the concrete policy issues that candidates were discussing in real time. But when the debates

[9] References to power-related words such as "beat" tended be negatively correlated with references both to the past (r(focuspresent,power)=-.03; p<.001) and present (r(focuspast,power)=-.04; p<.001).

were finished and the event was being reflected back on, Tweeters higher-level construals would favor a focus on such holistic issues such as who won, abstract themes that bridged the more specific comments, and moments that stood out as the most memorable. In this latter regard, the events that most differentiated these debates from others would have been the contentious exchanges that erupted involving Donald Trump—exchanges that would later come to define the debates in many post-election accounts (e.g., Lake 2016).

Do the effects generalize to other live political events?

While the evidence for our findings spans multiple debates and millions of Tweets, a natural question is whether the dynamics we observed may generalize to Tweeting during and after *any* comparable live event. To examine this, we repeated the same analyses reported here for a very different kind of political event that was widely watched by Americans: Barack Obama's final State of the Union address given on January 12, 2016. Similar to debates it was an event in which Twitter users where observing the same event and Tweeting about a rapidly changing array of exogenous topics. But it also differed in a number of potentially important ways: here there was no contest being waged among candidates, and the motive for watching was likely relatively more passive: that of listening to a scripted speech.

In Table 5 we report the mean linguistic features of the census of 449,817 unique unprotected Tweets that were created during and after Obama's speech, and in Table 6 the parameters of negative binomial regression models fit to retweet counts during and after. Although observed in a very different setting, the results show similarities to the dynamics observed in our debate data (Tables 3 and 4). Specifically, like in the debates, Tweets created after the address were longer, more likely to have been written in an analytic—and less authentic form--were less likely to contain quotes or make references to specific policy issues, but were more likely to make reference to superficial issues such as Obama's appearance and expressions. Likewise, these same effects were mirrored in what drove Tweet popularity after the debate compared to during; For example, Tweets displaying more emotion—here negative—were more likely to be shared (Table 6), and, as we observed in the debates, Tweets making reference to appearance were even less likely to be shared after the address than during. But here there was at least one important difference between the two contexts. A notable feature of Obama's final address was his frequent appeals to both the achievements of his administration, as well as abstract themes such as hope and healing (Baker 2016). Possibly because of this, references to these elements were more common among Tweets made during the debate (when it was being narrated) than after.

[Insert Tables 5 and 6 about here]

Policy implications

We see this work as holding a number of implications for political marketing. The most important is that it implies that very different Tweeting strategies would be called for depending on which target audience one intends to reach: during the debate when the audience is the largest but attentions are divided, or afterward when the audience is smaller but attention is undivided. In the former case the greatest success will likely be found in blogs that eschew elaboration and interpretation in favor of succinctness and timely references to the events at hand. In the latter case the opposite holds; the most successful tweets will be those that are rich in visuals, emotion, and focus on abstract themes over concrete details.

To illustrate these tactical differences, it is first important to note that the odds that any single Tweet will be shared in the Twittersphere are quite low. Of the 1.7 million Tweets that were created during the Presidential debate that we analyzed, for example, 83% were never retweeted, 1.6% earned 10 or more retweets, and only .01% --211 Tweets—earned more than 1,000 retweets. Having large followings helped, but not as much as one might hope: Tweets posted by the top 10% of users in terms of following sizes (3,789 or more followers) saw their postings retweeted 44% of the time, but only 14% earned more than 10 retweets. As such, contributors during the debates appeared to recognize that it was a numbers game, with the average user creating 16 original tweets during the debates and 7 retweets. While posting more indeed increases the odds that one's posts will be seen, in our data active Tweeters were no more skilled at formulating Tweets that had higher retweet rates than the average user (as shown in Table 3, the coefficient for number of debate Tweets on the retweet count is close to zero).

On the other hand, the odds go up for users who make the right choices in crafting their Tweet, choices that critically depend on timing. During the debate, for example, a Tweet that referenced the policy issues that were being discussed by the candidates, included a quote by a candidate, and supplemented these elements with graphics, would be predicted (from equation (2)) to have a 63% higher retweet rate than a Tweet that lacked these features (from 1.42 to 2.32 expected retweets). After the debate, in contrast, the formula for success changed. Although the base rate of post-debate retweeting of Tweets first posted during the debate was low (less than 1 expected retweet per Tweet), it was also much more influenced by design. A Tweet posted during the debate that included achievement words, included graphics, and referenced one of the contentious exchanges involving Donald Trump---but not policy topics---would be predicted to have a retweet rate 144% higher than a Tweet that lacked these features (from .03 to .42) retweets.

Limitations and future work

In this paper we reported large-scale descriptive evidence of how Tweeting evolved during and after a series of live political events. One of the limitations of analyzing social media use at this scale, however, is that it inhibited our ability to gain deep insights into psychological and social processes that may have underlain posting decisions. For example, at the outset we noted that consumers sometimes use Tweeting as a means to express their political identities (e.g., Sylwester and Purver 2015), and there is evidence that the process of viewing Tweets can work to activate those identities (e.g., Houston et al. 2013). An important next step in this research would be to investigate dynamics not just in the structure of Tweets themselves, but also changes in the mixture of political orientations that compose the contributor pool. For example, one might hypothesize that remarks by a candidate that trigger an upwelling of supportive comments by his or her followers could discourage contributions by supporters of opponents, who see few like-minded users posting on their feeds. But it also could have the opposite effect of activating the identities of opponents and encouraging them to post.

Another area for future research would be to further probe the degree to which the findings generalize beyond political contexts that would be of interest to marketers. Twitter is increasingly recognized as an important way to reach consumers during live events (such as Super Bowls and awards shows) outside of traditional paid media, and the current research could offer insights into how to best reach audiences during such events. One of the implications of the current research is that during such events many of the traditional principles of how to get heard on social media---such as by making content more emotional (e.g., Berger and Milkman 2012)—might need to be set aside. During live events there may be limited rewards to such tactics; it is a numbers game where frequent brief posts that narrate real-time action may be the best way to be heard. But while debates and political speeches have structural parallels to entertainment events, there are also differences that could qualify the findings. For example, such events typically have a predictable flow (e.g., the top awards appear at the end), and, of course, the motivation for Tweeting about one's favorite artist on the Grammys is likely different from the motivation to Tweet about candidate performance in a debate.

