Article

American Behavioral Scientist
2016, Vol. 60(14) 1816–1843

DOI: 10.1177/0002764216676245
abs.sagepub.com
SAGE

Dual Screening During Presidential Debates: Political Nonverbals and the Volume and Valence of Online Expression

Dhavan V. Shah[1], Alex Hanna[1], Erik P. Bucy[2], David S. Lassen[1], Jack Van Thomme[1], Kristen Bialik[1], JungHwan Yang[1], and Jon C. W. Pevehouse[1]

Abstract

The impact of presidential debates on candidate evaluations remains an open topic. Research has long sought to identify the factors that matter most in citizens' responses to debate content, including what candidates say, how they say it, and the manner in which they appear. This study uses detailed codings of the first and third 2012 presidential debates to evaluate the impact of candidates' verbal and nonverbal behaviors on viewers' "second screen" response—their use of computers, tablets, and mobile devices to express their reactions to the viewing experience. To examine the relationship between candidates' on-screen behaviors and the social media response, we conduct generalized least squares regression (Prais–Winstein estimation) relating two data sources: (a) a shot-by-shot content analysis coded for rhetorical/functional, tonal, and visual elements of both candidates' behavior during the debates, and (b) corresponding real-time measures of the volume and valence of online expression about the candidates on Twitter. We find that the nonverbal communication behaviors of candidates—their facial expressions, physical gestures, and blink rate—are consistent, robust, and significant predictors of the volume and valence of public expression during debates, rivaling the power of memes generated by candidates and contributing more than rhetorical strategies and speech tone.

[1]University of Wisconsin–Madison, WI, USA
[2]Texas Tech University, Lubbock, TX, USA

Corresponding Author:
Dhavan V. Shah, Mass Communication Research Center, School of Journalism and Mass Communication, University of Wisconsin-Madison, 5162 Vilas Communication Hall, 821 University Ave., Madison, WI 53706, USA.
Email: dshah@wisc.edu

Keywords
blink rate, computational communication science, facial expressions, gestures, machine learning, meme, political performance, sentiment analysis, Twitter

Since American presidential debates were first televised more than half a century ago, scholars, pundits, and citizens alike have been fascinated with their ability to crystallize political conflict and public discourse. Though there are growing questions about the impact debates have on elections themselves (Benoit, 2013; Druckman, 2003), interest in American televised debates has remained constant. There are few events in the American electoral cycle that focus national attention and public expression on the political process to the extent that televised debates do. In 2012, for example, the first presidential debate between Barack Obama and Mitt Romney drew an audience of more than 70 million Americans—the most in 32 years—a total viewership matched only by the Super Bowl (Carr, 2012).[1] Likewise, the third presidential debate in 2012 drew a total audience of nearly 60 million viewers. And during these two debates' 180-minute total duration, at least 16.5 million debate-related comments were posted to Twitter, making them among the most tweeted-about events in U.S. politics up to that time (Sharp, 2012a, 2012b). This "dual screening" of political events—moving between live broadcast viewing and social media interaction—is gaining academic attention (Vaccari, Chadwick, & O'Loughlin, 2015).

As all of this indicates, debates constitute moments of significant national attention and public expression (through social media) among the electorate. Although debates focus national attention in a powerful way, there is little understanding of how the public responds to these crystallizing moments. That is, there is disagreement among scholars about whether citizens' responses to debates are driven by what the candidates say, how they say it, or how they appear (Benoit, 2013; Cho, Shah, Nah, & Brossard, 2009; Nagel, Maurer, & Reinemann, 2012; Zhu, Milavsky, & Biswas, 1994). Despite scholarly interest in both televised debates and social media, only a handful of studies have examined how the volume and valence of public expression via social media may be shaped by the content of such focal moments (Shah, Hanna, Bucy, Wells, & Quevedo, 2015; Wells, Van Thomme, Maurer, Hanna, Pevehouse, Shah & Bucy, 2016).

Since at least the first televised debates between John F. Kennedy and Richard Nixon, students of presidential campaigns have discussed the role of visual content in political competition, arguing that viewers are heavily "influenced by appearances, gestures, or other nonverbal behaviors" (Kraus, 1996, p. 78). Druckman (2003), for example, found evidence that among subjects with no knowledge of the 1960 debate, the mode in which they first encountered the event—watching the television feed as opposed to only listening to it—affected debate evaluations, increased individual reliance on personality perceptions, and enhanced learning among those with the least preexisting knowledge. Thus, for many Americans, the influence of candidates' self-presentation goes beyond the rhetorical features of persuasive discourse available through their words alone (e.g., attacking, contrasting, or responding to an opponent; Benoit & Harthcock, 1999). Our research builds on this intuition and explores the wide range of verbal, tonal, and visual elements of presidential debates experienced by the audience.

We explore the power of the visuals in presidential debates by combining insights from biobehavioral research with computational methods tracking the public's real-time reaction to the 2012 U.S. presidential debates. Existing studies suggest that a debate audience may respond at least as much to a candidate's physical actions, facial expressions, and autonomic behaviors as they do to his or her words (Grabe & Bucy, 2009; Patterson, Churchill, Burger, & Powell, 1992). In this article, we explore this possibility directly by examining the influence of rhetorical/functional elements alongside voice tone, facial expressions, physical gestures, and blink rates of both candidates during the first and third U.S. presidential debates in 2012. Unlike many other examinations of candidate behavior, we study these elements in real time across debate settings. That is, we code for their existence at the level of the individual segment (ranging from 10 to 30 seconds) during two debates that addressed distinct topics hosted by different moderators.

We then relate differences in segment-level rhetorical/functional, tonal, and visual features to the volume—that is, the actual number of mentions of each candidate on social media or a proxy of the attention a candidate is receiving—and valence—that is, the sentiment posts mentioning each of the candidates or a proxy of the opinion embedded in that expression—of Twitter posts about Obama and Romney on a synchronous and lagged basis. Our social media data comes from an original harvest of 10% of all Twitter content posted during each debate (the Twitter "garden hose"). We then use machine-learning classifiers to score candidate-specific sentiment in all tweets that mention either Barack Obama or Mitt Romney. By comparing the words, tone, and visuals of the candidates to the online expression of debate viewers, we identify the candidate-specific features that most drive responsiveness via the "second screen"—that is, computers, tablets, and mobile devices that enable real-time reactions to the viewing experience (Tsekleves, Cruickshank, Hill, Kondo, & Whitham, 2007).

Literature Review

Our research builds on an extensive body of work examining visual portrayals and candidate nonverbal behavior (e.g., Hellweg & Phillips, 1981; Tiemens, Hellweg, Kipper, & Phillips, 1985) that runs parallel to studies on rhetorical strategies employed by candidates during modern presidential debates (see Benoit, 2013; Jamieson & Birdsell, 1990). As such, this research draws on studies of visual dominance in sensory processing (Colavita, 1974; McGurk & MacDonald, 1976) and considers historical changes in debate presentation and performance that increasingly highlight contentiousness and incivility in such settings.

Debates and the "Second Screen"

In many ways, the ebb and flow of debate formats has followed the evolution of television as a visual medium. The earliest televised debates were shown in black and white from a limited number of candidate perspectives and fixed-shot angles (Kraus, 1996). As the number of television sets and available channels grew rapidly, so too did the technical skill involved in the presentation of the debates. By the early 2000s, a dozen channels were broadcasting presidential debates, including CNN and Fox News, and used a broader variety of camera angles, moving shots, and split-screen displays, providing more visual information, and allowing viewers to continuously monitor the reactions of one candidate while the other was speaking (Cho et al., 2009). These increasingly sophisticated formats may have strengthened debates' well noted—though not always realized—potential to provide a range of information to the electorate (e.g., Jamieson & Adasiewicz, 2000; Zhu et al., 1994). Benoit, McKinney, and Holbert (2001) argue that while debates enhance the transfer of information, they may also change the nature of the criteria citizens use when deciding whether to vote and whom to support.

The evolution of televised media has also changed the nature of political discourse surrounding the debates themselves. Millions of debate viewers in 2012 discussed the event in real time using networked social media (Sharp, 2012a, 2012b). Emerging research suggests that the distracted nature of viewing induced by "second screen" activity may alter viewers' perceptions of candidate performance (Stroud, Stephens, & Pye, 2011). Furthermore, given the tendency for online networks to foster homophily and consensus (Conover et al., 2011), online exchanges during debates may support the finding that debate viewing "largely reinforces existing predispositions rather than substantially changing previously held images of candidates, issue orientations, or voting intentions" (Sigelman & Sigelman, 1984, p. 624; also see Abramowitz, 1978).

Similarly, social media use around the debate may reinforce Wald and Lupfer's (1978) observation that debates increase cynicism and reduce trust. Social media posts are often brief (a single tweet, e.g., can include no more than 140 characters) and may encourage users to focus on broad, generalized assessments of candidate performance, rather than more detailed evaluation of specific policies. This may lend itself to singling out one candidate as the winner of a particular exchange, similar to professional commentators' response to debates (e.g., Fridkin, Kenney, Gershon, & Woodall, 2008; Jamieson, 1992; Tsfati, 2003). Doing so, however, focuses on the strategic nature of candidate behavior and may thereby increase viewer cynicism (Cappella & Jamieson, 1997; Stroud et al., 2011).

Yet as Shah et al. (2015, p. 228) note, "these and most other analyses fail to consider the tonal and visual elements of candidate behavior that also contribute, if not outperform, the influence of candidate statements." Mutz and Reeves (2005; also see Mutz, 2007) examine this question in the context of heated exchanges among pundits. They find that when contention is highlighted through more aggressive exchanges and production decisions, despite holding the content constant, the conflict affects viewer evaluations, further indicating that tonal and visual elements matter.

