CHAPTER 3

IMAGE BITE ANALYSIS OF PRESIDENTIAL DEBATES

Erik P. Bucy
Zijian Harrison Gong

Presidential debates provide an ideal setting in which to observe and document the nonverbal behavior of political candidates in a competitive context. Because debates are televised, they present viewers with an extended opportunity to evaluate candidates not only for their issue positions, which are widely reported and available from a multitude of sources, but also for their communication efficacy and nonverbal communication style, from which viewers may infer a variety of politically relevant traits and associated evaluations. The use of the continuous split-screen presentation format by C-SPAN and other cable and broadcast networks during the 2012 general election debates between Barack Obama and Mitt Romney (Peters, 2012) highlighted the role of nonverbal behavior to an extent never before seen, as both candidates were in view continuously throughout each 90-minute encounter whether speaking or not. Indeed, the considerable consternation expressed by Obama supporters over the president’s subpar performance in
the first debate stemmed largely from his nonverbal communication, which
was criticized for seeming at times evasive, unconcerned, and disengaged
(Nagourney, Parker, Rutenberg, & Zeleny, 2012).

Analyzing the nonverbal dimension of presidential debates requires a
markedly different approach to derive meaning than does rhetorical or tex-
tual analysis. First, there is no nonverbal “transcript” that can be easily ex-
amined for recurring themes, rhetorical tropes, oratorical flourishes, memes
(pithy statements), or other language-based devices or strategies. The analyst
must look to completely separate literatures—on emotion, expression, and
even biobehavior—for inspiration and theoretical grounding (for a review, see
Bucy, 2011). Fortunately, a set of validated nonverbal categories now exists for
the coding of political debates and other televised messages, including news,
advertising, press conferences, presidential addresses, and other communi-
cation genres (see Grabe & Bucy, 2009). Second, nonverbal analysis encom-
passes not just facial expressions and bodily gestures but also tonal elements,
eye gaze and “shiftiness,” blink rate, head orientation, and other behaviors.
Production features that affect candidate presentation, such as camera angles,
shot lengths, vector orientation, and lead room, can be carefully coded as well
(see Bucy & Ball, 2010; Tiemens, 1978).

Regardless of variable selection, as a form of content analysis, nonverbal
coding should proceed from a firm conceptual foundation rather than rep-
resent an atheoretical exercise in merely generating data. While there is no
need to reinvent categories for coding, as a plethora of useful variables already
exist for this purpose (see Grabe & Bucy, 2009, 2011), there is a definite need
for work to proceed with solid conceptual grounding.

Notwithstanding the range of possible variables that could be coded,
nonverbal analysis of political communication recognizes a distinction be-
tween sound bites, defined as video segments in which candidates are shown
speaking, and image bites, defined as video segments in which candidates are
shown but not heard (Bucy & Grabe, 2007). Nonverbal behavior occurs in
both, but in a presidential debate setting a sound bite denotes the utterances
of the candidate who is speaking while an image bite denotes the visual por-
trayal of the candidate who is listening, or shown in a reaction shot. Because
political discourse requires dedicated attention and a cognitive framework
or schema for efficient processing of political information (see Ferejohn &
Kuklinski, 1990), the effects of nonverbal displays are sometimes more pro-
nounced for image bites, or reaction shots, than are candidate statements.
As shown in the analysis that follows, evidence from the 2012 presidential debates corroborates the general finding that viewers respond more to the candidates’ nonverbal displays than to their verbal arguments. Given the popularity of the split-screen format among the television networks (Peters, 2012), it is more important than ever for candidates to be aware of the nonverbal signals they are conveying.

In the sections that follow, we present a framework for “image bite” analysis of the 2012 U.S. presidential debates, highlighting the role of nonverbal behavior in political communication. C-SPAN’s footage of the 2012 presidential debates, which used a split-screen format and persistent camera shot in the first and third debates, is used to contrast the performances of Barack Obama and Mitt Romney. We review the major categories of nonverbal display behavior used in image bites analysis then present three applications of the approach to illustrate the utility of the coding scheme. First, the results of a detailed content analysis of the candidates’ nonverbal communication, including facial expressions, evocative gestures, and voice tone, are presented. In particular, particular emphasis is placed on the communicative style of President Obama, whose performance in the first debate represented a low point in the campaign and gave Mitt Romney momentum in the polls. Next, we show how memorable moments during debates can be utilized as conversation starters in focus groups to address broader questions about the appropriateness of candidate nonverbal behavior and how communication style affects viewer interpretations of leadership ability. Finally, we discuss how our biobehavioral coding of debates can be used to predict the valence and volume of candidate mentions on the social media platform Twitter. Issues in coding, including coder training, techniques to ensure accuracy, and achieving intercoder reliability, as well as future directions of image bites research, are also considered.

THE IMAGE BITES APPROACH

Though mentioned by Barnhurst and Steele (1997) in the 1990s, the image bite concept was not elaborated and systematically measured until Bucy and Grabe (2007) documented the ratio of image to sound bites in a longitudinal study of major network news coverage of presidential elections from 1992 to 2004. They found, perhaps not surprisingly, that candidates were being shown more than they were being heard and that as average sound bite time
was shrinking, the cumulative time that the broadcast networks (ABC, CBS, and NBC) were dedicating to visual coverage of the candidates per campaign story was increasing. Building a conceptual case for studying the content of image bites begins with an appreciation for the social information that nonverbal displays express. Facial expressions in particular are reliable indicators of a communicator’s emotional state and simultaneously serve as the basis of myriad judgments about politically relevant traits, such as competence, integrity, dominance, and appropriateness (Bucy, 2011). Indeed, inferences of competence from photographs of candidates’ faces alone can predict electoral outcomes at a rate significantly better than chance (Todorov, Mandisodza, Goren, & Hall, 2005).