Finally, we see the work as contributing to the rapidly-growing subfield of political marketing by illustrating the behavioral insights that can be gained by applying recent advances in natural-language processing tools to text data. For example, in this work we leveraged a new machine-learning algorithm for measuring the linguistic similarity of two different documents to draw inferences about how closely Tweets were following the actual debate as it progressed. While, in principle, such assessments could also be made on a small-scale basis using human judgments, *doc2vec* allowed us to perform the calculation on a complete census of millions of Tweets being paired with lengthy transcripts. A potential avenue for future research would be to explore further applications of the approach to a broader range of marketing problems, such as measuring the degree to which social media posts mimic the language in communications by firms.

References

Beliga, Slobodan (2016), “Keyword extraction: a review of methods and approaches”, accessed at: https://pdfs.semanticscholar.org/bdbf/25f3dcf63d38cdb527a9ffca269fa0b8046b.pdf

Baker, Peter (2016), “In Final State of the Union Address, Obama Aims to Set Tone for ’16 Campaign:”, New York Times, Jan. 10, 2016, accessed at: https://www.nytimes.com/2016/01/11/us/politics/obama-last-state-of-the-union.html

Berger, Jonah, and Katherine Milkman (2012), “What Makes Online Content go Viral”, *Journal of Marketing Research,* 49 (2), 192-205.

Bhatia, Sudeep, and Lukasz Walasek (2016), “Event construal and temporal distance in Natural Language”, *Cognition*, 152, 1-8.

Bramlett, J., McKinney, M.S., and Warner, B. R., (2017) “Processing the Political: Presidential Primary Debate 'Live-Tweeting' as Information Processing” in B. R. Warner, D. G. Bystrom, M. S. McKinney, and M. C. Banwart, Eds., *An Unprecedented Election: Media, Communication and the Electorate in the 2016 Campaign,* ABC-CLIO publisher.

Conover, M. D. J. Ratkiewicz, M. Francisco, B. Goncalves, A. Flammini, and F. Mencze (2011), “Political Polarization on Twitter”, *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 89-96.

Coppins, McKay (2016*), The Wilderness: Deep inside the Republican Party’s Combative, Contentious, and Chaotic Bid to Re-take the White House*, Little Brown and Company, New York.

Dang-Xuan, L., Stieglitz, S., Wladarsch, J., and Neuberger, C. (2913), “An investigation of influential and the role of sentiment in political communications on Twitter during election periods”, Information, Communication, and Society, 16(5), 795-825.

Dumenco, Simon (2016), “The Third Presidential Debate Summed Up in 31 Tweets”, Ad Age, 20 October, available at: https://adage.com/article/campaign-trail/presidential-debate-summed-31-tweets/306365/

Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., … P., M. E. (2015). Psychological Language on Twitter Predicts County-Level Heart Disease Mortality. *Psychological Science*, *26*(2), 159–169.

Ferrer, Ferrara E, Yang Z (2015) “Measuring Emotional Contagion in Social Media”, *PLoS ONE 10(11).*

Flores, Reena (2016), “Fox News Republican debate smashes 2016 viewership record”, CBS News, March 4, 2016, available at: https://www.cbsnews.com/news/fox-news-republican-debate-smashes-2016-viewership-record/

Fujita, Kentaro, Henderson, Marlone D.. Eng, Juliana, Trope, Yaacov, and Nira Liberman (2006). "Spatial distance and mental construal of social events," *Psychological Science* 17, 78-82 .

Gordon, B. R. and Hartmann, (2013) Advertising Effects in Presidential Elections. *Marketing Science* **32**:1, 19-35.

Gordon, Brett R. and Wesley R. Hartmann. (2016) Advertising competition in presidential elections. *Quantitative Marketing and Economics* **14**:1, 1-40.

Guadagno, Rosanna E., Rempala, Daniel M., Muphy, Shannon, and Bradley M. Okdie (2013), "Why do Internet Videos Go Viral? A Social Influence Analysis", *Computers in Human Behavior* 29, 2312-2319.

Han, Kyuhong, Jung, Jihye, Mittal, Vikas, Zyung, Jinyong Daniel, and Hajo Adam (2019), "Political Identity and Financial Risk Taking: Insights from Social Dominance Orientation." *Journal of Marketing Research*, (March).

Hansen, Lars Kai, Arvidsson, Adam, Nielsen, Finn Arrup, Colleoni, Elanor, and Michael Etter (2011), "Good Friends, Bad News-Affect and Virality in Twitter", in *Future Information Technology*, vol. 185 of *Communications in Computer and Information Science*, pp. 35-43.

Harber, Kent D., and Dov J. Cohen (2005), "The emotional broadcaster theory of social sharing", *Journal of Language & Social Psychology*, 24(4), 382-400.

Heath, Chip (1996), "Do People Prefer to Pass Along Good or Bad News? Valence and Relevance of News as Predictors of Transmission Propensity", *Organizational Behavior and Human Decision Processes*, 68(2), 79-94.

Heller, Nathan (2016), "The First Debate of the Twitter Election", *New Yorker*, September 27.

Hilbe, Joseph M. (2011). *Negative Binomial Regression* (Second ed.). Cambridge, UK: Cambridge University Press.

Hoang, Tuan-Anh, Cohen,, William W., Lim, Ee-Peng, Pierce, Doug, and David P. Redlawsk. (2013), "Politics, sharing and emotion in microblogs". *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (ASONAM '13). ACM, New York, NY, USA, 282-289.

Hoegg, J., and M.V. Lewis. 2011. The impact of candidate appearance and advertising strategies on election results. *Journal of Marketing Research* 48, no. 5: 895–909.

Houston, J. Brian & Hawthorne, Joshua & L. Spialek, Matthew & Greenwood, Molly & Mckinney, Mitchell. (2013). Tweeting During Presidential Debates: Effect on Candidate Evaluations and Debate Attitudes. *Argumentation and Advocacy*. 49. 302-213.

Jost, John T. "The marketplace of ideology: “Elective affinities” in political psychology and their implications for consumer behavior." *Journal of Consumer Psychology* 27, no. 4 (2017): 502-520.

Jung , M. H., & Critcher, C. R. (2018). How Encouraging Niceness Can Incentivize Nastiness: An Unintended Consequence of Advertising Reform.Journal of Marketing Research , 55, 147-161.