Presentation and Performance

Production practices may thereby significantly affect the way candidate performances are received and often heighten the sense of conflict perceived by television audiences

(Alexander, 2010; Davis, 1999; Zettl, 1990). To be sure, candidates have relatively little control over the format in which they are presented during a televised debate and must instead work within the constraints broadcast networks impose (Peters, 2012). Mutz (2007) demonstrates this point and argues that the effect of televised incivility—that is, being disrespectful, interruptive, and inattentive—is amplified by close-up camera shots that highlight interpersonal conflict. The split-screen format also emphasizes debate contentiousness and allows "viewers to constantly monitor the words, gestures, and reactions of each candidate, [intensifying the] perception of conflict in much the same manner as close-up camera shots" (Cho et al., 2009, p. 245).

While research on debate format has shown that presentation does matter, the effect of these changes is still not entirely understood. Split-screen shots during debates generally provide more information, allowing the public to better evaluate subtle aspects of the candidates' expressive behavior. Consistently and prominently filling the screen with candidates' faces allows viewers to monitor expressions, demeanor, and gaze, details individuals naturally account for when evaluating politicians' self-presentation. As Jeffrey Alexander (2012, para. 9) notes, "Political performances are also about eyes and energy, about looking and being looked at, about seeming eager and interested and caring." The intimate nature of debate coverage in 2012 facilitated this type of evaluation among Americans.

The first and third debates in 2012 thus offer an ideal opportunity to evaluate the influence of both spoken and nonverbal candidate communication. Camera shots in both debates were consistently tight and simultaneously displayed both candidates on screen. As suggested by prior work concerning format and presentation, viewers during these debates likely used candidates' body language, facial expressions, voice tone, and physical gestures in making political evaluations. Fortunately for our purposes, candidate performance also varied significantly across the two debates, with many commentators suggesting that Romney outperformed Obama in the first debate but that Obama was more successful in the third.[2] These debates thus offer a valuable opportunity to consider the nature and impact of candidate verbal and nonverbal behavior in different moments of the same general election environment, holding the candidates and presentational format constant. In addition to several measures of candidates' rhetorical style, our analysis therefore examines the influence of three general categories of candidates' nonverbal communication: voice tone, expressions and gestures, and blink rate.

Voice Tone

Voice tone is a paralinguistic cue present in all spoken communication. Tone modifies the content of speech by imparting emotion and potentially moderating the meaning of the message, informing audience reactions (Hall, 1980). Tone also signals social intent, expressing a variety of purposes such as disapproval or threat, in the case of an angry tone, or reassurance, as in the case of a friendly one. As Laplante and Ambady (2003) find, tone of voice shapes interlocutors' ratings of politeness or impoliteness. Longitudinal studies of presidential election coverage have found that challengers and debate losers tend to be more aggressive in tone than incumbents and front-runners (Bucy & Grabe, 2008; Grabe & Bucy, 2009), consistent with their tendency to attack (Benoit, 2013).

Expressions and Gestures

Often even more consequential than candidates' tone of voice is the quality of their facial expressions. The face, more than any other expressive feature, serves as the primary conveyance of emotional communication, signaling affective states and behavioral intentions to observers. Researchers identify a few key categories of nonverbal communication displays, two of which are relevant to political communication: happiness/reassurance and anger/threat (Sullivan & Masters, 1988). Happiness/reassurance displays are intended to facilitate a hedonic or friendly mode of social interaction, lowering the probability of an aggressive or antagonistic encounter. Anger/threat displays, on the other hand, are associated with hostile encounters and attempts to unseat the leader of a social group (Bucy & Grabe, 2008). Although reassurance is more commonly associated with effective leadership, in competitive contexts, partisans also respond positively to anger/threat (Masters, Sullivan, Lanzetta, McHugo, & Englis, 1986).

Similar to voice tone, trailing candidates are shown more often in political news coverage as exhibiting anger/threat and making defiant gestures, including finger pointing or wagging, making and brandishing a fist, or shaking their head in disagreement (Grabe & Bucy, 2009). Leading candidates, by contrast, are more likely to be shown engaging in positive affinity behaviors that imply bonding, compassion, or friendship. Less nuanced than facial expressions, hand gestures usually work in tandem with facial expressions, potentially amplifying their effect. The capacity of nonverbal displays to influence decision making depends, at least in part, on candidates' political status and expressive ability, as well as the context of the displays.

In close elections where much is at stake, leader displays are likely to take on added significance. High-status leaders are said to have an "attention-binding" quality that draws continued observance by other members of the social group (whether journalists, party activists, or interested voters). Indeed, Sullivan and Masters (1993) have characterized facial displays *as* leadership behavior. Thus, when the incumbent performs poorly or commits a violation of nonverbal expectations (see Burgoon & Hale, 1988), such as acting in an inappropriate or unexpected manner, we would expect a higher volume of attention to and communication in response to these actions, as well as an outpouring of negatively valenced reactions.

Eye Blink Rate

Another indirect indicator of a candidate's mental and emotional state during a debate is blink rate. Eye blinks typically involve the rapid closing and opening of the eyelids over the eyes (typically within 300-400 milliseconds) in an autonomic fashion, though blinking can also be deliberate. Physiologically, blinking helps provide moisture to the eye and protects the eye from irritants. But blinking has cognitive and emotional dimensions as well. Indeed, three types of eye blinks have been identified—reflex blinks that protect the eye from environmental intrusions, voluntary blinks, and spontaneous or "endogenous" blinks that occur in the absence of identifiable stimuli and are thought to be influenced by perceptions, affective orientation, or information processing (Stewart & Mosely, 2009).

A variety of physical and psychological factors influence this basic semi-autonomic response (Doughty, 2001). Those experiencing task-related fatigue or anxiety, for example, often blink at a faster rate than otherwise similar individuals (Barbato et al., 2007; Harris, Thackray, & Shoenberger, 1966). A number of studies have found that humans generally blink more when concentrating on a task or are under stress to manage dissonance[3] (Oh, Han, Peterson, & Jeong, 2012), especially when they deem the outcome to be important (Karson et al., 1981; Shultz, Klin, & Jones, 2011). As one study put it, the "eye blink [may be] triggered by aspects of information processing, and . . . blink latencies can be used as one tool for evaluating the level of complexity of such processing under a wide variety of task demands" (Fogarty & Stern, 1989, p. 35).

While laboratory research on blink rate has identified the underlying processes these behaviors signal, social scientists have also used blink rate to explore the social information they provide. Several studies have found that individuals monitor the blinking behavior of interaction partners and use it as a meaningful indicator of social information alongside facial expressions (Cunningham & Wallraven, 2009). Our brains instinctively respond to the social relevance of others' blink patterns in a manner similar to other facial features and gaze behavior (Brefczynski-Lewis, Berrebi, McNeely, Prostko, & Puce, 2011). Whatever the ultimate reason, people with unusually rapid blink rates are often perceived as careless, less intelligent, or nervous, while those who blink especially rapidly or slowly are considered "unfriendly" (Omori & Miyata, 2001). Thus, both average blink rate and deviations in blink rate may be noted.

Accordingly, in a study of the 1976 presidential debates, Exline (1985) found that increased blinking by both Gerald Ford and Jimmy Carter negatively affected judgments of candidate competency. Similarly, Patterson et al. (1992) showed that elevated blinking by Walter Mondale in the 1984 presidential debates with Ronald Reagan was associated with lower favorability scores. Thus, excessive blinking and significant variation in blink rate carries evaluative consequences in social as well as political settings.

Hypotheses and Research Questions

In this study, we measure communication behavior in response to actual presidential debates as they occurred, by tracking the volume and sentiment of real-time public expression on Twitter. By measuring actual responses from users of a popular social network, this approach benefits from a high degree of ecological validity. Although Twitter users are not representative of the population, they are nonetheless quite diverse, and their voluminous real-time comments allow us to trace, in a highly granular fashion, the connections between the first and second screens that characterize the television viewing experience in a social media age.

We expect that the tonal and visual elements of the candidates' debate performance, including blink rate, will shape reactions on social media more so than the verbal strategies candidates use to score political points, win favor with the public, and (at times inadvertently) create quotable moments in the form of memes. In light of these considerations, we predict that:

Hypothesis 1: Nonverbal elements of candidates' behavior, specifically (Hypothesis 1a) voice tone and (Hypothesis 1b) visual factors (i.e., facial expressions, physical gestures, and blink rate), will explain differences in the *volume* of online expression about each candidate beyond verbal elements.

Hypothesis 2: Nonverbal elements of candidates' behavior, specifically (Hypothesis 2a) voice tone and (Hypothesis 2b) visual factors (i.e., facial expressions, physical gestures, and blink rate), will explain differences in the *sentiment* of online expression about each candidate beyond verbal elements.

Method

To examine these relationships, we merged two data sets for each debate: (a) a detailed content analysis of the first and third presidential debate between Barack Obama and Mitt Romney that coded for rhetorical/functional, tonal, visual elements and (b) corresponding measures of the volume and sentiment of Twitter posts mentioning the candidates.

Debate Coding

For the verbal, tonal, and visual content analysis, we used C-SPAN's televised coverage of the first and third debates. Throughout the debate, this feed showed both candidates in split-screen format, using a medium shot from the waist up, enabling coding of all nonverbal responses and blink rate. Given that other broadcasters drew on the same nine camera feeds, but were free to use whichever images they chose, it is notable that all the major news outlets—ABC, CBS, NBC, CNN, and Fox News—favored split-screen shots similar to those used by C-SPAN. PBS, by contrast, used a single camera perspective. To standardize analysis across comparable units, the verbal, tonal, and visual aspects of the debate (including blink rates) were coded at the level of the individual camera shot. In cases where segments exceeded 30 seconds (which was typical), they were divided into 30-second increments. For the first 90-minute debate, this resulted in 169 codable segments ranging from 12 seconds to 30 seconds. For the third debate, this resulted in 181 codable segments ranging from 10 seconds to 30 seconds.

