Facing the Electorate: Computational Approaches to the Study of Nonverbal Communication and Voter Impression Formation

Constantine Boussalis[1] and Travis G. Coan[2]

[1]Trinity College Dublin

[2]University of Exeter & Exeter Q-Step Centre

September 3, 2020

Abstract

Politicians have strong incentives to use their communication to positively impress and persuade voters. Yet, one important question that persists within the fields of political science, communication, and psychology is whether appearance or substance matters more during political campaigns. To a large extent, this appearance vs. substance question remains open and, crucially, the notion that appearance can in fact effectively sway voter perceptions is consequential for the health of democracy. This study leverages advances from the fields of machine learning and computer vision to expand our knowledge on how nonverbal elements of political communication influence voters immediate impressions of political actors. We rely on video from the 4^{th} Republican Party presidential debate held on 10 November 2016, as well as continuous response approval data from a live focus group (n=311; 36,528 reactions), to determine how the emotional displays of political candidates influence voter impression formation. Our results suggest that anger displays can positively influence viewers' real-time evaluations. Happiness displays, on the other

hand, are much less effective in eliciting a response from the viewing public, while fear displays were extremely rarely projected by the candidates of the debate under study.

Keywords: political communication, nonverbal communication, machine learning, computer vision

1 Introduction

Televised debates have become widely anticipated political campaign events, not only in modern American politics (Friedenberg 1997), but also in most Western democracies (Coleman 2000). In the US context, Presidential debates reach large audiences. For instance, the first debate between Donald Trump and Hillary Clinton in 2016 was watched by a record 84 million viewers, while the average audience size of US Presidential debates over the 1960-2016 period was over 66 million. The highly contentious nature of the debate allows for a direct comparison of candidate positions and, importantly, images. In effect, the televised debate serves as the ultimate job interview for the voter. Debates also seem to matter in non-trivial ways: (1) debates help voters learn about candidate policy positions (political knowledge); (2) debates influence how voters form their impressions of a candidate's personality, character, and competence; (3) debates affect vote choice (Lanoue and Schrott 1989, Hellweg et al. 1992, Zhu et al. 1994, Benoit et al. 2003).

Importantly, debates provide voters with the opportunity to observe candidates "up close." Candidates, of course, attempt to employ effective *rhetorical* strategies to consolidate their bases of support and to even persuade undecided voters to join their side. From another perspective, though, viewers of debates are also exposed to a deluge of *nonverbal* signals which are conveyed by the political candidates. A steady stream of smiles, winks, nods, frowns, and hand movements, among other displays, flood the television screens of the viewing public during these events.

Do these nonverbal signals impact whether voters express support for a candidate? A growing multidisciplinary literature from the fields of political science, psychology, and communications seeks to shed light on the extent to which appearance plays a role in how voters form their impressions of candidates. Theoretical and empirical scholarship suggests that nonverbal signals may have more of an impact than one might think. One popular example of this tension between substance and appearance is the oft-cited differing perceptions reported by American voters in 1960 on who won the first televised debate between John F. Kennedy and Richard Nixon. Survey data from that era suggest that those who listened to the debate on the radio thought that Nixon had won, while viewers of the televised debate favored Kennedy (Kraus 1996). While a number of studies since that era have cast doubt on this contrast to the point of classifying it as a "myth" (Bruschke and Divine 2017, Vancil and Pendell 1987), an experiment that used content from the Kennedy-Nixon debate found that subjects who watched the debate selected the winner based on perceptions of candidate personality, while those who only listened to the debate evaluated the winner based on both the discussed issues as well as personality (Druckman 2003). This suggests that the medium of consumption (visual versus audio) may have differential effects on how the audience forms its impressions of political candidates.

This study asks to what extent nonverbal signals from a given candidate during a televised debate might influence how voters form their level of support for the candidate. Here we follow a long line of research, rooted in ethology (Chance 1976, De Waal and Waal 2007, Van Hooff 1967) and social psychology (Lanzetta and Orr 1980, Bush et al. 1989), and extended to political psychology (Masters et al. 1986, Lanzetta et al. 1985, Sullivan and Masters 1988, Bucy and Grabe 2008, Bucy and Gong 2018), which studies how facial displays made by political elites—e.g., competitive displays which convey anger or affiliative facial expressions that signal happiness—influence the emotions, attitudes and preferences of voters.

While there is considerable interest in this area across many disciplines, existing studies on the effect of nonverbal political communication face methodological barriers that limit a comprehensive examination of the effect of candidate appearance on voter opinions and behaviors. Extant literature has relied on manual methods to record the various facial, gestural, and vocal displays conveyed by candidates during debates. This is an arduous and painstaking task whose cost in terms of time and resources can reach prohibitive levels at large scales. For this reason, studies that have sought to gather empirical data from televised debates, while valuable, have had to limit the scope of the analysis. This limitation has constrained our ability to make externally valid inferences. We seek to push the literature further in this space by relying on computational methods to help overcome major measurement challenges in the field. Our study, therefore, joins a recently developed and growing movement in political science which seeks to study "images as data" (Anastasopoulos et al. 2016, Torres 2018, Cantú 2019, Casas and Williams 2019, Zhang and Pan 2019; see Joo and Steinert-Threlkeld 2018 for a review) and, in particular, the computational study of political candidate facial expressions (Joo et al. 2019).