Kanavos A., Perikos I., Vikatos P., Hatzilygeroudis I., Makris C., Tsakalidis A. (2014) Modeling ReTweet Diffusion Using Emotional Content. In: Iliadis L., Maglogiannis I., Papadopoulos H. (eds) Artificial Intelligence Applications and Innovations. AIAI 2014. IFIP Advances in Information and Communication Technology, vol 436. Springer, Berlin, Heidelberg.

Kashmiri, Saim, and Vijay Mahajan (2017), "Values that shape marketing decisions: influence of chief executive officers’ political ideologies on innovation propensity, shareholder value, and risk." *Journal of Marketing Research,* 54, 2, 260-278.

Kim, Jeehye Christine, Brian Park, and David Dubois (2018). "How Consumers’ Political Ideology and Status-Maintenance Goals Interact to Shape Their Desire for Luxury Goods." *Journal of Marketing,* 82, 6, 132-149.

Kollanyi, B., Howard, P. N., & Woolley, S. C. (2016). Bots and automation over Twitter during the first US presidential debate. *Comprop data memo, 1,* 1-4.

Lake, Thomas (2016), *Unprecedented: The Election That Changed Everything,* New York: Melcher Media

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents, *International Conference on Machine Learning* (pp. 1188-1196).

Levy, Gabrielle (2016), “Twitter Wins Big in 2016 Campaign”, U.S. News and World Report 8 November, available at: https://www.usnews.com/news/politics/articles/2016-11-08/more-than-1-billion-tweets-were-sent-about-the-election

Li, Junyi Jessy and Ani Nenkova (2015), “Fast and accurate prediction of sentence specificity”, *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (AAAI'15). AAAI Press 2281-2287.

Ludwig, Stephan, Ko de Ruyter, Mike Friedman, Elisabeth C. Brüggen, Martin Wetzels, and Gerard Pfann (2013), "More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates," *Journal of Marketing*, 77 (January), 87-103.

Maruyama, M., Robertson, S. P., Douglas, S. K., Semaan, B. C., and Faucett, H. A. (2014), “Hybrid media consumption: How tweeting during a televised political debate influences the vote decision”. *In Proceedings of the 17th ACM conference on Computer supported Cooperative work & social computing,* 1422-1432.

McInney, M.S., and Benjamin R. Warner (2013), “Do Presidential Debates Matter? Examining a Decade of Campaign Debate Effect”, *Argumentation and Advocacy*, 49(Spring), 238-58.

McGregor, Shannon C., and Logan Molyneux (2018). “Twitter’s influence on news judgment: An experiment among journalists” *Journalism*/ available at: https://journals.sagepub.com/doi/10.1177/1464884918802975

Melumad, Shiri, J. Jeffrey Inman and Michel Tuan Pham (2019), “Selectively Emotional: How Smartphone Use Changes User-Generated Content,” *Journal of Marketing Research*, 56(2), 259–275.

Ordabayeva, Nailya, Fernandes, Daniel (2018), “Better or Different? How Political Ideology Shapes Preferences for Differentiation in the Social Hierarchy,” *Journal of Consumer Research*, 45 (2), 227–50.

Pennebaker, James W., Boyd, Ryan L., Jordan, Kayla, and Kate Blackburn (2015), *The Development and Psychometric Properties of LIWC2015*, Austin, TX: University of Texas at Austin.

Phillips, J.M., J.E. Urbany, and T.J. Reynolds. 2008. Confirmation and the effects of valenced political advertising: A field experiment. *Journal of Consumer Research* 34, no. 6: 794–806.

Shamma, D.A., Kennedy, L., and E. F. Churchill (2009). “Tweet the debates: Understanding community annotation of uncollected sources”. In WSM ’09: Proceedings of the international workshop on Workshop on Social Media, pp. 3-10.

Shamma, D. A., Churchill, E. F., & Kennedy, L. (2010). Tweetgeist: Can the Twitter timeline reveal the structure of broadcast events? Paper presented at the 2010 ACM conference on computer supported cooperative work and social computing—CSCW (pp. 589–593). New York, NY: ACM.

Stieglitz, Stefan, and Lin Dang-Xuan (2013), “Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior”, *Journal of Management and Information Systems*, 29(4), 217-248.

Stromer-Galley, Jennifer (2014), *Presidential Campaigning in the Internet Age*. New York: Oxford University press.

Suh, Bongwon, Lichan Hong, Peter Pirolli, and Ed H. Chi (2010), “Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network”, *Social Computing (SocialCom), 2010 IEEE Second International Conference on* (pp. 177-184).

Sylwester, Karolina and Matthew Purver (2015). “Twitter Language Use Reflects Psychological Differences between Democrats and Republicans.” *PloS one* vol. 10, 16 Sep. 2015,

Toubia, Olivier and Stephen, Andrew T. (2013), *Intrinsic vs. Image-Related Utility in Social Media: Why Do People Contribute Content to Twitter? Marketing Science*, 32 (3). pp. 365-367.

Trilling, D. (2015), “Two Different Debates? Investigating the Relationship Between a Political Debate on TV and Simultaneous Comments on Twitter”, *Computer Science Review*, 33(3), 259-276.

Trope, Yaacov, and Nira Liberman (2010). *"Construal-level theory of psychological distance" Psychological Review,* 117 (2), 440–463.

Tumasjan, Andranik, Oliver-Sprenger, Timm, Sandner, Philipp and Isabell Welpe (2010). “Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment”. *Journal Of The International Linguistic Association*. 10.

Twiiter (2016), “Tonight's #debate was the most Tweeted ever, with more than 17 million debate-related Tweets sent” vailable at: https://twitter.com/twitter/status/785322512280276992

Varol, O. Ferrara, E., Davis, C., and Menxzer, F., and A. Flammini (2017), “Online Human-Bot Interactions: Detection, Estimation, and Characterization”, *International AAAI Conference on Web and Social Media*, North America, May.2017. Available at: <https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587/14817>.

White, Theodore H. (1961), *The Making of the President 1960*. Republished 2009 by Harper Collins, New York.

Wilson, Theresa, Wiebe, Janyce, and Paul Hoffmann (2005), “Recognizing contextual polarity in phrase-level sentiment analysis”, *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing* (HLT '05). Association for Computational Linguistics, Stroudsburg, PA, USA, 347-354.