Unlike verbal pronouncements and written communication, facial displays and other nonverbal modes of communication require no textual literacy to process and are not dependent on linguistic and syntactical modes of sense making (see Messaris & Abraham, 2001). They are readily recognized and, for the most part, understood by viewers regardless of educational level, socioeconomic status, or political sophistication (see Grabe & Bucy, 2009; Prior, 2014). This endows nonverbal communication with a universal quality that spoken words lack. Because of this accessibility, political nonverbals are arguably more important to study than rhetorical strategies because they are comprehended by a much broader segment of the electorate—the mass audience, basically, as opposed to the politically motivated and educated audience. Recent findings by Prior (2014) reinforce the importance of political visuals to the average citizen. When traditional political knowledge questions are accompanied by a related visual image for illustration, the sizeable differences normally observed for gender, age, and education dissipate dramatically in national samples. Prior concludes that visual knowledge is as indicative of civic competence as is verbal knowledge.

Beyond its equalizing potential, visual processing is also much more efficient than verbal processing. Facial expressions are accurately recognized within milliseconds of exposure, even sometimes outside of conscious awareness. During speeches and other televised events, viewers are even influenced by microexpressions—rapid and subtle displays of facial emotion lasting less than a second. In a study of President George H. W. Bush’s speech following Iraq’s invasion of Kuwait, participants exposed to Bush’s microexpressions of happiness/reassurance reported significantly less anger and threat compared
to an experimental group shown the same speech with the microexpressions removed (Stewart, Waller, & Schubert, 2009). Thus, nonverbal variations of even very short duration can affect emotional responding to a political speech. Consistent with the evolutionary preference for visual processing, political nonverbals are also better remembered than verbal information (Grabe & Bucy, 2009; Prior, 2014), and they are more reliable indicators of information acquisition during televised events than measures of verbal recall.

**Coding Nonverbal Behavior**

When coding nonverbal behavior, it is important to prioritize those gestures and displays that are politically consequential. For this, one must consult the research literature in human ethology, evolutionary psychology, and behavioral biology—branches of the life sciences interested in the interaction between social behavior and biology—to identify the nonverbal display repertoires that play the largest role in social organization, including politics. Fortunately, the first generation of empirically oriented scholars in politics and the life sciences, notably the political scientists Roger Masters and Dennis Sullivan, along with psychologist John Lanzetta (all colleagues at Dartmouth College in the 1970s and 1980s), synthesized the extant literature and carefully derived a set of enduring facial display categories with exacting, behaviorally derived definitions that lend themselves to continued use and application across both observed and televised settings (Lanzetta, Sullivan, Masters, & McHugo, 1985; Masters, Sullivan, Lanzetta, McHugo, & Englis, 1986). These categories have been used since their introduction in the mid-1980s in experimental and content analytic work involving television news, debates, and other televised political appearances (see Bucy & Grabe, 2007, 2008; Grabe & Bucy, 2009; Masters, 2001; Masters & Sullivan, 1993).

For our image bites research the 2012 presidential debates were recorded from C-SPAN’s telecast and the digital files were used for content analysis, focus group discussion, an eye-tracking experiment, and a public opinion study of social media expression during the first debate. To track the timing of candidate comments, we relied on the C-SPAN Archives’ online Video Library of the debates. C-SPAN’s consistent use of a two-shot split-screen technique and maintenance of a medium shot length throughout the first and third debates was advantageous for purposes of analysis. (The second debate
followed a town hall format and featured a moving camera perspective that was much less uniform.) In split-screen presentations the screen is divided into two equal-sized boxes. Each candidate occupies half of the screen, and two different cameras are used to feature each candidate continuously. This dual visual presentation allows viewers (and researchers) to assess the performance of one candidate without overlooking the reactions of the other.

Consistent with earlier visual analysis of political debates (e.g., Tiemens, 1978), an individual shot, defined as an uninterrupted piece of video between the beginning and ending of a camera change, was used as the unit of analysis. In our coding, shots less than 30 seconds in duration were regarded as single segments, while shots longer than 30 seconds were divided into 30-second increments for analysis. Once candidate introductions and opening statements commenced, shot changes were almost nonexistent. Thus, we identified 177 codable segments for the first debate and 180 codable segments for the third debate. Since both of these debates were broadcast uninterrupted, with a split screen featured throughout, we were able to capture the nonverbal behavior of both candidates for virtually the entire 90 minutes in each case.

**CODING SCHEME AND KEY VARIABLES**

To provide a sense of the more consequential nonverbal behaviors on display during presidential debates, we focus on categories for analysis that drive viewer response in predictive modeling (see Shah, Hanna, Bucy, Wells, & Quevedo, 2015). In particular, we code the candidates’ facial expressions, evocative gestures, and tone of voice. The following sections describe how these behaviors were operationalized and coded in the 2012 debates.