Another important area of research that has been understudied (see Bucy and Gong 2018) is how voters react to nonverbal displays of debate participants in real time. A number of studies have tracked how voters respond to candidates in real time in terms of agreement or disagreement with rhetoric (see Jarman 2005, Reinemann and Maurer 2005, Schill and Kirk 2014, Boydstun et al. 2014, Hughes and Bucy 2016, also see Schill et al. 2016 for an overview of real-time political communication research). Continuous response measurements of voter reactions to televised debates have also been used by media organizations when broadcasting debates (Kirk and Schill 2014). However, the integration of these real-time responses with nonverbal communication signals of US politicians is still in its infancy. Bucy and Gong (2018) investigates how voters react instantaneously to a selection of character and issue attack video clips from the third presidential debate of the 2016 general election. Again, while valuable, this study bases its inferences on a short temporal period because of the high cost of content analyzing nonverbal displays at the frame level of analysis for an entire debate.

Thus, our study contributes to the political nonverbal communication literature by applying computational methods to study the extent to which political candidates' nonverbal facial displays influence how voters construct and maintain their impressions of these candidates in real-time. By way of organization, Section 2 provides an overview of the theoretical foundations of the relationship between facial displays and impression formation of leaders, while also specifying our theoretical expectations. This is followed by a brief discussion in Section 3 of the particular campaign debate under study. Section 4 describes the data and statistical methods used to test our hypotheses. The results of these statistical tests are outlined in Section 5 and are followed by concluding remarks in Section 6.

2 Candidate facial displays and voter support

Can voter support really be influenced by the physical appearance of a candidate? The theoretical starting point in answering this question is an appreciation for the wealth of information contained within nonverbal signals. Nonverbal communication is ubiquitous in human communication. On a daily basis, we are faced with a deluge of direct and indirect nonverbal signals when interacting with others, be it a nod, a wink, a smile, a frown, breaking of eye contact, and so on. Studies have shown how upwards of two-thirds of all meaning in human interactions comes from nonverbal signals (Birdwhistell 1955, Burgoon et al. 2011). Seminal research from the field of psychology has emphasized the relative importance of nonverbal signals over verbal cues (e.g., Mehrabian and Wiener 1967, Argyle et al. 1970; 1971). Nonverbal communication can be understood as a fundamental form of human interaction, given that it likely predates verbal modes of communication within the species (McNeil 1970). The interpretation of nonverbal signals, such as facial expressions, is innate in humans and does not require any level of literacy or education—that is, nonverbal communication is accessible to all viewers, regardless of socioeconomic background or political sophistication (Messaris and Abraham 2001, Bucy and Stewart 2018 p. 4). Darwin (1872) describes how human facial expressions serve an evolutionary function as signals to be used for social behavioral coordination (Masters et al. 1986). Modern empirical research supports the notion that receivers use the emotive signal derived from a sender's facial expression to infer intention and trait characteristics of the sender (e.g., Camras 1980, McArthur and Baron 1983, Montepare and Dobish 2003, Todorov et al. 2008).

Translating this to the political world, voters use these accessible nonverbal signals as *heuristics*, or cognitive shortcuts, when forming their impressions of political candidates. In general, voters operate with relatively sparse information when forming their judgments of leaders (Carpini and Keeter 1996). In turn, people tend to rely on *heuristics* when making decisions with low information (Tversky and Kahneman 1974). Quick inferences of candidate traits based on visual cues conforms with the "dual-process" model of social cognition and decision-making (Kahneman 2003; 2011, Chaiken and Trope 1999), whereby initial judgments of people follow a "System 1" process that influences down-stream impression formation as a political campaign matures and more information about a candidate is gathered (see Todorov et al. 2005 p. 1623). Voters, in other words, can become "anchored" to initial impressions of candidates that are derived from nonverbal visual cues. Nonverbal signals conveyed by political candidates, therefore, can be interpreted as important heuristics for viewers of political debates when considering whom to support (Bucy and Newhagen 1999). As Masters (1989) puts it with respect to facial nonverbal signals, "since the capacity to decode and respond to facial displays is functionally necessary for normal social behavior, it should not be surprising that the facial displays of political leaders can influence observers' emotions and attitudes." (46)

There are two main strands of research on how nonverbal facial signals of candidates may influence voter support. The first area of research focuses on the static morphological features of candidates (Todorov et al. 2005, Olivola and Todorov 2010, Todorov et al. 2015). Researchers in this area investigate how structural facial attributes of candidates (e.g. attractiveness, dominance, and perceived trustworthiness) can serve as voting heuristics. Experimental studies have demonstrated that voters employ these visual heuristics to evaluate and select leaders (Efrain and Patterson 1974, Budesheim and DePaola 1994, Little et al. 2007; 2012, Antonakis and Eubanks 2017). Whether a voter is politically informed also seems to be a weak defense against these cognitive effects. Facial attractiveness, for instance, has been found to have a positive effect on how low-information voters form their impressions of candidates in general US elections; but this effect also persists among both low- and high-information voters in primary elections (Ahler et al. 2017). It has also been shown that voters are able to construct very rapid judgments (¡ 1 second) of candidate traits (e.g., competence) solely based on the image of the candidate and that these inferences predict actual electoral outcomes (Todorov et al. 2005, Olivola and Todorov 2010, Todorov et al. 2015). Research has even shown how children are able to successfully predict the outcome of elections from evaluations of candidate images (Antonakis and Dalgas 2009).