Winterich, Karen Page, Vikas Mittal, and Karl Aquino (2016) "Moral and political identity." *Cambridge Handbook of Consumer Psychology*: 589-618.

Zheng, Pei, and Shahin, S. (2018): Live tweeting live debates: How Twitter reflects and refracts the US political climate in a campaign season”, *Information, Communication & Society,* 1-20.

Authors	Year	Corpus	Sample Size	Key Finding	Coding Method
Using Sentiment to Predict Political Outcomes					
Bermingham and Smeaton	2011	Tweets During 2011 Irish General Election	32578	Positive association between Twitter sentiment and election outcomes	Human
Burnap et al.	2016	Tweets Concerning 2015 UK Election	13899073	Tweets sentiment predicted UK election result	Automatic
Choy et al.	2011	Tweets During 2011 Singapore Election	16616	Tweets pedicted top two contenders in Singapore.	Human
Jensen and Anstead	2013	Tweets During 2012 Iowa Caucuses	697065	Percentage of tweets mentioning a candidate correlates with vote percentage.	NA
McKelvey et al.	2014	Tweets During 2010 US House Election	113985	Tweet sentiment correlated with public voting choices.	Human
Mejova et al.	2013	Tweets During 2011 Republican Primaries	6400	Weak association betwwen Tweet sentiment and polls	Human
Nooralahzadeh et al.	2013	Tweets During 2012 Presdiential Election	197000	Sentiments and keywords for each candidate correlate with real time returns	Human
O'Connor et al.	2010	Political Tweets Between 2008 to 2009	1000000000	Tweet Sentiment correlated with presidential job approval polls.	Human
Sang and Bos	2012	Tweets During 2011 Dutch Senate Election	7000	Seat numbers predicted by Tweet sentiment close to the election results.	Human
Skoric et al.	2012	Tweets During 2011 Singapore Election	110815	Frequency of name mentions in Tweets predicts share of votes	NA
Thomson and Ehizokhale	2016	Tweets During 2016 Primary Debates	200000	Tweet sentiment oly weakly follows overall popularity	Automatic
Tumasjan et al.	2011	German election Tweets	104003	Party mentions and sentiment political preferences.	Automatic
Tunggawan and Soelistio	2016	Tweets During 2016 US Presidential Election	33708	Sentiment wrongly predicted that Sanders and Cruz would be 2016 nominees	Human
Jahanbakhsh and Moon	2014	Poltical Tweest leading up to 2012 election	39000000	Twitter volume and sentiment correlated with election outcomes	Automatic
Ramteke et al.	2016	Tweets During 2016 Presidential Election	30000	Proposal for ML approach to use Tweets to predict election outcomes	Automatic
Using Tweet feattures and Sentiment to Predict Retweet rates					
Dang-Xuan et al.	2013	Tweets During Election in Berlin	17778	Positive effect of positive emotionality on retweet counts	Human and Auto
Hansen et al.	2011	News and non-news Tweets	556644	Negativity enhances virality for news, but not for non-news Tweets	Human
Pancer and Poole	2016	August 2016 Tweets by Clinton and Trump	3600	Messge fluency more important thab valence in retweeting	Automatic
Pfitzner et al.	2012	Tweets in multiple contexts	35000000	Divergence of Sentiment predicts retweeting	Automatic
Stieglitz and Dang-Xuan	2012	Tweets Prior to 2011 Germany election	64431	Posiitve and negative sentiment associated with higher retweeting	Automatic
Stieglitz and Dang-Xuan	2013	German ElectionTweets	168748	Higher affect associated with higher retweeting	Automatic
Suh et al.	2010	General Tweets	74010000	Retweetiing positively associated with number of followers	n/a
Tsur and Rappoport	2012	Tweets in multiple contexts	400000000	The content of the idea plays a important role in its acceptance by the communi	n/a
Relationship between Tweeting and Triggering Events					
Ringsquandl and Petković	2013	Republican Candidates in 2011	54728	Evidence that there is a relationship between candidates, including their sentim	Human
Bollen et al.	2011	Tweets During 2008	1100000	Tweet sentiment reflets public reaction to political events	Automatic
Diakopoulos and Shamma	2010	2008 Presidential Debate	3238	Tweet sentiment varied during the course of a debate+E57	Human
Wang et al.	2012	Tweets During 2012 Election	17000	Tweet volume and sentiment covaries with campaign events. \	Human
Kalsnes et al.	2014	Tweets During 2011 Norway Debates	2391	Twitter not only provides a backchannel for reflections on topics discussed, bu	Human
Shamma et al.	2009	Tweets During 2008 Presidential Debate	3238	Tweet content unaligned with debate and news topics	NA
Trilling	2015	Tweets During 2013 German Election	120557	Tweet volume co-varied with external election events	NA
Predictors of Tweeting					
Houston et al.	2013	Tweets During 2012 Presidential Debates	768	Tweeting enhanced viewer engagement in a debate	Automatic
Kollanyi et al.	2016	Tweets During 2016 Presidential Debates	9000000	Found more pro-Trump twitter traffic driven by bots than pro-Clinton traffic	NA
Maruyama et al.	2014	Tweets During 2012 Hawaii Election	407	Twitter sentiment changed voting preferences	Human
Sylwester and Purver	2015	Political Tweets in June 2014	923758	Republican and Democratic Tweets have different linguistic markers	Automatic
Tweeting and Network Analysis					
Conover et al.	2011	Tweets Before 2010 US l Midterm Elections	252300	Content of political discourse on Twitter remains highly partisan.	NA
Larsson and Moe	2012	Tweets During 2010 Swedish Election	99832	Can infer political orientations by Tweet patterns	NA
Sylwester and Purver	2015	Twitter users who follow parties	456114	Can infler political clusters may analyzing following patterns	NA
Zheng and Shahin	2018	Tweets During 2016 Presidential Debates	300000	Evidence of distinct partison contributors of Tweets	NA
Methods with Political Applications					
Bakliwal et al.	2013	Tweets During 2011 Irish General Elections	2624	Able to classify a tweet as being positive, negative, or neutral towards	Automatic
Chin et al.	2016	Tweets During 2016 Presidential Election	3000	Uses emojis to train sentiment prediction	Human and Auto
Razzaq et al.	2014	Tweets During 2013 Pakistan Election	612802	Uses similarity metric to predict Tweet popularity	Human
Shamma et al.	2010	Tweets During 2008 Presidential Debate	8300	Shows feasibility of temporal tracking of commonly used text terms	Automatic