**Facial Displays**

In the biopolitics literature, at least four different categories of facial displays are associated with social dominance and subordination: happiness/reassurance, anger/threat, fear/evasion, and sadness/appeasement (Masters et al., 1986; Stewart, Salter, & Mehu, 2009). Of these, the two most common and, arguably, most consequential in political communication are happiness/reassurance and anger/threat. Because there was considerable concern among
partisans and the press about President Obama’s lack of engagement in the first debate (Alexander, 2012; Nagourney et al., 2012), we also consider fear/evasion. Each expressive category draws on research from primate and human ethology, which has found that different patterns of display behavior are associated with distinct roles in rivalry for dominance (see Bucy & Grabe, 2008). As composite terms, these emotion/display pairs reflect the duality of the emotion being expressed (e.g., happiness or anger) and the social signals communicated (e.g., reassurance or threat). On an interpretive note, it is worth appreciating that content analysis of nonverbal display behavior can only measure the visible display observed and cannot assume that the associated emotion was genuinely felt while being expressed by the communicator.

**Happiness/reassurance** displays, characterized by a smile or relaxed mouth position, are relatively fluid, smooth, and flexible. In these expressions, the eyes may be wide open or just slightly closed. Also evident are raised eyebrows and visible upper, or both upper and lower, teeth. Eye contact may be brief, followed by a cutoff or change of gaze to avoid staring. In addition, “crow’s feet” wrinkles may appear around the eyes, and the communicator’s head might be tilted to the side, back, or in a nodding position. Functionally, happiness/reassurance displays facilitate a hedonic or friendly mode of social interaction and in most situations lower the probability of an aggressive or competitive encounter. The exception is counterempathy, in which case a smile or other hedonic signal conveyed by a disliked other (e.g., a reviled politician or tormenting superior) may evoke a negative response in the observer (see Bucy & Bradley, 2004). As a sign of subordination, fearful smiles should be coded as an instance of fear/evasion.

**Anger/threat** displays evidence a more rigid pattern of facial expression, including a fixed stare, vertical head orientation, raised upper and tightened lower eyelids, brows that are pulled down and drawn together, lower or no teeth showing, and lowered mouth corners. When expressing anger/threat, the lips may be pressed firmly together or squared and tightened. The expression overall has a negative or tense quality and is coupled with a hostile communicative intent. Functionally, anger/threat displays are associated with competitive or hostile (agonic) encounters, aggressive behavior, and challenges to dominance hierarchies. Whereas political challengers and rivals are frequently aggressive, “the leader is usually the focus of attention, often engaging in hedonic or reassuring behavior” (Masters, 1981, p. 64). As Howard Dean
learned after his infamous “scream” during the 2004 primaries, excessive exhibitions of anger/threat on the campaign trail may attract intense media attention and quickly become characterized as nonpresidential.

_Fear/evasion_ displays feature furrowed brows and gaze aversion, a lowered head position, abrupt movement, and, at times, side-to-side head turning. In some cases, the emitter’s eyelids will be raised: the “deer caught in the headlights” look. Other times, the brows might be slightly furrowed and wrinkles may form in the middle of the forehead, suggesting worry; the lips may also stretch horizontally and the chin may be lowered. An evasive expression communicates an intention to avoid confrontation. Functionally, fear/evasion displays are also associated with agonic encounters, but instead of indicating aggression they signal subordination, avoidance, and inferior status—the social outcomes of effective aggression. Candidates who are forced to respond to allegations or difficult questions, who are caught in a contradiction or misrepresentation, or who are asked to justify contradictory statements might exhibit fear/evasion.¹

Figures 3.1 and 3.2 present screen captures of the candidates’ expressive variability in debate 3 across the prototypical display categories discussed above. Also included are images of both candidates’ neutral expressions, which characterize their reaction shots while not speaking and serve as reference points for determining when a prototypical display of happiness/reassurance, anger/threat, or fear/evasion is occurring.

**Evocative Gestures**

Evocative gestures were coded as body language that signals 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, moderator, or opponent. Examples included waving or giving a thumbs-up; winking or nodding knowingly to the camera, moderator, or other candidate; or using an open palm when referencing a policy, making a point, or appealing to the audience or opponent on a topic of mutual agreement (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 his or her opponent. Examples include finger pointing, wagging, or shaking; fist
raising; head shaking in disagreement; negative expressions accompanied by prolonged stares; and other behaviors signaling aggression.

From analysis of network news general election coverage we know that trailing candidates and debate losers are shown more often exhibiting anger/threat and making defiance gestures in news segments than are frontrunners and debate winners (Grabe & Bucy, 2009). Debate winners, by contrast, are more likely to be shown engaging in affinity behaviors that imply bonding, compassion, or friendship. While less nuanced than facial expressions, gestures typically work in unison with expressions and can amplify their effect.
Voice tone is a paralinguistic cue present in all spoken communication that imparts the emotion of the speaker while modulating the meaning of what’s being said. Voice tone also signals social intent, whether to communicate reassurance, as in the case of a friendly tone, or disapproval or even hostility, as in the case of an angry or threatening tone. In any competitive political encounter, a large part of nonverbal influence stems not just from semantic
content but also from voice tone and variability (Anderson & Klofstad, 2012; Klofstad, Anderson, & Peters, 2012).

In our analysis of the 2012 debates, we operationalized a voice tone evincing anger/threat as statements in which the candidate's vocal quality 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, or aggressive.

A voice tone evincing happiness/reassurance was operationalized as statements in which the candidate's vocal quality had an optimistic or cheerful feeling; where the candidate's voice was upbeat and positive and conveyed an affiliative or conciliatory 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.