Our study is situated in the second strand of research which is concerned with the fluid and rapid facial signals of political actors and how voters react to this information. While speaking, individuals produce quick changes in muscle movements in the face which lead to the formation of a variety of facial displays that express information of emotional state and behavioral intent (Schmidt and Cohn 2001, Cohn and Ekman 2005, Stewart et al. 2011). These signals are then processed by the recipient (voter) and are used to help form an impression of the leadership attributes of the sender (candidate).

Researchers from the field of human ethology have documented evidence of an innate human repertoire of social behavior. The human smile, for example, seems to be an ingrained facial expression given that it can be observed in infants—even blind children who could not have learned it from viewing others (Hass 1970). This innate social behavioral repertoire, much like many aspects of life, also applies to leadership contests for social dominance. Facial displays of emotion are understood as "ritualized signals" that play an important role in contestation and maintenance of dominance relationships (Eibl-Eibesfeldt 1979). In their efforts to gain followers, prospective leaders must mold their nonverbal communicative strategies in such a way as to conform with two conflicting expectations held by prospective supporters. First, the leader must credibly signal an ability to dominate others, and thus show how he or she is able to neutralize external as well as internal threats to the group (Somit and Peterson 1997). At the same time, however, the candidate must also demonstrate an ability to affiliate with in-group members. That is, humans have an innate aversion to leaders who would exert extreme levels of control over the group (Boehm 1999, Stewart et al. 2011 p. 168).

To gain support, therefore, leaders must be able to signal their ability to both dominate and affiliate. When communicating with potential supporters, candidates tend to display signals which correspond to three functional categories of behavior: (1) anger, threat, or aggression; (2) fear, evasiveness, or flight; and (3) happiness, affiliation, and social reassurance (Masters 1989, Salter 1995, Stewart et al. 2011). These behavioral categories, which are seen in both humans and nonhuman primates, relate to what are commonly referred to as either *agonistic* (competitive) or *hedonic* (affiliative) behaviors (Chance 1976). Further, these categories, which help determine the establishment and maintenance of dominance hierarchies, in turn, can be mapped onto a subset of basic facial expressions of emotion (see Ekman and Friesen 2003). Namely, agonistic emotions are understood as occurring in "anger/threat" and "fear/evasion" displays, while "happiness/reassurance" displays correspond to hedonic emotions (Lanzetta et al. 1985, Stewart et al. 2011). These composite terms convey the dual signals being broadcast at any given moment by a speaker. That is, an emotional signal is being expressed via an "expression" and also a social signal is conveyed through a "display," although we can not determine true intent of expression but can only observe the emotive display as such (Way and Masters 1996, Bucy and Grabe 2008).

The meaning and optimal usage of these facial displays of emotion depends on the social context in which they arise and the relative hierarchical position of the sender at the time of communication (see Masters et al. 1987a, Masters 1989). A happiness/reassurance display can convey dominance as well as submission, depending on the context in which it is presented. That is, it can be used to reassure a person lower in the social hierarchy that a punishment is not likely, thus increasing group cohesion and increasing the sender's dominant position in the group (Stewart et al. 2011). Likewise, it can be perceived as sign of weakness if it is displayed in a competitive environment between rivals for the dominant position. A political candidate who is taking part in an interview with a neutral participant, such as a journalist, will be more inclined to display a happiness/reassurance display rather than an anger/threat facial expression (Masters et al. 1987a). However, in a competitive context, i.e. political speeches and debates, an anger/threat expression is more likely to present itself given that it is more effective in forming a closer bond between leaders and followers (Lorenz 1966, Lanzetta et al. 1985, Masters et al. 1987a). Though, anger/threat expressions are also negatively associated with perceived trustworthiness (Richell et al. 2005). The competitive context of a political debate, in other words, reduces the risk that a potential follower will misinterpret the anger/threat display as a sign of insecurity or hostility (Masters et al. 1987a), but it may impact how a follower trusts the candidate. Therefore, given that our study investigates competitive political speech, in the form of political campaign debates, we have the following expectations with respect to candidates' agonistic facial displays. First, with respect to anger/threat displays:

Hypothesis 1 *Anger displays made by a political candidate will increase voter support for that candidate.*

The theoretical and empirical literature, therefore, provides a set of expectations on how various agonistic candidate facial expressions, conditional on being presented in a competitive context, might materialize given a candidate's strategic understanding of how a voter might react when forming an impression. Since televised political debates can be understood as contests for social dominance, it is expected that a candidate's nonverbal communication should convey signals that conform with the agonistic nature of the environment. Of course, as discussed above, there should be limits to the effectiveness of signaling an anger/threat display in a political debate. Candidates in a democratic election, we argue, still need to express some form of reassurance to supporters in order to maintain group cohesion and to also attract undecided voters who might be wary of an overly assertive form of communication. For instance, empirical work has shown how leaders who display a smaller brow-to-eye distance while smiling tend to elicit stronger levels of support from viewers (Campbell et al. 1999, Trichas and Schyns 2012, Trichas et al. 2017). This combination of a smile and lowered brows has been characterized as the "Bill Clinton effect" and has been associated with heightened attributions of leadership traits by potential followers (see Senior et al. 1999). Therefore, we argue that happiness/reassurance displays will have a positive effect on how viewers perceive debate participants:

Hypothesis 2 *Happiness displays made by a political candidate will increase voter support for that candidate.*

The literature is also quite clear in the expectation that fear/evasion displays are likely to have a consistent negative effect on how a potential leader is perceived by the viewing public, regardless of context (see Bucy 2016). Unfortunately, as will be discussed further later on, we were not able to test this hypothesis because almost no fear/evasion displays were detected in our sample.

3 Case selection

Our study focuses on the fourth Republican primary debate (of 12) which was held on the night of November 10^{th}, 2015 in Milwaukee, Wisconsin. The debate was sponsored by Fox Business Channel and the *Wall Street Journal* and was moderated by Maria Bartiromo, Gerard Baker, and Neil Cavuto. An estimated 13.5 million people viewed this debate.[1] The main theme of the debate was the economy, with discussions ranging from tax policy, international trade, jobs, wages, and government spending; although, there were a few diversions such as international politics, the media, and Hillary Clinton. Eight candidates for the Republican presidential nomination took part in the debate: Donald Trump, Ben Carson, Marco Rubio, Ted Cruz, Jeb Bush, Carly Fiorina, Rand Paul, and John Kasich. To be eligible, participants were required to have at least an average approval rating of 2.5% in the four nationally representative polls prior to the debate. Figure 1 shows the RealClearPolitics Republican Presidential nomination poll averages as of November 9^{th}, 2015.[2] Going into the debate, there was no clear front runner among the candidates. Donald Trump and Ben Carson were essentially tied for first at around one-fourth approval each, with Marco Rubio trailing behind at roughly 12% approval.

The selection of this particular debate offers a number of advantages for examining the influence of emotional displays on candidate support. First, we are able

[1] Kissel, Rick (2015) "GOP Debate Ratings: Fox Business Network Draws Record 13.5 Million Viewers." *Variety*, accessed at https://bit.ly/2KKYJMl

[2] RealClearPolitics, "2016 Republican Presidential Nomination," accessed at https://bit.ly/2iiYMSZ

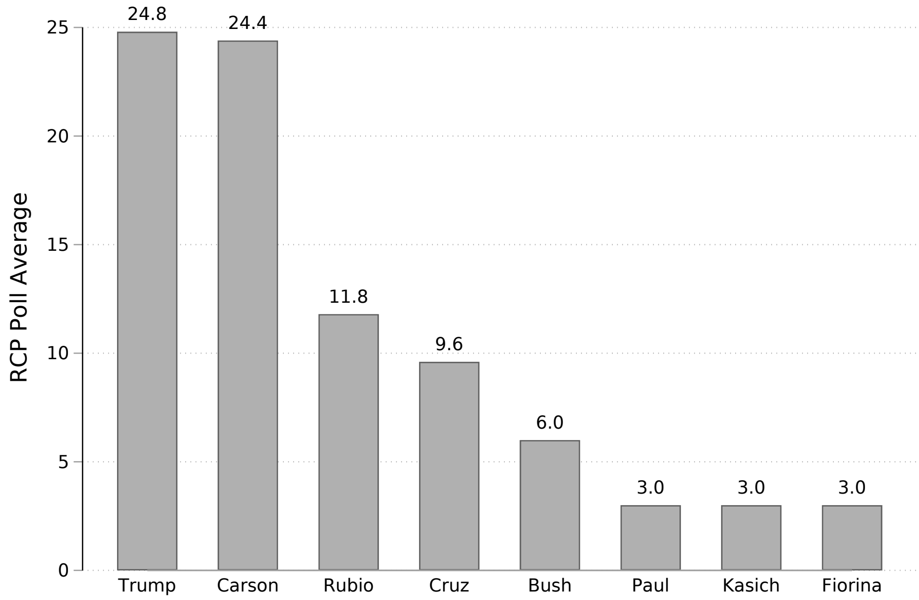

Figure 1: *Approval rating of candidates prior to the fourth GOP primary debate.* This figure displays the RealClearPolitics average approval rating for the eight debate participants as of the day before the debate (November 9th, 2015).

to observe reactions of viewers to a large number of candidates who were given more or less equal amount of speaking time.[3] This greatly expands the number of pairwise comparisons we can make between candidates and viewers. Second, our sample of candidates has at least some demographic variation, in that we observe three non-Whites (Ben Carson, Ted Cruz, and Marco Rubio) and one woman (Carly Fiorina). Third, by focusing on the Republican primary stage, we are able to control for party identification. By doing so we are making a clear trade-off between internal validity at the expense of external validity. Yet, this exchange allows us to normalize as much as possible prior attitudes towards the candidates, given their shared party identification.