Table 1: Synopsis of Prior Work Analyzing Social Media Data in Political Contexts

		Model 1		Model 2		Model 3		Model 4	
	Predictor	**Original**	**Retweet**	**Original**	**Retweet**	**Original**	**Retweet**	**Original**	**Retweet**
Topics	Intercept	0.22	0.11	0.16	-0.04	0.16	-0.03	0.18	-0.01
	Asia	0.00	0.18	0.11	0.17	0.11	0.21	0.11	0.20
	Europe	0.05	-0.09	0.02	-0.05	0.02	-0.06	0.02	-0.06
	Middle East	-0.13	-0.03	-0.07	0.04	-0.07	0.03	-0.08	0.02
	Environment	0.11	0.01	0.09	0.01	0.09	-0.06	0.12	-0.01
	Defense	-0.14	-0.16	-0.10	-0.09	-0.10	-0.10	-0.08	-0.10
	Education	-0.19	-0.17	-0.09	-0.01	-0.09	-0.03	-0.14	-0.04
	Immigration	0.07	0.21	0.05	0.13	0.05	0.14	0.06	0.12
	Terrorism	-0.09	0.02	-0.10	0.03	-0.10	0.05	-0.08	0.04
	Economy	-0.05	-0.03	-0.11	-0.04	-0.11	-0.05	-0.11	-0.03
	Health	-0.08	-0.18	0.12	-0.02	0.12	-0.01	0.11	-0.02
	Courts	0.03	0.41	-0.02	0.29	-0.02	0.27	-0.02	0.27
	Abortion	0.22	0.35	0.12	0.19	0.12	0.17	0.11	0.19
	Abstract theme	0.02	0.14	0.03	0.10	0.03	0.06	0.02	0.00
	Contentious	0.20	0.21	0.12	0.12	0.12	0.14	0.12	0.13
Emotion	Positive Emo	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	Negative Emo	0.00	-0.01	0.00	-0.01	0.00	-0.01	0.00	-0.01
Controls	Lag similarity			0.36	0.75	0.36	0.74	0.37	0.75
	February			-0.09	0.02	-0.09	0.04	-0.05	0.05
	March			-0.02	0.07	-0.02	0.10	0.02	0.11
	Presidential			0.03	0.04	0.03	0.09	-0.02	0.08
	Trump					-0.01	-0.16	-0.03	-0.17
	Clinton							0.12	-0.03
	Rubio							-0.20	-0.07
	Cruz							-0.21	-0.05
	Kasich							0.04	0.00
Model	N	2876957	4306892	2870799	4302821	2870799	4302821	2870799	4302821
	R-sq	0.01	0.03	0.01	0.04	0.01	0.04	0.01	0.04

Table 2: Least-squares estimates of a regression modeling the contemporaneous similarity between Tweets and the transcript as a function of transcript topics, pooled across all debates. Positive coefficients indicate that when a topic was discussed Tweets came in better alignment with the debate. All coefficients are significant at p<.001 except for shaded cells.

Category	Measure		LS Mean	Lower 95%	Upper 95%	Contrast F
Surface Features	**Word Count**	After	15.97	15.95	15.99	15920.20
		During	14.06	14.06	14.07	
	%Video or Photo	After	8.13	8.06	8.20	1235.41
		During	6.14	6.10	6.17	
	%Quotes	After	1.36	1.33	1.38	3399.41
		During	2.48	2.47	2.49	
Linguistic Style	**Specificity**	After	0.42	0.41	0.42	5058.07
		During	0.36	0.36	0.37	
	Analytic	After	78.27	78.17	78.37	7166.40
		During	74.55	74.50	74.59	
	Authentic	After	24.04	23.93	24.14	5268.49
		During	26.01	25.96	26.05	
LIWC Emotion and Drive	**Positive Emotion**	After	3.88	3.86	3.89	3521.76
		During	3.22	3.21	3.22	
	Negative Emotion	After	2.51	2.50	2.53	1679.96
		During	2.86	2.85	2.87	
	%Achievement	After	1.64	1.63	1.65	3734.46
		During	1.11	1.10	1.11	
	%Reward	After	1.74	1.73	1.75	2530.33
		During	1.31	1.31	1.32	
	% Power	After	2.93	2.92	2.95	1286.92
		During	2.86	2.86	2.87	
Topic	**%Policy**	After	4.41	4.31	4.51	6614.11
		During	11.38	11.33	11.43	
	%Contentious Exchange	After	7.03	6.95	7.11	5604.72
		During	5.30	5.26	5.34	
	%Abstract Theme	After	1.69	1.65	1.73	668.98
		During	1.81	1.80	1.83	
	Appearance or Expression	After	15.55	15.45	15.66	2988.85
		During	12.69	12.64	12.73	

Table 3: Mean features of Tweets created during versus after debates. The table shows Tweets created after the debate were likely to be longer, more graphic rich, written in a more analytic style, with a focus on achievement, and were less likely to mention policy issues. All reported F-tests of contrast during versus after the debate are significant at $p<.001$.

	Predictor	Retweets During Estimate	SE	Retweets After Estimate	SE	After-During Change	Pooled SE
User Features	log(followers)	0.350	0.000	0.762	0.003	0.412	0.003
	User Tweets	-0.001	0.000	-0.006	0.000	-0.005	0.000
Surface Features	Word Count	0.017	0.000	0.072	0.001	0.056	0.001
	Graphics	0.372	0.003	1.400	0.017	1.028	0.018
	Quote	0.002	0.000	-0.005	0.001	-0.007	0.001
Linguistic Style	Specificity	-0.010	0.003	0.655	0.018	0.665	0.018
	Analytic	0.000	0.000	0.002	0.000	0.002	0.000
	Authentic	0.000	0.000	-0.002	0.000	-0.002	0.000
LIWC Emotion and Drive	Pos Emotion	0.000	0.000	0.004	0.001	0.005	0.001
	Neg Emotion	0.002	0.000	0.003	0.001	NS	0.001
	Achievement	0.005	0.000	0.067	0.002	0.063	0.002
	Power	0.005	0.000	0.018	0.001	0.012	0.001
	Reward	-0.002	0.000	0.035	0.002	0.038	0.002
Topic	Policy Topic	0.081	0.002	-0.182	0.015	-0.262	0.015
	Cont. Exchange	0.132	0.003	0.902	0.018	0.770	0.018
	Abstract Theme	0.017	0.006	0.222	0.039	0.205	0.040
	Appearance	-0.034	0.002	-0.081	0.016	-0.046	0.017
Controls	Trump	0.036	0.002	0.059	0.011	0.024*	0.011
	Aug Debate	-0.006	0.002	-0.169	0.013	-0.163	0.013
	Feb Debate	-0.022	0.002	-0.402	0.016	-0.380	0.016
	March Debate	-0.011	0.002	-0.119	0.017	-0.107	0.017
Model	Intercept	-1.894	0.003	-9.212	0.025	-7.318	0.025
	Dispersion	0.462	0.001	26.141	0.145	25.679	0.145