Similar to the findings for facial displays, longitudinal analysis of presidential election coverage has found that challengers and debate losers tend to be more aggressive in tone than are incumbents and frontrunners (Bucy & Grabe, 2008; Grabe & Bucy, 2009), consistent with their secondary status. Candidates who are behind in the polls or recognized as having lost a debate are more often shown in news reports as angry or delivering statements that are negative in tone than frontrunners and recognized debate winners. This behavioral pattern is consistent with ethological observations that have documented aggression in second-ranking individuals or challengers to power (Masters, 1989).

Having described our main coding categories, we next turn to a discussion of issues that arise during coding, including the need for careful training, techniques to ensure accuracy, and achieving intercoder reliability.

**ISSUES IN CODING**

In visual content analysis, as with text-based analysis, coding consistency is a vital part of the research process (see Benoit, 2011). Even with precise variable definitions, as presented in the previous section, it takes more than simply
generating a list of definitions and having a group of coders accurately document the nonverbal content of a debate. Different viewers may watch the same debate, but without adequate training they may not “see” or recognize the same nonverbal behavior, particularly if two candidates are presented simultaneously. And even small discrepancies in interpretation can lead to sizeable differences in tabulated results, an unacceptable outcome. As is well known in content analysis, achieving a high degree of intercoder reliability is necessary before results can be reported—and even before intensive coding of content should proceed (see Krippendorff, 2012; Neuendorf, 2002).

Thus, training coders to recognize nonverbal behavior during political debates is an important first step in the process. Skilled coding of nonverbal behavior should begin with a systematic understanding of expressive displays from the biobehavioral literature to facilitate careful analysis of cues that are relevant to leadership. Appreciating the social and competitive significance of nonverbal behavior should produce more attentive coding, and with that more accurate data. Starting from the terra firma of the research literature also avoids unnecessary (and sometimes haphazard) attempts at inventing new categories of visual variables that have no real conceptual grounding or social meaning, or are so broad as to lack predictive validity. Additionally, some understanding of television production is beneficial to appreciate the visual environment of televised debates to know when production features matter.

Interestingly, our initial coding of anger/threat displays did not produce acceptable intercoder reliability figures, potentially because the candidates’ head and body movements were restrained and rarely changed direction during both sound bites (candidate statements) and image bites (reaction shots). Thus, a clear differentiation between a more neutral reaction shot looking at the opponent and a staring gaze indicative of anger/threat could not be made with situational cues from the communication setting alone. To differentiate neutral looking from more menacing staring, we expanded the definition of anger/threat to include widened eyes and a reduced blinking rate (see Garland-Thomson, 2009). We also defined the generally rigid or tense facial expression indicative of anger/threat in more detail, adding dilated nostrils and vertical lines between the eyebrows (see Knapp, Hall & Hogan, 2013).

To ensure that each candidate’s full range of nonverbal behavior is recognized, including small gestures and momentary (or micro) expressions, it
is important to code each candidate individually, either from the start of the debate to the finish or for each 30-second segment, before moving onto the opponent on the other side of the split screen. This sequential approach allows coders to understand an individual candidate's unique expressive range and tendencies, remain focused on the candidate in question, and not be distracted by the verbal give-and-take of political debate. Focusing on a single candidate at a time also allows coders to become familiar with a candidate's unique nonverbal repertoire (Bucy & Grabe, 2008) or expressive range, which may have a small but important influence on coding decisions where some coder judgment is required. For instance, during the third presidential debate, Mitt Romney, when not speaking, had the tendency to look at President Obama with a slight grimace. Although Romney's rigidity and prolonged gaze might have indicated anger/threat, for Romney it was really a default expression and was more accurately coded as a neutral display.

Another technique for ensuring accurate nonverbal coding is to turn the sound off. With the exception of voice tone, which of course must be coded with the sound on, coding with the sound off ensures that coders are responding to the visual channel of televised expressive behavior and not what the candidates are saying. And, indeed, the most accurate coding of visual variables occurs without the distraction of listening to the candidates’ verbal utterances, when the audio track is muted. Ekman and Friesen (1975) recognized the utility of this approach early in the development of systematic nonverbal coding techniques.

Finally, to ensure that individual shots or segments are coded accurately within the designated unit of analysis (e.g., the 30-second segment), it is advantageous to use a digital timer or auto clicker to precisely start and stop the recording at desired time points and to pause the recording to type numeric codes into a spreadsheet (or handwrite marks on a coding sheet). Auto clickers take the guesswork out of when to code for content and can be used in conjunction with most media players, such as Media Player for Windows or QuickTime for Mac. Analysis of audiovisual content should be approached as a precision endeavor to produce reliable data suitable for use in both descriptive comparisons and predictive modeling.

Because it requires precision, quality coding of nonverbal behavior is necessarily time consuming. To ensure quality and precision when coding a
debate, for example, we recommend making *multiple passes* of the same candidate during the same debate for different categories of variables. The first pass of coding may focus on voice tone and inappropriate displays, the second pass on facial displays and evocative gestures, and the third pass on eye blinks. Coding should begin after the moderator completes candidate introductions and turns the floor over to the first candidate for his or her opening statement and continue until the end of the candidates’ closing statements at the close of the debate.