Coding of the debates proceeded in two stages. In the first stage of the analysis, the candidates' statements were coded for their primary rhetorical function and "memes"—a reproducible unit of culture that is widely shared (Dawkins, 1976). In the second stage, the candidates' nonverbal behavior was coded for tonal and visual elements, as well as blink rate.

Rhetorical Functions

Major rhetorical functions identified in individual debate segments included *attacks* on the opponent, *contrast statements* that highlighted differences between the speaker and his opponent, direct *responses* to statements (typically, answers to attacks) made by the opponent, and *personal narratives* from the candidate's own experience (see Benoit & Harthcock, 1999; Green & Brock, 2000, for details on these categories).[4]

Memes

As verified by our reading of the news coverage and social media content surrounding each debate, we identified the statements and phrases used to create one or more memes—that is, intended and inadvertent utterances that produced memorable and highly referenced moments. We identified five memes in the first debate, including Romney's promise to cut funding for public television, even though he declared, "I love Big Bird"; and Obama's statement to moderator Jim Lehrer, "I had five seconds before you interrupted me." Similarly, we coded five memes in the third debate, including Obama's "fewer horses and bayonets" comment, and Romney's statement, "I love teachers." Because memes are memorable debate encounters that flow from the utterances of the candidates, they were included with the verbal elements we coded and were simply identified as occurring during the period in which they were uttered. Via digital media, memes can be reproduced almost immediately, and are discussed by pundits in postdebate analyses and rebroadcast in news reports; as such, these simple utterances become highly quotable.

Tonal Elements

In any persuasive encounter, a large part of nonverbal influence stems not just from semantic content but also from voice tone and variability (Bucy & Grabe, 2008). We therefore coded for the presence or absence of two emotion/intention pairs that play a central role in political competition: anger/threat and happiness/reassurance. These categories reflect the felt emotion and presumed behavioral intention of the communicator (Way & Masters, 1996).

Anger/threat was operationalized as statements in which the candidate's tone had a menacing or hostile feel; where the candidate used confrontational verbal tactics to challenge his rival; where the candidate revealed a desire to do political battle, or took exception to and forcefully rebutted a claim by his opponent; or, where the overall tone of a segment could be characterized as enraged, feisty, bold, or aggressive.

Happiness/reassurance was operationalized as statements in which the candidate's tone had an optimistic or cheerful feeling; where the candidate's voice was upbeat, positive, and conveyed an affiliative intent; where the candidate offered hopeful predictions about what will happen to the country if elected; or, where the tone suggested an attempt at bonding or reinforcing a sense of goodwill with potential supporters.

Visual Elements

Next, we coded aspects of the candidates' nonverbal behavior, particularly facial expressions and body language. Consistent with voice tone, we coded for the presence of anger/threat and happiness/reassurance in facial expressions. We also documented the occurrence of gestures that signaled affinity (bonding) or defiance (aggression).

Consistent with the biobehavioral approach to nonverbal communication (see Masters et al., 1986), facial expressions that contained one or more of the following key elements were classified as anger/threat displays: lowered eyebrows, a staring gaze, the visibility of lower teeth, lowered mouth corners (frowning), facial rigidity that showed little to no movement, lips pressed firmly together, or an overall expression that was negative or hostile.

Similarly, happiness/reassurance displays were operationalized using very specific behavioral criteria as expressions containing one or more of the following elements: a smile with relaxed mouth position, the visibility of upper or both rows of teeth, nodding up and down, brief eye contact to avoid staring, open or just slightly closed eyes, "Crow's feet" wrinkles around the eyes, or an overall expression that was welcoming.

Gestures were coded as body language that signaled affinity or defiance (see Grabe & Bucy, 2009). Affinity gestures consisted of hand, body, or facial movements that suggest a friendly relationship or attempt at bonding between the candidate and the audience, opponent, or moderator. Examples include waving or giving a "thumbs-up"; winking or nodding knowingly to the camera, moderator, or other candidate; or, using an open palm when referencing the audience or opponent (rather than a closed fist or pointed finger).

Defiance gestures consisted of hand, body, or facial movements that suggest a threatening or antagonistic relationship between the candidate and their opponent. Examples include finger pointing, wagging, or shaking; raising a fist; shaking one's head in disagreement; negative expressions accompanied by prolonged stares; or, other behaviors signaling aggression. Frequencies of these manifest tonal and visual elements are provided in Table 1.

Blink Rate

In addition, eye blinks of both candidates were manually coded at 5-second intervals, consistent with previous research (see Stewart & Mosely, 2009). Altogether, this resulted in 1,060 codable segments for Debate 1 and 1,083 segments in Debate 3. Obama (M = 4.55, SD = 2.98) exhibited a slightly higher average rate of blinking than Romney in Debate 1 (M = 4.09, SD = 2.31) as well as Debate 3 (Ms = 5.22, 4.64, SDs = 1.73, 1.31, for both candidates, respectively). Obama showed more variability than Romney, with variance scores of 8.87 and 3.01 in Debates 1 and 3, compared with Romney's 5.34 and 1.72. The maximum number of blinks in any given segment was also highest for Obama, with 13 in Debate 1 (compared with 10 for Romney). Both candidates had a maximum number of 10 blinks in any segment in Debate 3. To facilitate analysis with our other measures, blink rate data were collapsed into corresponding

Table 1. Nonverbal Display Frequencies, Key Variables.

	Debate 1		Debate 3	
	Obama	Romney	Obama	Romney
Verbal tone				
Anger/threat	32.4% (60)	43.8% (80)	44.2% (80)	49.7% (90)
Happiness/reassurance	24.9% (46)	21.6% (40)	22.1% (40)	31.5% (57)
Facial display				
Anger/threat	23.8% (44)	35.7% (66)	35.9% (65)	32.6% (59)
Happiness/reassurance	43.2% (80)	40.0% (74)	16% (29)	37% (67)
Body gestures				
Affinity	6.5% (12)	16.2% (30)	10.5% (19)	28.2% (51)
Defiance	14.6% (27)	36.8% (68)	37.6% (68)	17.1% (31)

Note. Percentages may not add up to 100 because more than one display type may occur in any coded segment. Frequency counts in parentheses.

segments (which ranged from 10 to 30 seconds) to parallel the verbal, tonal, and visual variables. For each segment, average blink rates and standard deviations were calculated for candidates.

Intercoder Reliability

For the first debate, the presence or absence of each verbal, tonal, and visual element was coded for in each shot or 30-second segment. For all 169 codable segments, each element was coded 1 if present and 0 if absent. A similar procedure was followed in the third debate, with each verbal, tonal, and visual element rated for the 181 codable segments, with the exception that visual displays were also coded for degree of intensity (2 = *vivid display*, 1 = *visible display*, 0 = *no display*) given the generally negative and aggressive affect of both candidates throughout much of the third debate.

For the verbal elements, a second trained coder assessed each of the 350 retained segments, agreeing on all but 88 of the 3,500 individual codings and producing initial Krippendorff's alpha values ranging from .89 for Obama's use of contrast to .99 for Romney's use of response. Instances of disagreement were discussed between coders until a consensus was achieved. For the tonal and visual variables, 11% of the debate footage, or 40 segments, were assessed by a second coder. Tonal variables had a similarly high level of agreement, with α values of .91 and 1 for Obama and Romney's anger/threat and happiness/reassurance voice tone, respectively. Gestures' α values ranged from .90 for Obama's defiance gestures to 1 for Romney's defiance gestures. For facial expressions, again of anger/threat and happiness/reassurance, there was .95 agreement, which dropped to α values of .83 and .78. Reliability testing of the blink rate data was conducted on about 15% of the segments from each debate (152 from Debate 1 and 219 from Debate 3) by two trained coders. Intercoder reliability was high, with α values of .91 for Debate 1 and .94 for Debate 3.

Twitter Harvesting

For this study, we archived the Twitter garden hose as a sample of social media activity through its Streaming API.[5] Twitter describes the garden hose as a continuous 10% sample of the 300 to 500 million global tweets per day. Tweet information includes a variety of different fields, including user information, tweet time, geolocation (if available), and platform used to post the tweet.[6] From this archive, we drew posts within a 50-day window before the election, from September 19, 2012, to November 8, 2012, that included "Obama," "Romney," "Biden," or "Paul Ryan" (not case-sensitive). We then identified all tweets from our sample that referenced Obama or Romney and were posted during either the first (October 3) or the third (October 22) debate. We generated volume measures from mentions of only "Obama" or only "Romney" and not both within the text of the tweet. While this undercounts the total candidate mentions, our intent was to label whether a tweet was about a particular candidate and to understand over-time variability in those mentions.

Sentiment analysis was performed using a supervised machine-learning method, in which a function for inferring the positivity or negativity of a post was generated by the labeled training data. To generate a training set, we randomly sampled 1,000 tweets from each debate period, 500 mentioning Romney and 500 mentioning Obama. Two coders then coded these tweets for three possible sentiment values: positive, negative, or neutral. After coding separately, coders met to resolve any disagreements. In the first debate, we found that most tweets (87%) were either positive or negative. The final training set for sentiment analysis of the first debate therefore used only this subset of tweets that were coded as positive or negative, numerically represented as 1 and −1, respectively. In the third debate, we used all three values to assign a sentiment score. Last, we trained a regularized logistic regression classifier using the Python scikit-learn library. The classifier obtained an F1-score of .77 for the first debate and a F1-score of .71 for the third debate, both within acceptable levels of accuracy (Yang & Liu, 1999).