[3]Speaking time was as follows, in descending order: Ted Cruz (13.6 min), John Kasich (11.9 min), Donald Trump (11.3 min), Carly Fiorina (11 min), Marco Rubio (10.3 min), Rand Paul (10.1 min), Jeb Bush (9.8 min), and Ben Carson (9.4 min).

4 Research Design

4.1 Data

We combine frame-level facial display data of political candidates with second-by-second continuous response measures of viewer reactions to debate participants to determine the extent to which facial expressions of candidates might influence voter support. This section offers a detailed description of the data collection process.

4.1.1 Real-time voter evaluations of candidate performance

Continuous response measures of respondents were gathered by G2Analytics, a focus group research firm, on the night of the fourth Republican presidential nomination debate which took place on November 10, 2015. The focus group consisted of 311 viewers who identified as either Republican or Independent. Respondent demographic summary statistics are provided in Table 1.

Viewers were provided a hand-held device and instructed to record through the device whether they agree ("bell") or disagree ("buzzer") with anything they heard while viewing the debate live. In total, the dataset includes 36,528 continuous responses from the focus group: 26,753 (73.24%) bells and 9,775 (26.76%) buzzers. Roughly 98% of all respondent-seconds were absent a reaction. We employ these second-by-second responses as the dependent variable of our analyses. Specifically, we operationalize signals of increased "support" for a candidate as a bell and reduced support as a buzzer, while a neutral stance is defined as a second in time with neither a bell nor a buzzer.

4.1.2 Facial displays of emotion

Currently, scholarship has been hampered by the extremely high cost associated with the manual measurement of facial displays of emotion. The manual coding of these nonverbal signals at the frame-by-frame level is a painstakingly long process which requires non-trivial investments in time and resources. It is estimated that the coding of a relevant frame using the industry standard Facial Action Coding System (Ekman and Friesen 2003) can take up to 10 minutes on average (Stewart

Variable		**Frequency**	**Percent**
Affiliation	Republican	199	63.99
	Independent	112	36.01
Gender	Female	180	57.88
	Male	131	42.12
Race	White non-Hispanic	251	80.71
	Hispanic	25	8.04
	African-American	18	5.79
	Asian	11	3.54
	Other	6	1.93
Age	18-24	25	8.04
	25-29	43	13.83
	30-39	84	27.01
	40-49	44	14.15
	50-65	88	28.3
	65+	27	8.68
Region	West	60	19.29
	Mid-West	69	22.19
	South	125	40.19
	North East	57	18.33

Table 1: *Focus group participant demographics.* This table provides demographic summary statistics of the 311 Republican and Independent voters who participated in the November 10, 2015 focus group and whose continuous response measures are used in the study.

et al. 2011). In many instances, it may take a manual coder multiple passes over the same content to ensure measurement validity (Bucy and Stewart 2018). At large scales, such as every frame in an entire debate or set of debates, the cost associated with measurement becomes prohibitive. As such, scholars have had to constrain the scope of inquiry to still images (Olivola and Todorov 2010), short video segments (Bucy 2016), or very coarse periods of time (Bucy and Gong 2018).

We seek to overcome these challenges by drawing on methods from computer vision and machine learning to automatically extract facial displays of emotion of the participants in the fourth 2016 Republican primary debate. In general terms, the task involved 1) extracting all the frames from the video and 2) passing each frame though a detection system which produces confidence estimates of the nonverbal facial displays of emotion for each candidate.

While there are a number of approaches available to automatically detect emotive facial displays, we relied on the Microsoft Face API, which was developed by Microsoft Research and is offered as a cloud service via Microsoft Azure. In contrast to some other proprietary software developed to detect facial emotional displays, the Microsoft Face API is based on discrete emotion theory and is thus consistent with much of the theoretical and empirical literature on nonverbal communication of debate participants. In particular, Microsoft Face API provides a robust emotion recognition capability that uses a modified form of the Ekman and Friesen (2003) model of discrete facial expressions. For each frame, the system is able to automatically recognize a given speaker and assign to them a confidence score for each of the following emotions: anger, contempt, disgust, fear, happiness, neutral, sadness and surprise. It should be noted that Microsoft has added "contempt" and "neutral" categories to the Ekman and Friesen model.

Although the source code underlying the model is not available in the public domain, it is known that the emotion detection functionality of the Microsoft Face API relies on a deep convolutional neural network (CNN) (Bargal et al. 6 10). CNN is a popular model used when analyzing and classifying visual information (LeCun et al. 1998; 2015). While we do not know the exact architecture of the Microsoft Face API algorithm, generally speaking, a convolutional neural network consists of an input layer (i.e. an array of image pixel values, typically with red, blue, and green channels), that then moves on to a series of hidden layers, which include a variety of convolutional, pooling, and fully connected layers, and ends with an output layer that contains the confidence of a given class (Krizhevsky et al. 2012). As is the case with other neural networks, the CNN model is trained with a set of labeled data. Learning is achieved through backpropagation, which commonly uses gradient descent optimization (LeCun et al. 1998).