Once coding is completed, a sample of the debate should be double-coded by a second trained coder to enable intercoder reliability analysis. Most content analysis textbooks recommend that at least 10 percent of the sample be double-coded for this purpose (Krippendorff, 2012; Neuendorf, 2002). In our coding of the debates, two coders were assigned to perform the coding for all three debates, rotating candidates between debates. To perform reliability analysis, each coder also analyzed 10 percent (roughly eighteen 30-second segments) of the other candidate’s nonverbal behavior for each debate. The double-coded segments were then subjected to reliability analysis. In cases in which an alpha coefficient of less than .80 was produced (using Krippendorff’s reliability program), the variables in question were recoded by the principal investigator for the entire debate until an acceptable level of agreement was attained.

**RESEARCH APPLICATIONS**

With reliable data in hand, coding of nonverbal behavior can be used in research for both descriptive comparisons and predictive modeling. Descriptively, nonverbal repertoires or frequency of different display types can be compared between candidates, between different stages of a debate (opening 30 minutes, middle 30, final 30), or across different debates.² In general, challengers or rivals to power are expected to act more aggressively than incumbents and exhibit more anger/threat in facial expressions and vocal tone while signaling defiance and rejection of the incumbent’s policies, legislative record, and vision for the country. This is particularly true during a president’s reelection bid, as with Bob Dole’s attacks on Bill Clinton’s character during the
1996 presidential debates or John Kerry’s aggressive attempt to unseat George W. Bush in 2004 (see Clines, 1996; Nagourney, 2004).

By contrast, in the “happy warriors” tradition of reassuring leadership (Sullivan & Masters, 1988), incumbents should embrace a more confident, empathetic style of communication typical of power holders, evidenced by happiness/reassurance displays, a reassuring tone of voice, and the use of affinity gestures. If sitting presidents are indeed more confident and self-assured than challengers, we would also expect their blinking rate to be lower overall, although blinking may fluctuate considerably across different topics, rhetorical exchanges, and even debate stages. Among recent presidents, perhaps Bill Clinton was the most empathetic of the era, performing feats of projected compassion and understanding that reached viewers on an emotional level through television in a way that few political communicators can (see Bucy & Newhagen, 1999).

Display Repertoires of Obama and Romney, 2012

Using these observations as a point of reference, we next analyze the display repertoires of Obama and Romney on a set of key variables during the first and third presidential debates of 2012. The second debate is not examined because of its unique town hall format, which featured candidates who roved from the podium to engage with the studio audience and a moving camera perspective. The coding issues presented by a moving camera are not insurmountable, but the different format makes direct comparisons more difficult. The first and third debates are highly comparable because they both featured a fixed camera perspective and split-screen presentation format throughout. In the first debate the candidates each stood behind a podium, while in the third debate they sat across from each other at a table.

Nonverbal presentation style has been recognized as an important aspect of political competition since the first televised presidential debates between John F. Kennedy and Richard Nixon in 1960 (see Kraus, 2001). The 2012 debates underscored the importance of effective nonverbal communication in politics, as observers widely criticized President Obama for a lackluster and disengaged nonverbal communication style in the first debate, which the more assertive and assured Mitt Romney was credited with winning. Our
data confirm the accuracy of these impressions and show that Obama had changed his approach considerably by the time of the third debate.

As shown in Table 3.1, Obama exhibited some degree of evasion (namely, by looking down or away from Romney while shown in a reaction shot, or image bite) in almost 40 percent of all segments in the first debate. By the third debate he had eliminated this avoidance behavior from his nonverbal repertoire almost entirely, and signs of evasion appeared in just 4.3 percent of segments. Obama’s more assertive approach in debate 3 was also evident in a greater percentage of anger/threat displays (in 35.9 percent of segments, compared to 23.8 percent in debate 1) and a more aggressive tone of voice. At the same time, he was more defiant in his use of gestures, using more rigid and emphatic hand movements while speaking. A different picture emerges for Romney, who projected a more upbeat tone and used more affinity gestures in debate 3.

### Table 3.1 Nonverbal Display Frequencies, Debates 1 and 3

<table>
<thead>
<tr>
<th></th>
<th>Debate 1</th>
<th>Debate 2</th>
<th>Debate 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obama</td>
<td>Romney</td>
<td>Obama</td>
</tr>
<tr>
<td><strong>Facial displays</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anger/threat</td>
<td>23.8% (44)</td>
<td>35.7% (66)</td>
<td>35.9% (65)</td>
</tr>
<tr>
<td>Happiness/reassurance</td>
<td>43.2% (80)</td>
<td>40.0% (74)</td>
<td>16% (29)</td>
</tr>
<tr>
<td>Fear/evasion</td>
<td>39.5% (73)</td>
<td>17.8% (33)</td>
<td>4.3% (8)</td>
</tr>
<tr>
<td><strong>Verbal tone</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anger/threat</td>
<td>32.4% (60)</td>
<td>43.8% (80)</td>
<td>44.2% (80)</td>
</tr>
<tr>
<td>Happiness/reassurance</td>
<td>24.9% (46)</td>
<td>21.6% (40)</td>
<td>22.1% (40)</td>
</tr>
<tr>
<td>Fear/evasion</td>
<td>4.3% (8)</td>
<td>2.7% (5)</td>
<td>1.1% (2)</td>
</tr>
<tr>
<td><strong>Gestures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affinity</td>
<td>6.5% (12)</td>
<td>16.2% (30)</td>
<td>10.5% (19)</td>
</tr>
<tr>
<td>Defiance</td>
<td>14.6% (27)</td>
<td>36.8% (68)</td>
<td>37.6% (68)</td>
</tr>
</tbody>
</table>