Table 4: Parameter estimates of a negative binomial regression of retweets posted during the debate as observed during the debate and after. Positive coefficients indicate that a given feature was associated with higher retweet counts. NS denotes a non-significant difference in coefficients during versus after, * denotes a small effect p<.05. All other differences are significant at p<.001.

Category	Measure		LS Mean	Lower 95%	Upper 95%	Contrast F
Surface Features	**Word Count**	After	16.26	16.21	16.30	3845.34
		During	14.60	14.58	14.62	
	%Video or Photo	After	9.48	9.28	9.68	762.85
		During	6.38	6.29	6.46	
	%Quotes	After	1.38	1.33	1.42	2317.59
		During	2.65	2.63	2.67	
Linguistic Style	**Specificity**	After	0.41	0.40	0.41	578.40
		During	0.38	0.37	0.38	
	Analytic	After	68.07	67.81	68.33	313.99
		During	65.51	65.41	65.62	
	Authentic	After	20.59	20.32	20.85	1656.21
		During	26.57	26.46	26.68	
LIWC Emotion and Drive	**Positive Emotion**	After	3.49	3.45	3.54	14.39
		During	3.59	3.57	3.61	
	Negative Emotion	After	2.37	2.33	2.41	47.97
		During	2.23	2.21	2.24	
	%Achievement	After	1.07	1.05	1.10	258.76
		During	1.30	1.29	1.31	
	%Reward	After	1.18	1.15	1.20	49.42
		During	1.28	1.27	1.29	
	% Power	After	2.79	2.75	2.83	776.58
		During	3.46	3.45	3.48	
Topic	**%Policy**	After	7.46	7.20	7.72	1092.70
		During	12.16	12.05	12.26	
	%Abstract Theme	After	3.59	3.41	3.78	559.05
		During	6.04	5.96	6.11	
	Appearance or Expression	After	14.78	14.50	15.06	34.03
		During	13.87	13.76	13.99	

Table 5: Mean features of Tweets created during versus after Obama's final State of the Union address. Like in the debates Tweets created after the address were likely to be longer, more graphic rich, written in a more analytic style and made fewer references to policy issues. All reported F-tests of contrasts during versus after the address are significant at $p<.001$.

		Retweets During		Retweets After		After-During	
	Predictor	Estimate	SE	Estimate	SE	Change	Pooled SE
User Features	log(followers)	0.363	0.001	0.769	0.006	0.406	0.006
	User Tweets	-0.002	0.000	-0.013	0.000	-0.011	0.000
Surface Features	Word Count	0.012	0.000	0.069	0.002	0.057	0.002
	Graphics	0.627	0.006	1.591	0.033	0.965	0.034
	Quote	0.007	0.000	-0.014	0.002	-0.022	0.002
Linguistic Style	Specificity	0.107	0.006	0.776	0.037	0.669	0.037
	Analytic	0.000	0.000	-0.004	0.000	-0.004	0.000
	Authentic	-0.001	0.000	-0.006	0.000	-0.005	0.000
LIWC Emotion and Drive	Pos Emotion	-0.002	0.000	-0.024	0.002	-0.021	0.002
	Neg Emotion	0.002	0.000	0.016	0.003	0.014	0.003
	Achievement	-0.002	0.001	0.041	0.004	0.043	0.004
	Power	0.003	0.000	0.008	0.002	0.005*	0.002
	Reward	-0.001	0.001	-0.005	0.004	NS	0.004
Topic	Policy Topic	0.090	0.005	0.262	0.031	0.172	0.031
	Abstract Theme	-0.021	0.007	-0.214	0.044	-0.193	0.045
	Appearance	-0.042	0.005	-0.417	0.032	-0.375	0.032
Model	Intercept	-2.052	0.008	-8.726	0.059	-6.674	0.059
	Dispersion	0.437	0.002	15.906	0.185	15.469	0.185

Table 6: Parameter estimates of a negative binomial regression of retweets posted during the debate as observed during the State of the Union Speech and after. Positive coefficients indicate that a given feature was associated with higher retweet counts. All effects and parameter differences are significant at p<.0001; N=449,817.