To align Twitter activity to archived video material of the debate, we identified points during both debates where we knew there was high Twitter volume: Romney's quip, "I love Big Bird"; Obama's "I had five seconds" remark; Romney's "apology tour" comment, and Obama's "horses and bayonets" retort. Assuming a small gap between broadcast and Twitter reaction, we then used the first mention of these sound bites as the measure of the lag. We found a consistent gap between the debate clock on the C-SPAN feed and the UTC (Coordinated Universal Time) timestamp on the Twitter posts. Accordingly, we synchronized Twitter data to the debate feed. Each coded shot of the debate was the unit of observation and analysis, with the volume and sentiment of expression on Twitter normed by the length of the unit.

For each shot, we next generated a synchronous volume metric and average sentiment score for both candidates, along with lagged versions of these variables, at 15, 30, and 45 seconds. These lagged values account for any delayed reaction by Twitter users and observe the robustness of observed effects. This resulted in the generation of 16 variables for each debate, matched or lagged to the start-stop times on the debate

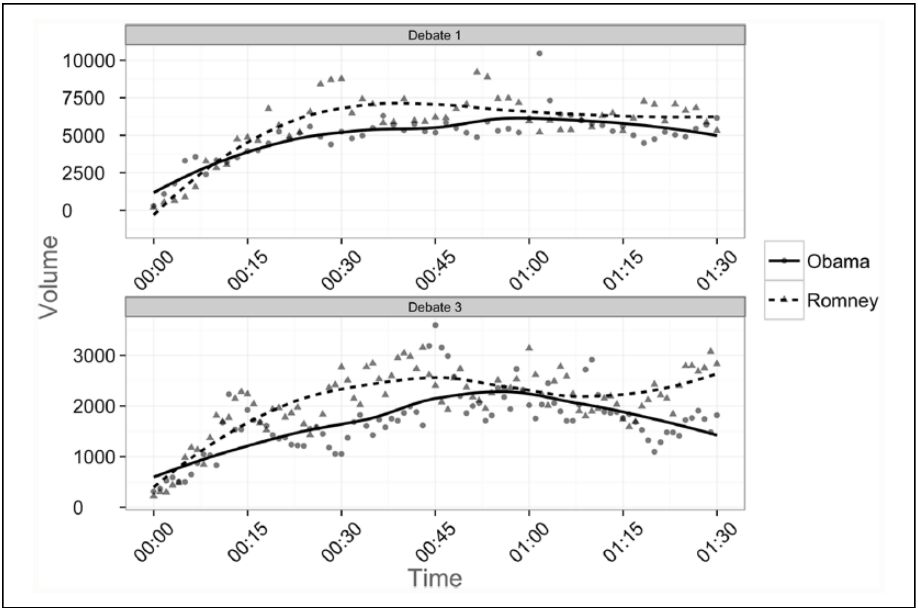

Figure 1. Volume per minute by candidate, with LOESS regression smoothed average.

clock (volume and sentiment for Obama and Romney synchronous to the debate clock, and at 15-, 30-, and 45-second lags). Because the results at 30 seconds are very consistent with the 15-second lag, we omit presentation of those analyses, opting to make them available as an online appendix (available at http://abs.sagepub.com/content/by/supplemental-data).

Figure 1 shows the volume per minute by candidate for each debate, while Figure 2 shows the average sentiment per minute by candidate. The lines on each graph represent LOESS regression smoothed averages for the candidates over their data points, which provide the minute-by-minute averages. The plots in Figure 1 are comparable with Twitter's own graphs of total debate volume, not distinguished by candidate.[7] The sentiment plots reveal that Twitter favored Obama over Romney during both debates, a result that held throughout each event, albeit with considerable variation over time. Notably, both candidates averaged sentiment scores below 0, indicating generally negative expression on this social media platform.

Analysis

Before testing our hypotheses, we thought it important to verify that the debates were indeed moments of national attention and expression. To do so, we charted the volume of keyword references to Obama, Romney, Biden, and Ryan during the 50-day window of our archive draw. The results are presented in Figure 3 and clearly indicate that

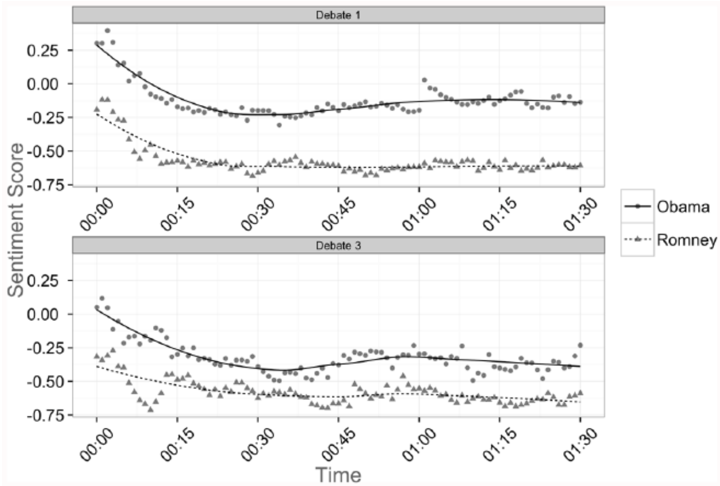

Figure 2. Sentiment per minute by candidate, with LOESS regression smoothed average.

the debates were high-intensity moments during the campaign, with the first debate between Barack Obama and Mitt Romney being the highest volume moment other than Election Day itself.

Next, to estimate the effects of verbal, tonal, and visual factors on the volume and valence of online expression about the debates, we estimate time series regression models that take into account the autoregressive nature of the data-generating process. Durbin–Watson tests on each of our dependent variables reveal significant autocorrelation. To take account of this temporal dependence, we use generalized least squares (Prais–Winstein) regression to estimate our models. These models adjust the variance–covariance matrix to account for a first-order autoregressive process in the error term of the model, allowing for a more conservative test of our hypotheses. We calculate the adjustment (rho) using the single-lag ordinary least squares estimate of the residuals from the original estimating equation. We begin with saturated models with all predictors included. We then removed variables that were consistently not significant predictors across the models, leading us to drop candidates' use of contrast, response, and narrative verbal strategies and happiness/reassurance facial expressions. Retaining these variables yields no changes on the model estimates discussed below.

We conducted these regression analyses on the normalized volume measures (ratio of number of posts to seconds in the segment) and average sentiment scores (the mean value of scores across seconds in the segment) of the tweets mentioning Obama or Romney in Debate 1 followed by Debate 3. As noted above, these analyses were run with two different versions of the dependent variables: synchronous and 15 seconds after the segment.

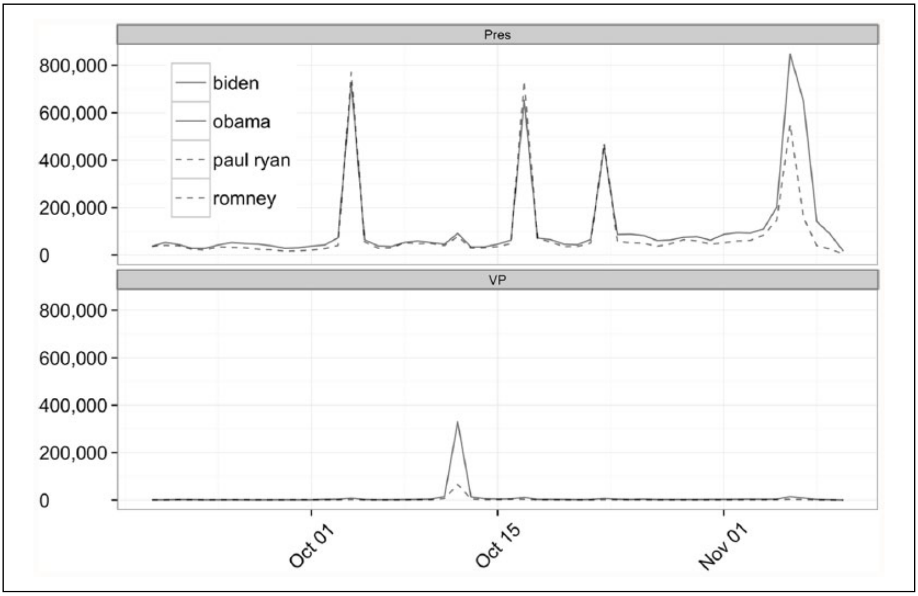

Figure 3. Volume of name mentions of presidential and vice-presidential candidates on Twitter.

For each debate, independent variables were grouped in three blocks that were sequentially added to each regression: verbal, tonal, and visual. Models separately tested the volume and valence of mentions for each candidate. The first block consisted of candidates' use of attacks and indicators of whether either candidate sparked a meme during the shot. The next block added variables assessing each candidate's voice tone (anger/threat, happiness/reassurance). The third block added Obama's and Romney' facial expressions (anger/threat), affinity and defiance gestures, average blink rate, and standard deviation of that blink rate per segment. Both blink rate variables were included because it is not only the rate but also large deviations that are notable.

Results

The results of the generalized least squares regression analyses present a clear picture of the factors that shaped the volume and valence of responses on Twitter. Beginning with the volume of Obama mentions in the first debate, the results are consistent across the synched and 15-second lag models (see Table 2; columns 1 and 2). The verbal factors account for a sizable amount of variance in the volume of Obama mentions, with *F*-change scores significant in the synchronized tests, accounted for mainly by Obama's generation of memes and to a lesser degree by Romney's generation of memes. Comparatively, when the tonal block is added, the models' performance does not

Table 2. Synchronous and Lagged Models Predicting Normalized Volume of Obama and Romney Mentions for Debate 1—Generalized Least Squares (Prais–Winstein) Regression.