**Note:** Percentages may not add up to 100 because more than one display type may occur in any 30-second segment, or not at all. Frequency counts are shown in parentheses.
Figure 3.3 shows the dramatic difference in Obama’s fear/evasion displays compared to Romney’s in the first debate, and Romney’s more aggressive posture as evidenced through anger/threat displays, a critical voice tone, and increased use of defiance gestures. Figure 3.4 illustrates the candidates’ nonverbal behaviors in the third debate, showing Obama’s increased use of defiance gestures and greatly reduced fear/evasion displays. Figures 3.5 and 3.6 show
within-candidate comparisons across the two debates. Here, the differences 
between each candidate’s communication style between debate 1 and debate 3 
become quite obvious. By the third debate, Obama is noticeably more aggres-
sive and defiant and much less concerned about imparting a reassuring, con-
ciliatory tone than he was in the first debate, while Romney remains aggressive 
but also becomes more reassuring in tone while using more affinity gestures.
Display Appropriateness in the 2012 Debates

In the next phase of research, we convened a series of focus groups and asked participants to assess the appropriateness of President Obama's display behavior during the first and third debates. For this analysis, responses to the two best performing (i.e., most “appropriate” and most “inappropriate”) clips from an experimental study (see Gong & Bucy, 2014) were analyzed qualitatively. The two clips, again from C-SPAN's telecast of the debates (archived in the Video Library), both featured Obama and Romney in a split-screen presentation. A 53-second excerpt from debate 1 starting at 33:44 (see C-SPAN, 2012a) represented the inappropriate condition, and a 70-second excerpt from debate 3 starting at 80:16 (see C-SPAN, 2012b) represented the appropriate condition.

Both clips were shown in the context of several other memorable moments from televised politics, so that each group considered a range of different political encounters and was not focused solely on Obama and Romney. In the inappropriate clip from the first debate, Obama was shown glancing downward with a slight smirk while being verbally attacked by Romney in the adjacent split-screen window. In the appropriate clip, from the third debate, Obama is shown much more engaged and visually focused on Romney, interjecting out of turn with short objections, and denying several of Romney's claims. Rather than avoiding or enduring Romney's verbal barrages, Obama in this exchange makes it difficult for Romney to complete his point.

Themes Emerging from the Discourse

Overall, focus group participants judged the clip from the third debate, showing an assertive and combative Obama (see C-SPAN, 2012b), to be more presidential and confidence inspiring than the excerpt from the first debate. The clip from the first debate depicted a demur Obama who shied from eye contact and appeared to smirk, while looking downward, in response to some of Romney's comments. Interestingly, some initial interpretations depended on clip order: When Obama's inappropriate clip was shown after an evasive and stone-faced Michael Dukakis from the 1988 presidential debates, for example, Obama was at first rated positively—in comparison to Dukakis. However, in comparison to his own performance in the third debate, Obama's performance
in the first debate was judged more critically by focus group participants, and not just for ceding the floor to his opponent by refusing to visually engage Romney. As well, Obama was judged harshly for the way he was perceived as treating Romney. As Douglas, a focus group participant, put it, Obama deserved to experience a dip in the polls following his first debate performance, not because he appeared weak and ineffectual but because of his body language and his attitude—it was like, “you don’t know what you’re talking about and I don’t care what you have to say.” For me, he did not take what Romney was saying seriously. (Douglas)

Obama was unprepared. He came across like...he could care less about what Romney had to say. (Leah)

Focus group participants articulated a two-sided view of Obama’s inappropriate style from the first debate. First, they acknowledged that he seemed passive and disengaged, as was widely reported in news coverage, a posture that clearly violated nonverbal expectations. But also, and perhaps more interestingly, they criticized his nonverbal demeanor for seeming disinterested, dismissive, and disrespectful toward Romney—to the point of perceiving an element of sarcasm in Obama’s expression. The phrases unprofessional and unpresidential were used in describing Obama’s communication. “From a visual standpoint, I … noticed that Obama was kind of smirking,” a female participant observed. “That bothered me,” added a second female viewer. “He needed an attitude adjustment.” The perception of sarcasm in a partial, downcast smile illustrates the sensitivity that viewers have to even small, incomplete expressions.

By contrast, focus group discussion of Obama’s clip from the third debate reflected positive assessments of an engaged and assertive leadership style—one that conformed to expectations even while violating norms of polite discussion through interjection and speaking out of turn. Through such positive nonverbal expectancy violations, Obama’s communication was uniformly viewed as appropriate and fitting to the rhetorical situation. Even as Romney held the floor and attempted to score points for his economic record and policies toward the auto industry, Obama acted as a nonverbal auditor of Romney’s comments and was quick to correct his opponent upon hearing a questionable claim or statement inconsistent with the historical record. Such
active auditing and visual accounting of Romney by Obama was described by focus group participants as signaling dominance. “You’ve got one guy who is on the defense here [Romney] saying stuff. And the other guy who is saying, ‘No that’s not what your words or the print says.’ It’s kind of obvious and easy that Obama is winning this,” noted Kliff. “I think Obama derailed Romney quite a bit and kind of negated what he said,” added Bob.