A video file of the debate (without commercial breaks) which matches the timestamps for all bells and buzzers in the continuous response measure dataset was provided by G2Analytics. The total runtime of the debate video is 1:58:41 at 29.97 frames per second. With this video in hand, we were able to automatically classify 1) the speaker in a given frame and 2) the primary facial display of emotion using the following procedure. First, we used the `OpenCV` library in Python to extract all frames from the video ($n = 213,422$) as JPEG files. Next, using a simple Python script, we passed each frame to the Microsoft Face API. For each frame we were able to retrieve: 1) a face identification estimate[4]; and 2) a confidence score, which ranges within the interval $[0, 1]$ for each of the eight facial displays of emotion described above.[5] For each frame, we then generate a dummy variable for each emotion which takes on the value of 1 if it has the highest confidence score. Using these dummy variables, we take the average maximum emotion score for each second (30 frames). We then standardize the mean emotion scores and use them as covariates in our statistical models which are discussed below.

Descriptive statistics of the extracted facial displays of emotion for the candidates from the above procedure are displayed in Figure 3. The most prevalent

[4]We uploaded 10 photos of each of the eight candidates (Donald Trump, Ben Carson, Marco Rubio, Ted Cruz, Jeb Bush, Carly Fiorina, John Kasic, and Rand Paul) and three journalists (Maria Bartiromo, Gerard Baker, and Neil Cavuto) to the Microsoft Identify API to train a person identification model. To validate the accuracy of the Identity API, the authors annotated a random sample of 250 frames, coding for the presence of the eight candidates and three moderators. If the frame included a face other than the candidates or moderators, the face was coded as "unknown," while frames without a face were coded as "None". Frames with multiple faces (e.g., more than two candidates) were labeled to include each face that was present. The results of this exercise suggest that the face detection model was extremely accurate: with a micro-averaged precision of 0.96, recall of 0.94, and an F1-score of 0.95. This model was ultimately used to identify each face in a given frame and provide a confidence score for the match.

[5]Note that we also were able to recover the head position of the detected face in each frame (yaw, roll, and pitch), however we do not use this information in the current study.

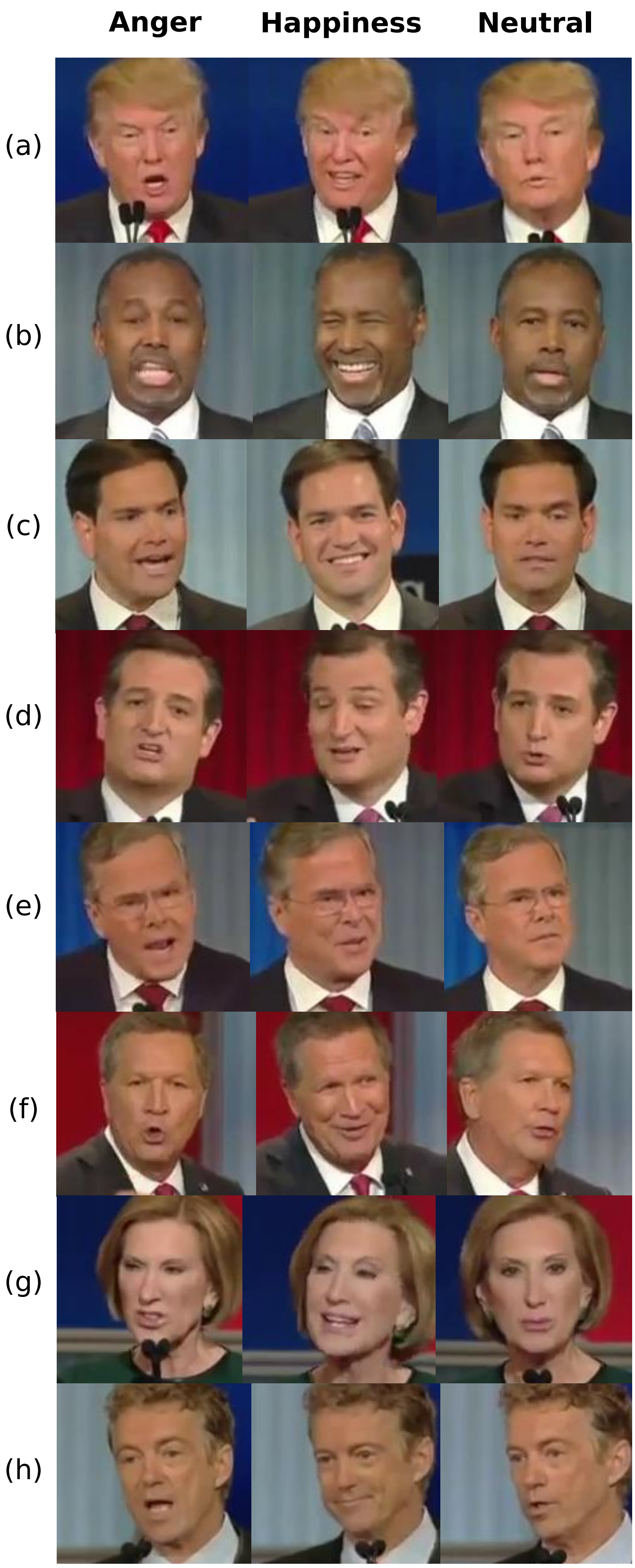

Figure 2: *Candidate images with high confidence scores for anger, happiness, and neutral displays.* This plot provides examples of frames with high confidence scores of anger, happiness, and neutral displays for (a) Donald Trump, (b) Ben Carson, (c) Marco Rubio, (d) Ted Cruz, (e) Jeb Bush, (f) John Kasich, (g) Carly Fiorina, and (h) Rand Paul. We note that there were no frames in the debate where the dominant display of emotion was anger for Marco Rubio and Rand Paul. Therefore, for these candidates we selected the frame with the highest level of confidence of anger to be displayed in the figure.