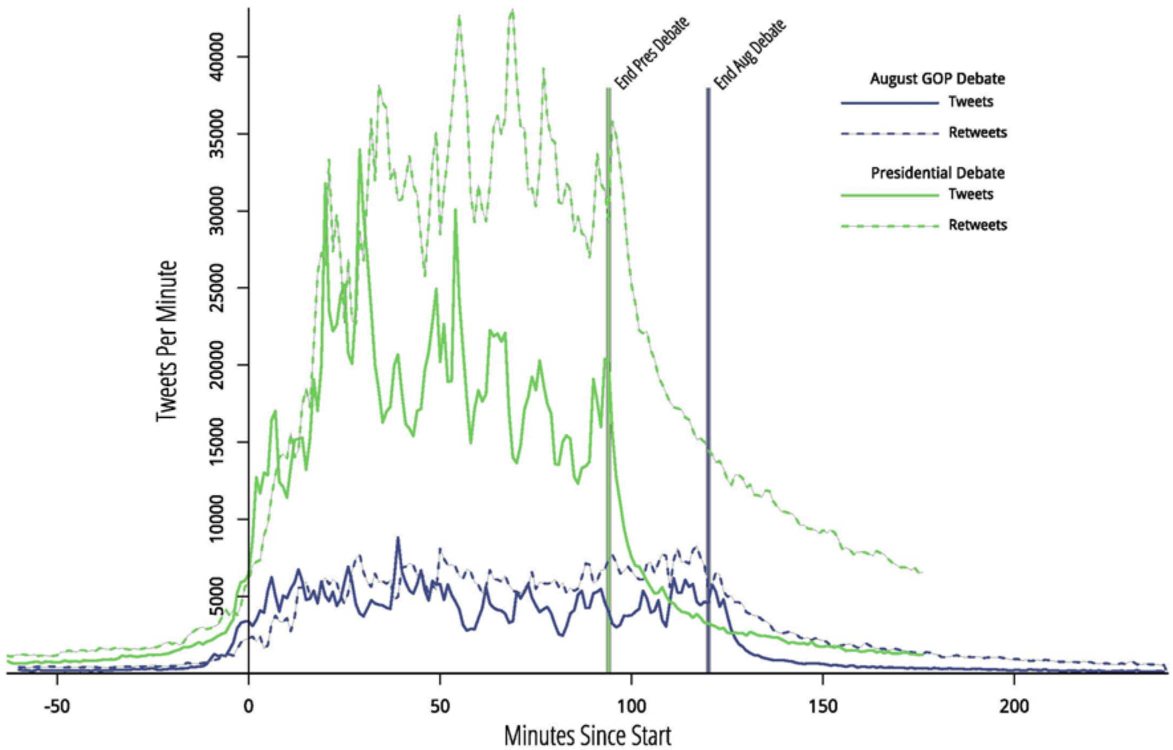

Figure 1: Frequency of Tweets and retweets per minute for the August 2015 GOP primary debate and the October 2016 Presidential debate

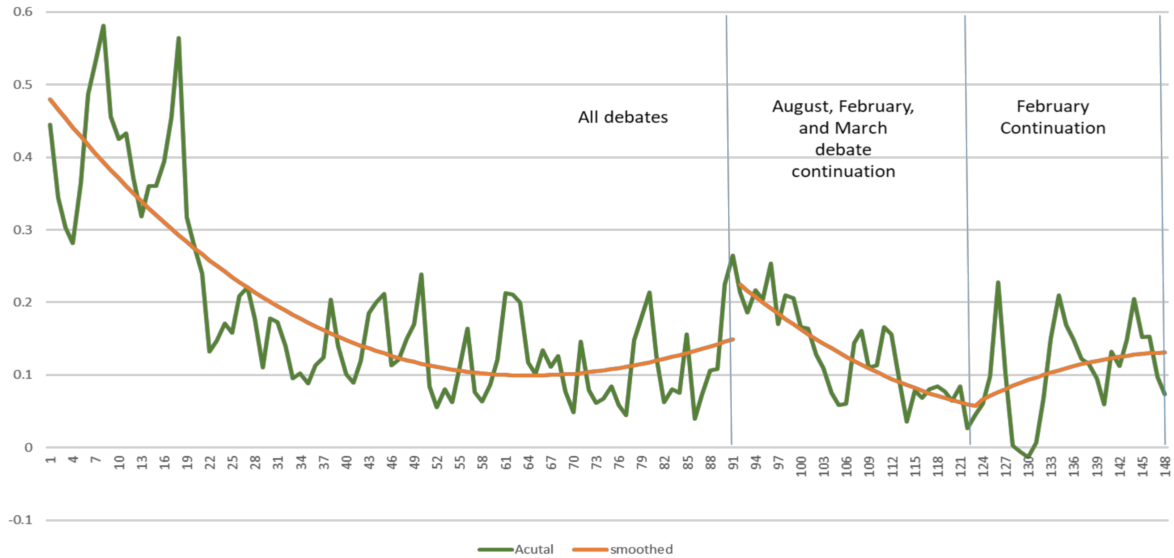

Figure 2: Average actual and polynomial-smoothed cosine similarity between all Tweets and contemporaneous 4-minute transcript windows. Vertical lines indicate the differing end points of the debates

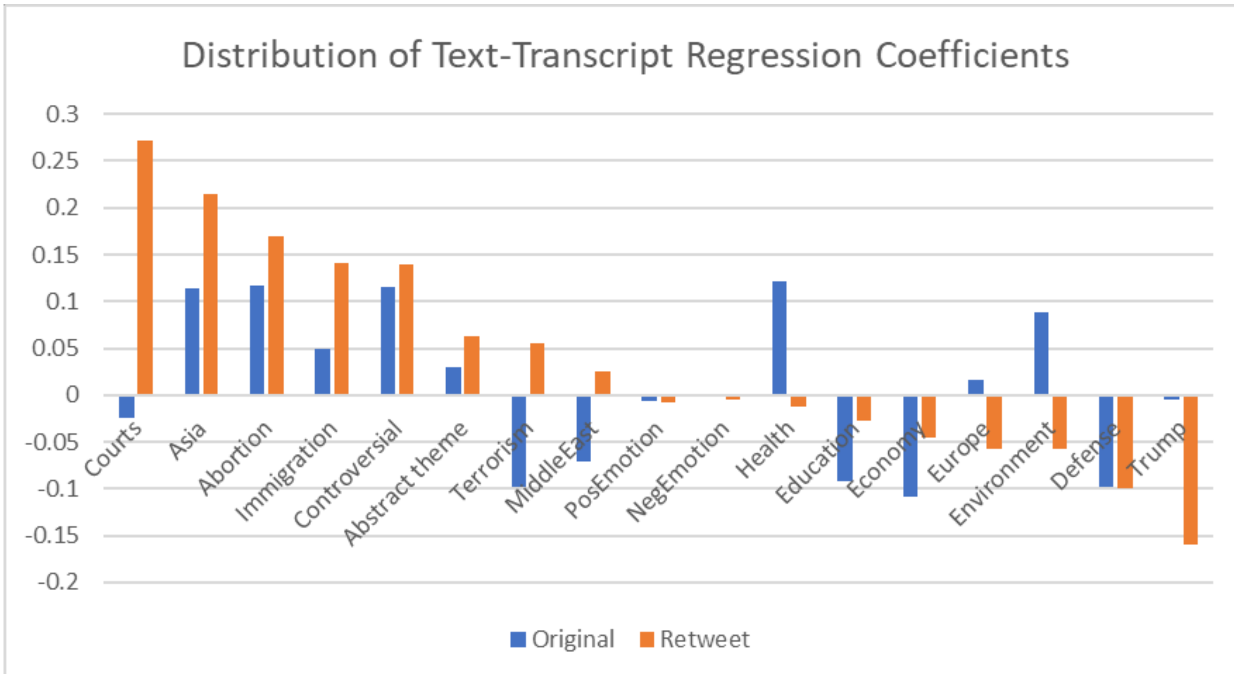

Figure 3: Least-squares estimates of the effect of topic on Tweet-transcript similarity, pooled across all debates (N for retweets=4.30 million; N for original Tweets=2.87 million)