Discussion about Romney’s performance in the clip from the third debate centered on the hurried pace and rehearsed quality of his statements. In Romney’s rush to recite facts, viewers discerned an ill-at-ease attempt to command the stage by rotely filling the air with an overstuffed recitation of rehearsed lines. “I think he could have cut out probably 50 percent of what he said and slowed down,” observed Bob. Participants characterized Romney’s rapid recitation of facts as having a defensive and inauthentic quality. “Romney came across as being very phony and he clearly was just following a script and going through the motions,” Donavin commented. “Obama, even with his few words, was able to keep [Romney] in the lane of being a governor versus being a chief commander of a nation,” added Sarah. Viewers also noticed Romney’s rapid blink rate, a potential sign of stress (or intensive mental activity).

From this we conclude that, when everyday citizens are given the opportunity to reflect on and share impressions of televised political encounters, small visual cues begin to speak loudly.

**Linking Biobehavioral and Social Media Approaches to Debate Effects**

Thus far we have considered nonverbal communication as a dependent variable or outcome measure to compare the candidates’ nonverbal communication styles between debates, and then we used key encounters between Barack Obama and Mitt Romney during the first and third presidential debates of 2012 as stimulus clips to generate focus group discussion about the appropriateness of the candidates’ display behavior. These clips were also used experimentally in eye-tracking research to assess the extent to which viewers fixate on inappropriate or unexpected candidate behavior (see Gong & Bucy, 2014).

Another way in which nonverbal coding can be applied in political communication research is predictively, as independent variables in a causal model showing debate effects. In an innovative project linking our biobehavioral coding of the first presidential debate to Twitter responses by viewers—perhaps the first study to formally link the content of first and second screens during
a political event—Shah and colleagues (2015) compared the ability of nonverbal communication and rhetorical strategies to predict the valence and volume of candidate mentions on the social media platform.

Considerable controversy surrounds the study of presidential debates, particularly efforts to connect their content and impact. Researchers have long debated whether the citizenry reacts to what candidates say, how they say it, or simply how they appear (see Druckman, 2003; Kraus, 1996). Using our detailed coding of the first debate as independent variables in a series of regression models, we examined the relative influence of the candidates' verbal persuasiveness and nonverbal style on viewers' “second-screen” behavior—their use of computers, tablets, and mobile phones to enhance or extend the televised viewing experience. To test these relationships, we merged our coding of the candidates' nonverbal communication, their rhetorical strategies during the debate (coded separately), and corresponding real-time measures, synched and lagged, of the volume and sentiment of Twitter expression about Obama and Romney.

Performing such an analysis requires a “Big Data” approach to parsing the publicly available Twitter postings on the evening of the debate, which numbered in the millions, and ultimately involved a machine learning method for determining the emotional valence, or sentiment, of the Twitter postings. Thus, advanced programming skills are requisite for this kind of large-scale analysis. Once the data are in hand, interesting questions can be asked. In close elections where much is at stake, leader displays and other nonverbal behaviors are likely to take on added significance. And in cases where the incumbent performs particularly poorly or commits a violation of nonverbal expectations (see Burgoon & Hale, 1988), or where the challenger surprises by exceeding expectations, we would expect a higher volume of audience attention and second-screen communication in response to these developments, as well as an outpouring of valenced reactions.

Consistent with evolutionary arguments and the image bites outlook on political communication, our analysis found that the candidates’ facial expressions and physical gestures (namely, displays of either affinity or defiance) were more consistent and robust predictors of the volume and valence of Twitter expression than the candidates’ persuasive strategies, pithy statements or memes, and voice tone during the debate. Thus, we were able to confirm on a large scale the consequentiality of nonverbal political behavior in driving social media responses during a debate.
Despite their inherent messiness, social media provide political communication researchers with an index of audience behavior that offers unprecedented precision and scope. Like continuous response measures that are typically gathered in small group settings, tweets and other social media posts provide moment-to-moment audience feedback that can be traced down to the second. But unlike continuous response measures, social media responses can have local, national, or international reach depending on how the search parameters are set. Auspiciously, the data they produce (at least publicly available Twitter posts) lend themselves to effects modeling as a promising new form of public reaction to televised political performances. Although Twitter users are not necessarily representative of the population, Shah and colleagues (2015) note that, “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” (pp. 229–230).

In interpreting the significance of our findings, we posit that second-screen responses to the candidates’ nonverbal behaviors indicate greater reliance on social rather than factual information or rhetorical efforts (see Grabe & Bucy, 2009, pp. 274–276). This interpretation is consistent with evolutionary analyses of political behavior, in which nonverbal communication is regarded as a more reliable predictor of leader traits than are verbal utterances (see Bucy & Grabe, 2008; Masters et al., 1986). If nonverbal cues facilitate inferences about such politically relevant traits as competence and integrity (see Olivola & Todorov, 2010; Rahn, Aldrich, Borgida, & Sullivan, 1990), we may need to rethink our assumptions about the informational cues that voters actually use, as opposed to the bases of information that normative theorists would prefer the public to rely on.

**FUTURE DIRECTIONS**

Though new to debate analysis, the image bites approach to political communication has already been validated as a robust and adaptable methodology for assessing the nonverbal dimension of candidate communication (Bucy & Grabe, 2007, 2008; Grabe & Bucy, 2009, 2011), easily extending to international political contexts (see Esser, 2008). The application of biobehavioral coding to televised presidential debates opens new vistas for investigating
Chapter 3

this time-honored campaign tradition because it introduces a systematic and easily replicated framework for documenting the nonverbal communication elements that are a continuous feature of competitive candidate encounters. Previous research on the visual dimensions of political debates tended to highlight the role of production techniques, camera angles, and other perceived presentational biases (e.g., Morello, 1988; Tiemens, 1978) rather than the social significance of political display behavior. Moreover, candidate expressions were documented primarily to assess the extent to which they reinforced the verbal message (e.g., Hellweg & Phillips, 1981; Tiemens, Hellweg, Kipper, & Phillips, 1985) rather than the extent to which they wielded persuasive influence on their own. Coding was also highly specific to debates and not generally transferable to other communication contexts.