	(1)	(2)	(3)	(4)
	Obama volume, synchronous	Obama volume, 15-second delay	Romney volume, synchronous	Romney volume, 15-second delay
Obama Attack	−0.619 (1.045)	−1.042 (1.321)	−0.633 (1.441)	−1.928 (1.447)
Romney Attack	0.00229 (0.966)	−0.446 (1.221)	−0.391 (1.332)	−1.082 (1.337)
Obama Meme	37.81*** (3.809)	10.97** (4.814)	0.916 (5.249)	0.103 (5.272)
Romney Meme	−4.475* (2.605)	2.927 (3.292)	−13.36*** (3.590)	1.972 (3.605)
Obama Tone—Angry/Threatening	0.508 (0.834)	1.462 (1.053)	−1.886 (1.150)	−2.236* (1.155)
Obama Tone—Happy/Reassuring	0.402 (0.809)	−0.835 (1.022)	−2.461** (1.115)	−3.148*** (1.120)
Romney Tone—Angry/Threatening	−1.232 (0.953)	−2.426** (1.204)	−1.825 (1.314)	−2.002 (1.320)
Romney Tone—Happy/Reassuring	−1.502 (0.960)	−0.387 (1.213)	−0.110 (1.324)	0.812 (1.329)
Obama Facial—Angry/Threatening	1.320*(0.707)	0.218 (0.893)	1.366 (0.974)	0.187 (0.978)
Obama Affinity Gesture	−1.087 (1.430)	−0.552 (1.805)	0.0912 (1.972)	−0.569 (1.980)
Obama Defiance Gesture	−0.0839 (0.901)	1.926* (1.139)	−0.235 (1.243)	−0.0744 (1.248)
Romney Facial—Angry/Threatening	1.930 (1.173)	2.143 (1.481)	0.554 (1.619)	1.025 (1.625)
Romney Affinity Gesture	0.913 (0.898)	1.406 (1.135)	1.631 (1.238)	0.726 (1.243)
Romney Defiance Gesture	−0.281 (0.996)	1.041 (1.258)	1.404 (1.373)	3.230** (1.378)
Obama Blink Rate—Average	0.199 (0.186)	0.0300 (0.235)	−0.177 (0.257)	−0.280 (0.258)
Romney Blink Rate—Average	0.263 (0.304)	−0.0783 (0.383)	0.146 (0.420)	−0.152 (0.422)
Obama Blink Rate—*SD*	0.143 (0.439)	−0.654 (0.554)	0.519 (0.605)	−0.442 (0.608)
Romney Blink Rate—*SD*	−1.252** (0.488)	−0.617 (0.617)	−0.00201 (0.674)	0.973 (0.676)
Constant	48.57*** (3.968)	51.59*** (4.463)	58.41*** (6.172)	60.52*** (5.838)
Observations	169	169	169	169
R^2	0.446	0.126	0.121	0.107
Block F tests				
Verbal	26.45***	1.78	3.54***	0.69
Tonal	1.18	1.83	2.19*	3.42**
Visual	1.47	1.10	0.73	1.07

Note. Standard errors in parentheses.
***p < .01. **p < .05. *p < .1.

improve substantially, with F-change scores insignificant for both tests. Only an angry or threatening tone by Romney was negatively related to volume of Obama mentions. Likewise, the addition of the visual block does not contribute significantly to the models' performance across both F-change tests. Nonetheless, Obama's angry or threatening facial expressions and defiance gestures contribute to the synchronized and lagged models, respectively. The analysis also indicates that the variability of the blink rate within segment reduced the reaction in terms of the volume of social media posts.

Moving next to the volume of Romney mentions during the first debate (see Table 2; columns 3 and 4), the verbal factors account for significant variance in the immediate volume of Romney mentions, with F-change tests achieving statistical significance, driven mainly by Romney's memes. The tonal block adds to overall performance of the models, with F-change scores significant in both tests, largely a function of the use of a happy or reassuring tone by Obama. The addition of the visual block does not contribute to the performance of either model. Romney's angry/threatening facial expressions and defiance gestures account for most of the explained variance within the block, although the latter barely miss conventional levels of statistical significance. These results, in combination with those observed for Obama volume, provide some support for Hypothesis 1 in the context of the first debate.

For the sentiment analysis of posts mentioning Obama during the first debate, we again ran the synched and 15-second lag models (see Table 3; columns 1 and 2). The verbal factors account for a significant amount of variance in Obama sentiment, with the F-change statistic significant for the 15-second lag model, driven mainly by Obama's and Romney's memes. The tonal block contributes minimally, failing to reach significance in any of the tests. Conversely, the visual block contributes significantly, again in the 15-second lag model based on the F-change test, mainly a function of Obama's gestures, both affinity and defiance.

These findings contrast somewhat with the models predicting sentiment of Romney mentions (see Table 3; columns 3 and 4). The verbal and tonal factors account for little variance in the sentiment scores for Romney, failing to achieve significance in any F-change tests. In sharp contrast, the addition of the visual block contributes substantially and significantly to the synchronous models' performance. Two visual factors stand out: Obama's gestures and Romney's blink rate. These results, in combination with those observed for Obama sentiment during this first debate, provide some support for Hypothesis 2, especially concerning the visual elements.

Turning to the volume of Obama mentions in the third debate, the results are highly consistent across the synched and 15-second lag models (see Table 4; columns 1 and 2). Yet, unlike the first debate, it is not Obama's memes but his *attacks* that account for significant variance explained by verbal factors when predicting the volume of Obama mentions. Notably, the F-change is significant in both models. When the tonal block is added, the models' performance does not improve. Yet, unlike the first debate, the addition of the visual block contributes considerably to the models' performance, with F-change scores significant for both models. This result is largely due to Obama's angry or threatening facial expressions and his use of affinity gestures. Romney's blink rate is also a significant predictor, contributing to the synched and lagged

Table 3. Synchronous and Lagged Models Predicting Sentiment of Obama and Romney Mentions for Debate 1—Generalized Least Squares (Prais–Winstein) Regression.

	(1)	(2)	(3)	(4)
	Obama sentiment, synchronous	Obama sentiment, 15-second delay	Romney sentiment, synchronous	Romney sentiment, 15-second delay
Obama Attack	0.0131 (0.0104)	0.00994 (0.00909)	−0.0118 (0.00890)	0.00109 (0.00842)
Romney Attack	0.00108 (0.00959)	−0.000773 (0.00838)	0.00164 (0.00819)	0.00151 (0.00774)
Obama Meme	0.0676* (0.0379)	0.107*** (0.0331)	0.0596* (0.0325)	0.0487 (0.0307)
Romney Meme	−0.00967 (0.0259)	−0.0599*** (0.0226)	0.00259 (0.0222)	−0.00323 (0.0210)
Obama Tone—Angry/Threat	−0.00121 (0.00825)	−0.00856 (0.00722)	0.00765 (0.00702)	−0.000279 (0.00663)
Obama Tone—Happy/Reassuring	0.00775 (0.00800)	−0.00750 (0.00701)	0.00105 (0.00682)	−0.00309 (0.00645)
Romney Tone—Angry/Threat	0.0103 (0.00943)	0.00280 (0.00825)	−0.00600 (0.00803)	−0.00242 (0.00759)
Romney Tone—Happy/Reassuring	0.00456 (0.00950)	0.00398 (0.00832)	−0.00495 (0.00808)	−0.00795 (0.00764)
Obama Facial—Angry/Threat	0.00533 (0.00702)	−0.00392 (0.00614)	0.00158 (0.00600)	−0.00135 (0.00567)
Obama Affinity Gesture	0.0244* (0.0141)	−0.0413*** (0.0123)	0.0366*** (0.0120)	0.0140 (0.0113)
Obama Defiance Gesture	−0.0183** (0.00893)	−0.00968 (0.00781)	0.0190** (0.00762)	0.0127* (0.00720)
Romney Facial—Angry/Threat	−0.00174 (0.0115)	0.00928 (0.0101)	0.00127 (0.00976)	−0.0153* (0.00922)
Romney Affinity Gesture	0.00919 (0.00890)	0.00447 (0.00779)	−0.00386 (0.00759)	0.00580 (0.00717)
Romney Defiance Gesture	0.00158 (0.00986)	−0.00525 (0.00863)	−0.00384 (0.00840)	−0.00580 (0.00794)
Obama Blink Rate—Average	−0.00120 (0.00181)	−0.00133 (0.00160)	−0.000431 (0.00152)	−0.00102 (0.00144)
Romney Blink Rate—Average	0.00231 (0.00291)	−0.00116 (0.00258)	−0.00460* (0.00242)	−0.00424* (0.00228)
Obama Blink Rate—*SD*	−0.000911 (0.00435)	0.00285 (0.00381)	−0.00221 (0.00371)	−0.00193 (0.00351)
Romney Blink Rate—*SD*	0.00220 (0.00482)	0.00660 (0.00422)	−0.00627 (0.00410)	−0.00601 (0.00387)
Constant	−0.215*** (0.0253)	−0.201*** (0.0241)	−0.606*** (0.0195)	−0.591*** (0.0182)
Observations	169	169	169	169
R^2	0.071	0.226	0.274	0.330
Block F tests				
Verbal	1.19	4.85***	1.36	0.64
Tonal	0.59	0.75	0.61	0.36
Visual	1.17	1.76*	2.04**	1.47

Note. Standard errors in parentheses.
***$p < .01$. **$p < .05$. *$p < .1$.

Table 4. Synchronous and Lagged Models Predicting Normalized Volume of Obama and Romney Mentions for Debate 3—Generalized Least Squares (Prais–Winstein) Regression.