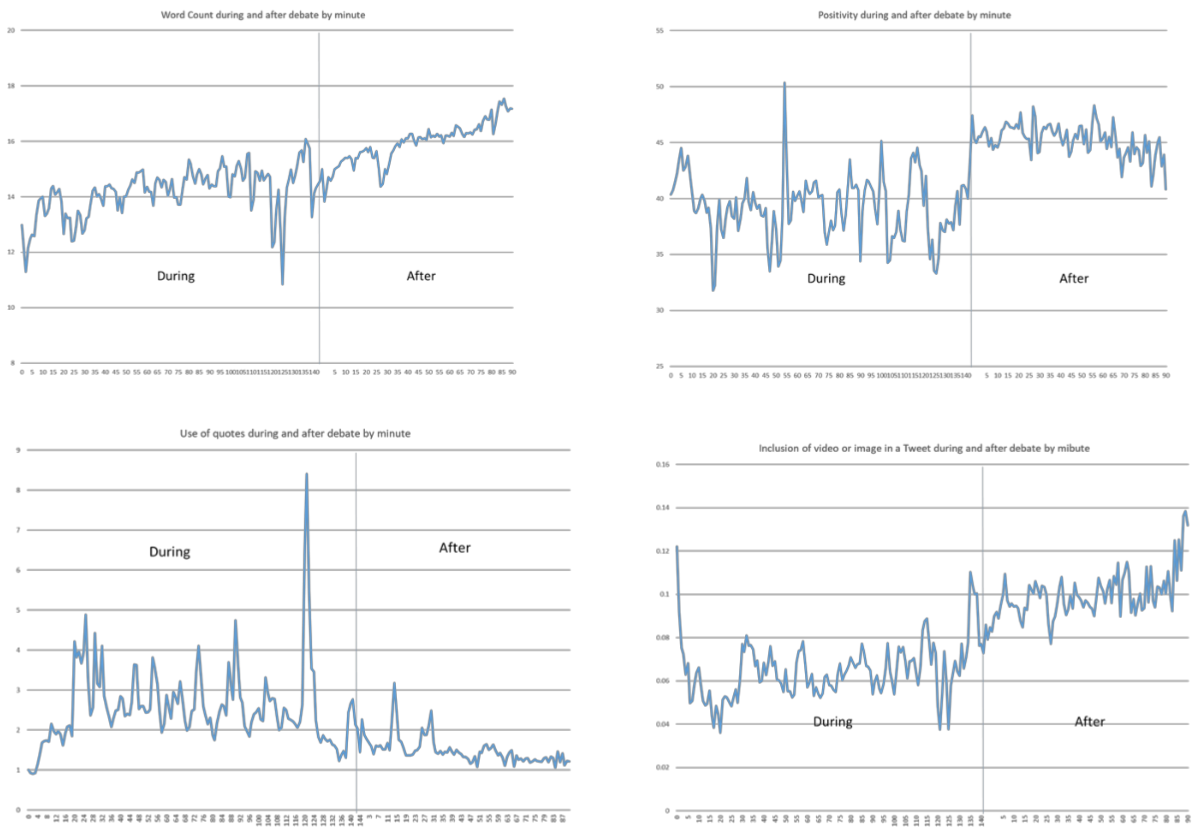

Figure 4: Evolution of four features of newly-created (original) Tweets by minute during and after the debates. Upper left plots word count, upper right sentiment, lower left use of quotes, lower right inclusion of enriched content (video or photo)

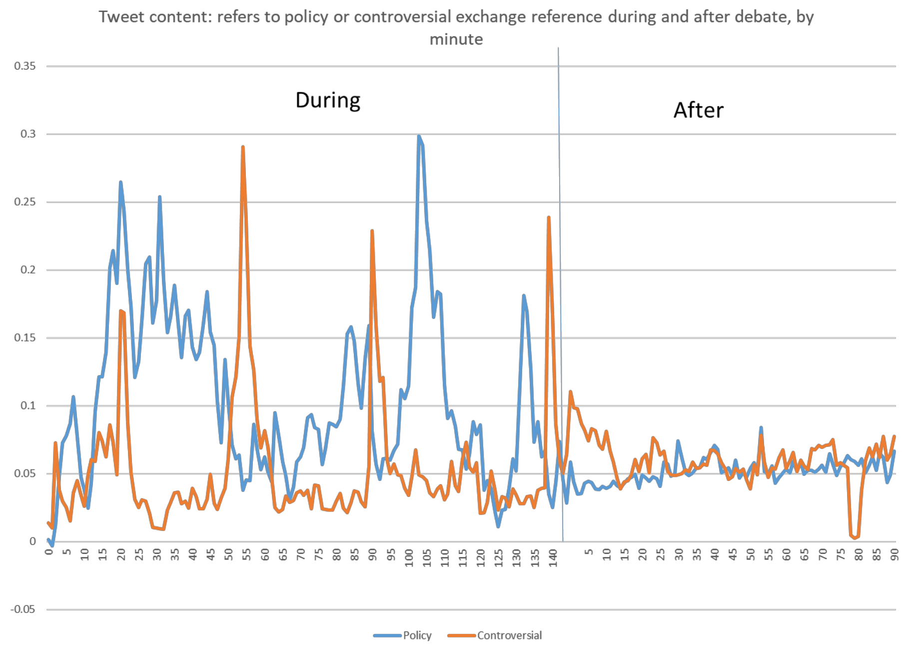

Figure 5: Evolution of the relative frequency with which a Tweet referenced a substantive policy issue (blue) and/or a contentious exchange (orange) by minute during and after the debates