Happiness		Anger	
Candidate	*RMSE*	*Candidate*	RMSE
Jeb Bush	0.33	Rand Paul	0.33
Marco Rubio	0.37	Ted Cruz	0.47
Ted Cruz.	0.39	Ben Carson	0.47
Donald Trump	0.40	Marco Rubio	0.60
John Kasic	0.50	John Kasic	0.69
Rand Paul	0.55	Carly Fiorina	0.81
Ben Carson	0.62	Donald Trump	0.86
Carly Fiorina	0.66	Jeb Bush	0.88

Table 2: *Comparing the Microsoft Face API to Human Judgements.* This table provides root mean squared error (RMSE) estimates for happiness and anger displays across each candidate. The RMSE estimates were constructed using the Face API confidence score to predict the human judgements. As such, the RMSE these estimates use a 5-point emotion scale.

facial display for most candidates throughout the debate, unsurprisingly, was a neutral display; though, the classifier coded a significant proportion of frames as non-neutral displays for Trump and Cruz. Turning to the emotive displays of interest in this study—i.e., anger, happiness, and fear—we find that: 1) Trump, by far, showed the most anger displays (6.2% of his frames were coded as having a dominant anger display); 2) Marco Rubio showed the most happiness displays (4.9% frames) while Trump displayed the least happiness (0.9% frames); and 3) fear displays were essentially non-existent in the debate and therefore we were not able to test its effect on support.

Validating the Face API. How well does the Face API perform in extracting the relevant emotions from our video content? We attempt to answer this question in two steps. First, to get an overall sense of the models ability to detect happiness and anger, as well as providing a visual illustration of the expression of these emotions, Figure 2 presents an array of example frames that were classified with high confidence scores for anger, happiness and neutral displays for the eight candidates. Here we can see that, generally speaking, Figure 2 provides *prima facie* evidence of the model's ability to detect facial displays of emotion.

While the results in Figure 2 are encouraging, this is clearly not enough—it is necessary to estimate the extent to which the API predictions correspond to human judgements. We therefore expand the validation exercises by using the following four step procedure (for details on the validation procedure, see Appendix ??). First, we select all the frames where the computer recognizes a given speaker and collapsed these frames into five second sequences in GIF format (i.e., 150 frames per sequence). Second, in order to ensure that all candidates and relevant emotions are sampled, we take a stratified random sample of five-second clips for each candidate and emotion ($n = 240$). Third, we recruited six research assistants to judge the extent to which a given clip contains anger and happiness displays, using a 5-point Likert scale ranging from "Not at all" to "Extremely". Lastly, we assess the model's ability to predict the human judgements using the root mean squared error (RMSE) (see Table 2). Overall, the model appears to perform reasonably well. In general, we observe lower prediction error for happiness displays (with RMSEs ranging from 0.33 (Bush) to 0.66 (Fiorina)) than for anger (with RMSEs ranging from 0.33 (Paul) to 0.88 (Bush)). However, even at its worst (i.e., anger displays for Bush), the model misses by an average of less than one point (0.88) on a 5-point scale (e.g., predicting "moderately angry" instead of "very angry"). While there is certainly room for improvement in terms of performance, the results provided in Table 2 allow us to proceed with a suitable level of confidence in the emotion measures.

4.1.3 Negative sentiment of speech

Following work by Nagel et al. (2012) and Bucy and Grabe (2008), our study controls for the verbal tone, specifically the negative sentiment, of candidate utterances. We measure this variable via a dictionary approach. The first task was to obtain the text of candidate utterances. We relied on the Google Cloud Speech-to-Text service to transcribe the audio of the debate video. The Google API returns a transcript with a range of timestamps for each transcribed word. Using these timestamps we were able to reconstruct every sentence along with their time range. These sentences were then merged with our second-by-second dataset (including duplicates of a given sentence over its range of seconds). Next, we calculate the share of words in each sentence that match terms that are part of the "Negative emotion" category of the Linguistic Inquiry and Word Count (LIWC) dictionary