The categories of analysis reviewed in this chapter overcome these limitations while lending themselves to repurposing in the form of stimulus clips for focus groups and eye-tracking experiments, as well as variables appropriate for use in predictive modeling of public response to campaign events via social media. As research utilizing biobehavioral measures of presidential debates and other political communication progresses, it will be important to enhance coding precision by reducing the standard length of individual coding segments from 30 seconds to as few as, say, 5 seconds to enable finer-grained analysis of communication behavior. A lot of expressive variability can happen in 30 seconds. Thus, in our coding scheme some segments were coded for multiple expressive behaviors—sometimes affiliative and reassuring, at other times defiant and threatening—all occurring in the same segment. Shortening segment lengths would give each segment more distinct meaning and discrete value. Another advance would be to record the durations of expressive displays rather than simply noting whether a given behavior or expression is present or absent, as we were able to do in our original image bites research (see Bucy & Grabe, 2007, 2008; Grabe & Bucy, 2009).

Here, automated coding would represent an enormous leap forward. At present, visual coding of presidential debates and other audiovisual content is a time-consuming and painstaking process. To ensure accuracy, multiple passes of the same content are required, and for every hour of content it takes at least twice that long to perform reliable coding manually. The ability to run a computer program that automatically documents the range of a candidate's expressive behavior would greatly accelerate the coding process while obviating the need for coder training or intercoder reliability checks. Instead of
training human coders, the program would simply need to be trained and validated (with human coding) to confirm recognition of the expressive variables of interest. Face-tracking software that locates and tracks facial features in video sequences in real time, such as that offered by Visage Technologies, hold considerable promise for the next wave of image bites research. And debates are a good testing ground for this because they feature an unobstructed, well-lit view of the candidates in relatively fixed positions with an invariant background.

Computer analysis of vocal variables, including voice inflections, tone, pace, and even decibel level, would also bring added precision and leverage to research on nonverbal behavior, in this case bioacoustics. The freely available Praat speech analysis software is well suited to this endeavor (see Boersma & van Heuven, 2001). Prior research has shown that presidential debates promote learning about candidate issue positions and influence evaluations of candidate traits (see Benoit, 2013). In both instances, behavioral indicators matter. In terms of the former, candidates can use changes in pitch and decibel level in tandem with evocative gestures to draw attention to an issue. In terms of the latter, candidates can raise and lower the tone of their voice to demonstrate a higher or lower level of emotional engagement. Facial displays of anger/threat combined with defiance gestures, or happiness/reassurance displays combined with affinity gestures, work in concert with voice to convey the communicator’s emotional state with added certainty. In either instance, changes in vocal inflection and nonverbal communication can influence how voters perceive presidential candidates.

To date, little research has considered such relationships. The advent and availability of social media archives, however, now makes it possible to model real-time responses to candidate communication on a mass scale. One possibility is to examine whether there is a biobehavioral aspect of agenda-setting influence. When candidates raise their tone of voice, for example, do these verbal punctuation marks act as an attentional cue that predicts Twitter response? Similarly, do vocal factors complement or compete with facial expressions and evocative gestures to enhance perceptions of leadership? Experimental research shows that candidates who speak with a lower pitch tend to receive higher leadership ratings and, consequently, are more likely to garner votes (Klofstad et al., 2012). In these studies vocal pitch is typically measured in isolation. However, pitch often covaries with decibel level (Gramming, Sundberg, Ternström, Leanderson, & Perkins, 1988). A higher vocal pitch combined with a higher
decibel level may lead to changes in perceptions of leadership, influencing the evaluation of traits associated with a dominant or lackluster performance.

Ultimately, the ability to perform automated coding of televised nonverbal behavior will continue to present new possibilities for research, allowing reconsideration and much more rigorous testing of long-accepted theories and assumptions—and development of new concepts and insights that were not evident before. As the small data of manual coding gives way to the big data of automated analysis, we should also be able to develop norms for political communication across different settings and contexts, so that a more precise form of discrepancy analysis may be performed on whether candidates are communicating within or outside the average range of typical political behavior, and how these fluctuations affect audience response.

NOTES

We wish to thank Riley Davis for his assistance with coding the 2012 presidential debates.

1. For examples of these candidate displays drawn from presidential election news coverage, see Grabe and Bucy (2009).

2. For an application of this approach to network news coverage of presidential campaigns, see Bucy and Grabe (2008) and Grabe and Bucy (2009).

3. The groups were organized and conducted in the context of a graduate seminar in political communication. No incentive was offered beyond course credit for each moderator. Thus, opinions offered were voluntary and not made in exchange for payment.

4. For details on the method, see Shah et al. (2015).

5. For an even finer-grained approach, see Nagel, Maurer, and Reinemann (2012).


7. www.praat.org

REFERENCES


and visual analysis of the Carter-Reagan debate. *Communication Quarterly, 33*(1), 34–42. http://dx.doi.org/10.1080/01463378509369576