	(1)	(2)	(3)	(4)
	Obama volume, synchronous	Obama volume, 15-second delay	Romney volume, synchronous	Romney volume, 15-second delay
Obama Attack	2.864*** (0.924)	2.634*** (0.913)	1.626 (1.243)	1.763 (1.262)
Romney Attack	−1.437 (1.218)	−1.059 (1.204)	1.611 (1.641)	1.754 (1.666)
Obama Meme	−0.764 (2.463)	0.536 (2.434)	−−2.167 (3.324)	0.429 (3.374)
Romney Meme	−2.674 (2.380)	1.453 (2.352)	0.708 (3.198)	2.542 (3.246)
Obama Tone—Angry/Threat	1.122 (0.735)	0.765 (0.726)	0.402 (0.987)	−0.432 (1.002)
Obama Tone—Happy/Reassuring	0.444 (0.761)	0.442 (0.752)	1.066 (1.025)	1.335 (1.040)
Romney Tone—Angry/Threat	0.104 (0.769)	−0.894 (0.760)	−1.133 (1.033)	−2.158** (1.048)
Romney Tone—Happy/Reassuring	0.395 (0.691)	−0.190 (0.683)	0.340 (0.928)	−1.413 (0.941)
Obama Facial—Angry/Threat	−1.115* (0.575)	−1.678*** (0.568)	−1.284* (0.774)	−1.349* (0.785)
Obama Affinity Gesture	1.894** (0.821)	1.999** (0.812)	1.329 (1.104)	0.743 (1.121)
Obama Defiance Gesture	0.407 (0.605)	0.0453 (0.598)	−0.00552 (0.815)	0.208 (0.828)
Romney Facial—Angry/Threat	0.840 (0.607)	−0.253 (0.600)	0.109 (0.819)	0.188 (0.831)
Romney Affinity Gesture	0.230 (0.679)	−0.151 (0.671)	−0.128 (0.910)	0.563 (0.923)
Romney Defiance Gesture	0.121 (0.773)	0.136 (0.764)	2.150** (1.039)	3.115*** (1.054)
Obama Blink Rate—Average	0.0194 (0.321)	−0.250 (0.317)	0.0881 (0.429)	0.306 (0.435)
Romney Blink Rate—Average	−1.053** (0.417)	−1.052** (0.412)	−0.655 (0.559)	−0.995* (0.567)
Obama Blink Rate—*SD*	−1.204* (0.652)	−0.223 (0.644)	−0.504 (0.879)	−0.0653 (0.893)
Romney Blink Rate—*SD*	0.670 (0.618)	0.368 (0.611)	0.137 (0.832)	–0.156 (0.845)
Constant	34.07*** (3.573)	36.26*** (3.525)	40.31*** (4.140)	41.47*** (4.180)
Observations	180	180	180	180
R^2	0.148	0.159	0.063	0.133
Block F tests				
Verbal	2.97**	2.71**	0.70	0.94
Tonal	0.65	0.77	0.59	2.31*
Visual	1.87*	2.08**	1.20	1.78*

Note. Standard errors in parentheses.
***p < .01. **p < .05. *p < .1.

models. We want to be careful not to read too deeply into the differences between the two debates, but the increasing potency of the visual features is notable.

Moving next to the volume of Romney mentions during the third debate, the results are fairly consistent across the synched and 15-second lagged models (see Table 4; columns 3 and 4). The verbal factors account for little variance in the immediate or lagged measures of Romney mentions, with F-change tests failing to achieve significance in either model. The tonal block adds to overall performance in the lagged model, driven by Romney's expressions of anger and frustration. The addition of the visual block contributes substantially to the models' performance, but again, only in the lagged model. Romney's defiance gestures account for most of the explained variance, linked to greater volume of Romney mentions, whereas Obama's angry and threatening facial expressions and Romney's heightened blink rate yields fewer mentions, the latter in 15-second lag model only. These results, in combination with those observed for Obama volume, provide strong support for Hypothesis 1, particularly for the visual elements in the third debate, building on results from the first debate. Indeed, visual elements appear to play a somewhat larger role in third debate than in first debate.

For the sentiment of Obama posts during the third debate, we again ran the synched and 15-second lag models (see Table 5; columns 1 and 2). The verbal factors account for immediate variance in Obama sentiment, with F change significant for the synched model, driven mainly by Obama's and Romney's use of attacks. The tonal block does not consistently contribute variance across the models, though Romney's angry or threatening tone is negatively related to positive sentiment. The addition of the visual block contributes significantly to the synchronous model, with Obama's angry/threatening facial expressions and his blink rate as individual predictors.

The models predicting sentiment of Romney mentions (see Table 5; columns 3 and 4) in the third debate do not perform well. The verbal factors account for minimal variance in the sentiment scores for Romney, failing to achieve significance in any F-change tests. The addition of the tonal block is significant in just one case, for the synchronized model, with Romney's angry and threatening tone diminishing sentiment in the synched model but enhancing it in the lagged model. In contrast, the addition of the visual block does not contribute to the models' performance, with few individual predictors contributing significantly, namely Obama's angry/threatening facial expressions and defiance gestures enhancing Romney's sentiment. These results, in combination with those observed for Obama sentiment, provide limited support for Hypothesis 2 for the third debate, with nonverbal elements occasionally predicting sentiment. The lower potency of nonverbal elements in explaining the sentiment of responses in the third debate compared with the first debate runs counter to the pattern observed for the volume of mentions.

Taken as a whole, however, the results between these two candidates across two different debates indicate that the candidates' nonverbal communication shaped the volume and valence of expression on Twitter, providing somewhat stronger support for Hypothesis 1 than for Hypothesis 2. Without overemphasizing the differences between the two debates, attention across the tests of incremental variance explained

Table 5. Synchronous and Lagged Models Predicting Sentiment of Obama and Romney Mentions for Debate 3—Generalized Least Squares (Prais–Winstein) Regression.

	(1)	(2)	(3)	(4)
	Obama sentiment, synchronous	Obama sentiment, 15-second delay	Romney sentiment, synchronous	Romney sentiment, 15-second delay
Obama Attack	−0.0256** (0.0119)	0.00436 (0.0126)	−0.00611 (0.00995)	−0.00574 (0.0102)
Romney Attack	0.0283* (0.0157)	0.0224 (0.0166)	0.00441 (0.0131)	0.00683 (0.0135)
Obama Meme	−0.0260 (0.0317)	0.00784 (0.0335)	0.0434 (0.0266)	0.0127 (0.0273)
Romney Meme	0.000811 (0.0306)	−0.0181 (0.0324)	−0.00712 (0.0256)	−0.0146 (0.0262)
Obama Tone—Angry/Threat	0.00724 (0.00945)	−0.00880 (0.0100)	−0.00724 (0.00789)	−0.00795 (0.00808)
Obama Tone—Happy/Reassuring	–0.00548 (0.00979)	0.0141 (0.0104)	−0.0129 (0.00820)	0.00229 (0.00840)
Romney Tone—Angry/Threat	0.0165* (0.00990)	−0.00157 (0.0105)	−0.0230*** (0.00826)	−0.00484 (0.00845)
Romney Tone—Happy/Reassuring	–0.00384 (0.00889)	0.00806 (0.00941)	−0.00106 (0.00741)	−0.00300 (0.00759)
Obama Facial—Angry/Threat	0.0136* (0.00739)	0.000946 (0.00782)	−0.00148 (0.00619)	0.0106* (0.00635)
Obama Affinity Gesture	−0.0163 (0.0106)	−0.00756 (0.0112)	−0.00613 (0.00883)	−0.00854 (0.00904)
Obama Defiance Gesture	0.000223 (0.00778)	−0.000842 (0.00824)	−0.00564 (0.00653)	−0.00378 (0.00669)
Romney Facial—Angry/Threat	−0.000301 (0.00780)	−0.00124 (0.00826)	−0.00235 (0.00656)	−0.00497 (0.00673)
Romney Affinity Gesture	−0.00718 (0.00874)	0.00531 (0.00925)	−0.00706 (0.00726)	0.00996 (0.00743)
Romney Defiance Gesture	−0.0129 (0.00995)	−0.0154 (0.0105)	0.00817 (0.00831)	0.00389 (0.00851)
Obama Blink Rate—Average	0.00736* (0.00413)	0.00344 (0.00437)	0.000449 (0.00342)	0.00348 (0.00350)
Romney Blink Rate—Average	0.00706 (0.00536)	0.00168 (0.00567)	0.00485 (0.00446)	0.00410 (0.00456)
Obama Blink Rate—*SD*	0.0208** (0.00838)	0.000116 (0.00887)	0.00194 (0.00705)	0.0112 (0.00722)
Romney Blink Rate—*SD*	−0.0128 (0.00795)	−0.0123 (0.00841)	−0.00554 (0.00666)	−0.00556 (0.00682)
Constant	−0.275*** (0.0510)	−0.200*** (0.0518)	−0.514*** (0.0321)	−0.557*** (0.0324)
Observations	180	180	180	180
R^2	0.171	0.069	0.169	0.128
Block F tests				
Verbal	2.48**	0.66	0.73	0.28
Tonal	0.97	1.02	2.73**	0.39
Visual	2.22**	0.71	0.47	1.08

Note. Standard errors in parentheses.
***$p < .01$. **$p < .05$. *$p < .1$.

finds that the candidates' rhetorical attacks and meme generation contributed to the models in 4 of 8 tests for the volume of second screen expression about the candidates and in 2 of 8 tests for the sentiment of this expression. With these verbal elements assessed, the tonal elements contributed incremental variance to the predictive models in 3 of 8 tests for volume (Hypothesis 1a) and in 1 of 8 tests for the sentiment (Hypothesis 2a). After accounting for these verbal and tonal elements, the visual elements contributed significant incremental variance to the predictive models in 3 of 8 tests for volume (Hypothesis 1b) and in 3 of 8 tests for sentiment (Hypothesis 2b). This pattern of findings suggests a meaningful and consistent influence of nonverbal communication elements on volume and sentiment. Taken together, the tonal factors contribute incremental variance in only 4 of 16 tests beyond verbal factors. In contrast, candidate expressions, gestures, and blink rate contributed incremental variance to the models in 6 of 16 tests even after accounting for verbal and tonal factors. Overall, this pattern lends support to Hypothesis 1, with tonal (Hypothesis 1a) and visual elements (Hypothesis 1b) jointly contributing to the predictive models. In contrast, results revealed more modest support for Hypothesis 2, with visual markers (Hypothesis 2b) performing better than the tonal variables (Hypothesis 2a) in terms of consistency of contribution to the predictive models.

Discussion

Consistent with theoretical expectations and previous experimental findings, the nonverbal behavior of candidates is consequential in driving social media responses, rivaling what candidates actually say during debates (Benoit, 2013; Cho et al., 2009; Zhu et al., 1994). These findings expand on previous work connecting debate performance to reactions on social media (Shah et al., 2015) by looking across two debates, one focused on domestic policy and the other on foreign policy, and integrating additional nonverbal markers, specifically, average blink rate and its standard deviation, into predictive models. Equally important, the analysis reported here employs generalized least squares (Prais–Winstein) regression, a time series modeling technique that addresses the significant autocorrelation in our data. Notably, these results yield substantially more conservative estimates than ordinary least squares regression.

Given this more conservative approach, it is notable that across both of these debate contexts, and two different dependent variables for two different candidates, the tonal and visual aspects of debate presentations explained variance in the volume and valence of second screen expression beyond the verbal aspects of debates. By linking the content of first and second screens during major political events and examining the immediate and lagged connection between candidates' behavior and users' social media expression, this work provides a novel method for analyzing real-time effects of broadcast media on social media. In doing so, it shares some similarity with prior debate analysis by Nagel et al. (2012) and Shah et al. (2015), while improving on the methodological approach employed in these earlier efforts to understand audience responses to debate verbal, tonal, and visual features.

And what does this effort to connect debate presentations and social expression reveal? Certainly, this work confirms that debates are moments of "national conversation," with peaks of expressive activity that dwarf all other campaign events in the 50 days before the election. But if public expression corresponds to nonverbal cues as much as verbal statements, we may have to rethink what constitutes a debate effect. This analysis suggests that the Twitter-using public, or at least that subset of users who choose to be active during presidential debates, responds as strongly to the visual elements of candidate behavior, including facial displays, expressive gestures, and their blink rates, as they do to verbal elements, particularly candidate memes or disputes. Future research should look beyond the volume and valence of online expression to examine other discursive outcomes.

Given the capacity of candidate nonverbal behavior to spark conservation, the quality of democratic discourse resulting from reactions to debate performances merits closer attention. When witnessed by audiences, do facial expressions, physical gestures, and autonomic responses (blink rates) trigger more attentive processing and encourage members of the viewing public to generate meaningful responses that productively contribute to a democratic dialogue? Or do these visual cues spur superficial utterances, focusing on the appearance or distress of the candidate? These remain open questions and warrant further attention from research following this line of research. Previous work has shown that candidates' tonal and visual behaviors provide relevant social information, with nonverbal communication treated as a more reliable predictor of leader traits than verbal utterances (see Bucy & Grabe 2008; Masters et al., 1986). Nonverbal cues, for instance, facilitate accurate inferences about candidates' competence and integrity (Olivola & Todorov, 2010).

Our analysis also reveals that while candidate attack strategies and memes explain some variance in the amount and favorability of expression, they were matched by the influence of nonverbal factors, especially candidate expressions, gestures, and blinking. It is notable that across these two debates, the most powerful verbal elements were memes—the quotable and malleable utterances that are easily repeated or referenced on a short messaging platform such as Twitter. Moreover, it appears that visual factors often absorb the effects attributed to tonal factors, suggesting a complex interplay between some of these interrelated nonverbal elements. Subsequent analyses should consider accumulated or conditional effects, such as how an angry or threatening tone works alongside corresponding facial expressions, or how defiance gestures are enhanced by a rapid blink rate or sudden variance in this response.

Future research should also differentiate among social media users to examine how subgroups respond to discrete moments during debates or other televised events. Classifying social media users based on their ideology or geography would allow their expressions to be differentiated and their responses to be modeled separately. By using profile information, geographic tags, and previous tweets to classify users, these relationships can be examined within subgroups rather than in the aggregate, revealing how different user segments respond to the verbal, tonal, visual, and autonomic aspects of candidates' debate performance. This would be an appropriate extension of the novel methods advanced in this article, connecting the power of television images to the responses of social media audiences via the second screen.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2013S1A3A2055285).

Notes

1. Eleven television channels broadcasted the debates live, as did YouTube.
2. We did not include the second debate in this analysis because it was conducted in a Town Hall format, in which citizens posed questions to the candidates and they moved freely around a stage. Given these differences from the more conventional staging of the first and third debates, we focused our analysis on these two instances.
3. This may explain in part why blink rates often decrease when an individual lies (Mann, Vrij, & Bull, 2002).
4. We also coded for *agreement* with a position just taken, *pleasantries* exchanged as a matter of routine between the candidates, and *policy statements* about what each candidate would do if elected (beyond responses and contrast statements). These were not as common as the four functions included in our models, however, and appeared too infrequently for analysis.
5. See https://dev.twitter.com/docs/api/streaming
6. For available platform objects, see https://dev.twitter.com/overview/api
7. See blog.twitter.com/2012/dispatch-from-the-denver-debate and blog.twitter.com/2012/the-final-2012-presidential-debate

References

Abramowitz, A. I. (1978). The impact of a presidential debate on voter rationality. *American Journal of Political Science*, *22*, 680-690.

Alexander, J. C. (2010). *The performance of politics: Obama's victory and the democratic struggle for power* [Reprint edition]. New York, NY: Oxford University Press.

Alexander, J. C. (2012, April 10). Obama's downcast eyes. *The Huffington Post*. Retrieved from http://www.huffingtonpost.com/jeffrey-c-alexander/obama-debate-performance_b_1938755.html

Barbato, G., De Padova, V., Paolillo, A. R., Arpaia, L., Russo, E., & Ficca, G. (2007). Increased spontaneous eye blink rate following prolonged wakefulness. *Physiology & Behavior*, *90*, 151-154.

Benoit, W. L. (2013). *Political election debates: Informing voters about policy and character*. Lanham, MD: Lexington Books.

Benoit, W. L., & Harthcock, A. (1999). Functions of the great debates: Acclaims, attacks, and defenses in the 1960 presidential debates. *Communication Monographs*, *66*, 341-357.

Benoit, W. L., McKinney, M. S., & Holbert, R. L. (2001). Beyond learning and persona: Extending the scope of presidential debate effects. *Communication Monographs*, *68*(3), 259-273.

Brefczynski-Lewis, J. A., Berrebi, M. E., McNeely, M. E., Prostko, A. L., & Puce, A. (2011). In the blink of an eye: Neural responses elicited to viewing the eye blinks of another individual. *Frontiers in Human Neuroscience*, *5*, 1-11.

Bucy, E. P., & Grabe, M. E. (2008). “Happy warriors” revisited: Hedonic and agonic display repertoires of presidential candidates on the evening news. *Politics and the Life Sciences*, *27*, 78-98.

Burgoon, J. K., & Hale, J. L. (1988). Nonverbal expectancy violations: Model elaboration and application to immediacy behaviors. *Communication Monographs*, *55*, 58-79.

Cappella, J. N., & Jamieson, K. H. (1997). *Spiral of cynicism: The press and the public good*. New York, NY: Oxford University Press.

Carr, D. (2012, 14 October). TV debates that sell more than just drama. *The New York Times*. Retrieved from http://www.nytimes.com/2012/10/15/business/media/televised-debates-that-sell-more-than-just-drama.html?_r=0

Cho, J., Shah, D. V., Nah, S., & Brossard, D. (2009). “Split screens” and “spin rooms”: Debate modality, post-debate coverage, and the new videomalaise. *Journal of Broadcasting & Electronic Media*, *53*, 242-261.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*, 409-412.

Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A., & Menczer, F. (2011). Political polarization on Twitter. In *Proceedings of the 5th International Conference on weblogs and social media* (pp. 89-96). Menlo Park, CA: AAAI Press.

Cunningham, D. W., & Wallraven, C. (2009). Dynamic information for the recognition of conversational expressions. *Journal of Vision*, *9*(13), 1-17.

Davis, S. (1999). Effects of audience reaction shots on attitudes towards controversial issues. *Journal of Broadcasting & Electronic Media*, *43*, 476-491.

Dawkins, R. (1976). *The selfish gene* (1st ed.). New York, NY: Oxford University Press.

Doughty, M. J. (2001). Consideration of three types of spontaneous eyeblink activity in normal humans: During reading and video display terminal use, in primary gaze, and while in conversation. *Optometry and Vision Science*, *78*, 712-725.

Druckman, J. N. (2003). The power of television images: The first Kennedy-Nixon debate revisited. *Journal of Politics*, *65*, 559-571.

Exline, R. V. (1985). Multichannel transmission of nonverbal behavior and the perception of powerful men: The presidential debates of 1976. In S. L. Ellyson & J. F. Dovidio (Eds.), *Power, dominance, and nonverbal behavior* (pp. 183-206). New York: Springer-Verlag.

Fogarty, C., & Stern, J. A. (1989). Eye movements and blinks: Their relationship to higher cognitive processes. *International Journal of Psychophysiology*, *8*, 35-42.

Fridkin, K. L., Kenney, P. J., Gershon, S. A., & Woodall, G. S. (2008). Spinning debates: The impact of the news media’s coverage of the final 2004 presidential debate. *International Journal of Press/Politics*, *13*, 29-51.

Grabe, M. E., & Bucy, E. P. (2009). *Image bite politics: News and the visual framing of elections*. New York, NY: Oxford University Press.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*, 701-721.

Hall, J. A. (1980). Voice tone and persuasion. *Journal of Personality and Social Psychology*, *38*, 924-934.

Harris, C. S., Thackray, R. I., & Shoenberger, R. W. (1966). Blink rate as a function of induced muscular tension and manifest anxiety. *Perceptual and Motor Skills*, *22*, 155-160.

Hellweg, S. A., & Phillips, S. L. (1981). A verbal and visual analysis of the 1980 Houston Republican presidential primary debate. *Southern Speech Communication Journal*, *47*, 23-38.

Jamieson, K. H. (1992). *Dirty politics: Deception, distraction, and democracy*. New York, NY: Oxford University Press.

Jamieson, K. H., & Adasiewicz, C. (2000). What can voters learn from election debates? In S. Coleman (Ed.), *Televised election debates: International perspectives* (pp. 25-42). New York, NY: St. Martin's Press.

Jamieson, K. H., & Birdsell, D. S. (1990). *Presidential debates: The challenge of creating an informed electorate* [Reprint edition]. New York, NY: Oxford University Press.

Karson, C. N., Berman, K. F., Donnelly, E. F., Mendelson, W. B., Kleinman, J. E., & Wyatt, R. J. (1981). Speaking, thinking, and blinking. *Psychiatry Research*, *5*, 243-246.

Kraus, S. (1996). Winners of the first 1960 televised presidential debate between Kennedy and Nixon. *Journal of Communication*, *46*(4), 78-96.

Laplante, D., & Ambady, N. (2003). On how things are said: Voice tone, voice intensity, verbal content, and perceptions of politeness. *Journal of Language and Social Psychology*, *22*, 434-441.

Mann, S., Vrij, A., & Bull, R. (2002). Suspects, lies, and videotape: An analysis of authentic high-stakes liars. *Law and Human Behavior*, *26*, 365-376.

Masters, R. D., Sullivan, D. G., Lanzetta, J. T., McHugo, G. J., & Englis, B. G. (1986). The facial displays of leaders: Toward an ethology of human politics. *Journal of Social and Biological Structures*, *9*, 319-343.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746-748.

Mutz, D. C. (2007). Effects of "in-your-face" television discourse on perceptions of a legitimate opposition. *American Political Science Review*, *101*, 621-635.

Mutz, D. C., & Reeves, B. (2005). The new videomalaise: Effects of televised incivility on political trust. *American Political Science Review*, *1*, 1-15.

Nagel, F., Maurer, M., & Reinemann, C. (2012). Is there a visual dominance in political communication? How verbal, visual, and vocal communication shape viewers' impressions of political candidates. *Journal of Communication*, *62*, 833-850.

Oh, J., Han, M., Peterson, B. S., & Jeong, J. (2012). Spontaneous eyeblinks are correlated with responses during the Stroop task. *PLoS One*, *7*, 1-11.

Olivola, C., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior*, *34*, 83-110.

Omori, Y., & Miyata, Y. (2001). Estimates of impressions based on frequency of blinking. *Social Behavior and Personality*, *29*, 159-67.

Patterson, M. L., Churchill, M. E., Burger, G. K., & Powell, J. L. (1992). Verbal and nonverbal modality effects on impressions of political candidates: Analysis from the 1984 presidential debates. *Communication Monographs*, *59*, 231-242.

Peters, J. W. (2012, 9 October). Networks like split-screens in debates, even if candidates don't. *The New York Times*. Retrieved from http://www.nytimes.com/2012/10/10/us/politics/networks-like-split-screens-in-debates-even-if-candidates-dont.html

Shah, D. V., Hanna, A., Bucy, E. P., Wells, C., & Quevedo, V. (2015). The power of television images in a social media age: Linking biobehavioral and computational approaches via the second screen. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 225-245.

Sharp, A. (2012a, 4 October). *Dispatch from the Denver debate* [Twitter blogs]. Retrieved from https://blog.twitter.com/2012/dispatch-from-the-denver-debate

Sharp, A. (2012b, 23 October). *The final 2012 presidential debate* [Twitter blogs]. Retrieved from https://blog.twitter.com/2012/the-final-2012-presidential-debate

Shultz, S., Klin, A., & Jones, W. (2011). Inhibition of eye blinking reveals subjective perceptions of stimulus salience. *Proceedings of the National Academy of Sciences*, *108*, 21270-21275.

Sigelman, L., & Sigelman, C. K. (1984). Judgments of the Carter-Reagan debate: The eyes of the beholders. *Public Opinion Quarterly*, *48*, 624-628.

Stewart, P. A., & Mosely, J. (2009). Politicians under the microscope: Eye blink rates during the first Bush-Kerry debate. *White House Studies*, *9*(4), 373-388.

Stroud, N. J., Stephens, M., & Pye, P. (2011). The influence of debate viewing context on political cynicism and strategic interpretations. *American Behavioral Scientist*, *55*, 270-283.

Sullivan, D. G., & Masters, R. D. (1988). Happy warriors: Leaders' facial displays, viewers' emotions, and political support. *American Journal of Political Science*, *32*, 345-368.

Sullivan, D. G., & Masters, R. D. (1993). Nonverbal behavior, emotions, and democratic leadership. In G. E. Marcus & R. L. Hanson (Eds.), *Reconsidering the democratic public* (pp. 307-332). University Park: Pennsylvania State University Press.

Tiemens, R. K., Hellweg, S. A., Kipper, P., & Phillips, S. L. (1985). An integrative verbal and visual analysis of the Carter-Reagan debate. *Communication Quarterly*, *33*, 34-42.

Tsekleves, E., Cruickshank, L., Hill, A., Kondo, K., & Whitham, R. (2007). Interacting with digital media at home via a second screen. In *Proceedings of the 9th IEEE International Symposium on multimedia workshops* (pp. 201-206). Washington, DC: IEEE Computer Society.

Tsfati, Y. (2003). Debating the debate: The impact of exposure to debate news coverage and its interaction with exposure to the actual debate. *International Journal of Press/Politics*, *8*, 70-86.

Vaccari, C., Chadwick, A., & O'Loughlin, B. (2015). Dual screening the political: Media events, social media, and citizen engagement. *Journal of Communication*, *65*, 1041-1061.

Wald, K. D., & Lupfer, M. B. (1978). The presidential debate as a civics lesson. *Public Opinion Quarterly*, *42*, 342-353.

Way, B. M., & Masters, R. D. (1996). Political attitudes: Interactions of cognition and affect. *Motivation and Emotion*, *20*, 205-236.

Wells, C., Van Thomme, J., Maurer, P., Hanna, A., Pevehouse, J., Shah, D. V., & Bucy, E. (2016). Coproduction or cooptation? Real-time spin and social media response during the 2012 French and US presidential debates. *French Politics*, *14*(2), 206-233.

Yang, Y., & Liu, X. (1999). A re-examination of text categorization methods. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on research and development in information retrieval* (pp. 42-49). New York, NY: ACM.

Zettl, H. (1990). *Sight, sound, motion: Applied media aesthetics*. Boston, MA: Wadsworth.

Zhu, J.-H., Milavsky, J. R., & Biswas, R. (1994). Do televised debates affect image perception more than issue knowledge? A study of the first 1992 presidential debate. *Human Communication Research*, *20*, 302-333.

Author Biographies

Dhavan V. Shah is the Louis A. & Mary E. Maier-Bascom Professor at the University of Wisconsin–Madison, where he is director of the Mass Communication Research Center. His work focuses on framing effects on social judgments, digital media influence on civic engagement, and the impact of health information and communication technologies.

Alex Hanna is an assistant professor in the Institute for Communication, Culture, Information and Technology at the University of Toronto. Her research focuses on social movements, media, and computational social science, with a particular interest in protest coverage and political participation.

Erik P. Bucy is the Marshall and Sharleen Formby Regents Professor of Strategic Communication in the College of Media and Communication at Texas Tech University. His research includes nonverbal aspects of political news and public opinion about the press. He is the outgoing editor of *Politics and the Life Sciences*, an interdisciplinary journal.

David S. Lassen is a doctoral candidate in the Department of Political Science at the University of Wisconsin–Madison. His research focuses on the role of humor in American politics, mass political participation, as well as media coverage of government officials and institutions.

Jack Van Thomme is a doctoral candidate in the Department of Political Science at the University of Wisconsin–Madison. His research focuses on rhetorical, political, and institutional change in European far right parties as it relates to European elections and political behavior.

Kristen Bialik is a master's student in the Department of Journalism and Mass Communication at the University of Wisconsin–Madison. Her research interests focus on the relationship between news framing of global issues, public opinion, and foreign policy.

JungHwan Yang is a doctoral candidate in the School of Journalism and Mass Communication at the University of Wisconsin–Madison. His research employs computational and quantitative approaches to explore partisans' political perceptions and behaviors in new media environments.

Jon C. W. Pevehouse is Vilas Distinguished Achievement Professor of Political Science at the University of Wisconsin–Madison. His research interests center on international relations, international political economy, American foreign policy, international organizations, and political methodology. He is a specialist in time series analysis in the social sciences.