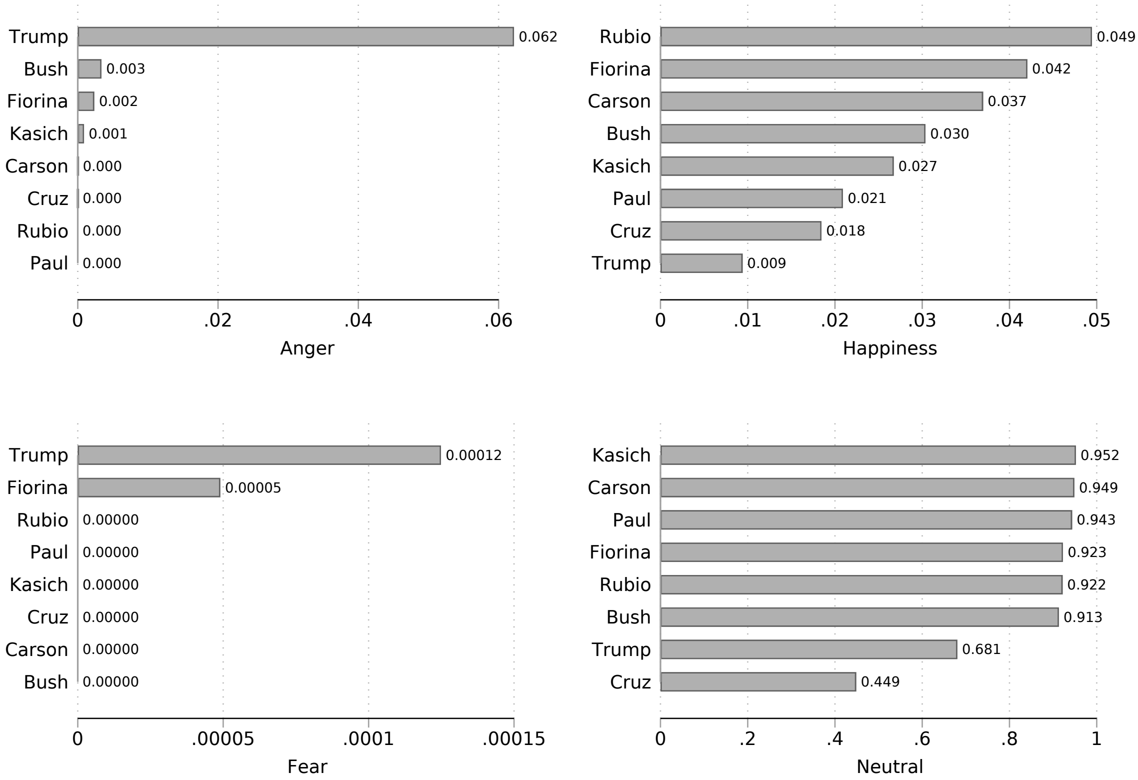

Figure 3: *Average frame-level facial displays of emotion.* These graphs display the proportion of frames that were classified as having a dominant anger, happiness, fear, or neutral display for all eight candidates. Note that these plots are not on a common scale.

(Pennebaker et al. 2001). Similar to facial displays of emotion, Donald Trump is also a clear leader in terms of spoken negative sentiment (2.39%), and is followed by Kasich (1.91%), Paul (1.83%), Bush (1.73%), Fiorina (1.65%), Carson (4.41%), Rubio (1.33%), and Cruz (1.30%).

4.2 Methods

The focus group data present a number of challenges for estimating the influence of facial expressions on candidate support. The structure of these data is complex: the second-by-second ordinal response data (disagree, neither agree nor disagree, or agree) are nested within a particular respondent, directed at a specific candidate, and divided across a number of discrete "segments" (e.g., periods of continuous debate, which are then interrupted by commercial breaks or the debate moderators). We account for this structure in a number of ways. First, following previous scholarship on the use of continuous response measures in political debates (Nagel et al. 2012), we a) estimate separate models for each of the eight candidates included in the debate and b) treat each individual "segment" as a discrete period of time.[6] For instance, Donald Trump's first segment starts at 5:54 and runs for 54 seconds, his next segment starts at 28:13 and runs for 103 seconds, and so on. Second, to account for the multilevel structure of the response data, we estimate a series of mixed effects ordered logistic regression models, which include a random intercept for each focus group participant and fixed effects for the theme being uttered by a candidate. To gather data on the thematic substance of the utterances, we perform a content analysis of each candidates' transcript and classify each sentence into one of 16 themes of discussion.[7] We also include key political and demographic covariates available in the G2Analytics data, including party identification, age, ethnicity, and gender.[8]

An added complication is that these data are inherently dynamic and the lag time between a candidate's facial expression and a focus group participant's behavioral response (if there is one at all) is not known *a priori*. In their analysis of the 2005 German National Election, Nagel et al. (2012) find that a three second lag best describes the dynamics between non-verbal messages and viewer responses. Nagel et al. (2012) assume a quadratic distribution for the implied lag weights and assume—rather than estimate—the underlying lag distribution (i.e., $t_0 = 0.5$, $t_{-1} = 1$, and $t_{-2} = 0.5$). In the analysis described in the next section, we follow a similar process as Nagel et al. (2012), but instead estimate the lag distribution. First, we follow the standard procedure when estimating distributed lag models

[6] Debate participant segments were coded manually by a researcher using the following decision rules. A new segment begins when a candidate is first heard or seen. Likewise, the segment ends when the same candidate is finished answering a question and there are no follow up questions from one of the journalists. If there is a follow-up question, this time is included in the segment. In the very rare circumstance that another candidate interrupts, we decided to attribute the segment to the original candidate unless the interruption is prolonged, whereby the segment changes to the new candidate.

[7] In particular, the themes include: Climate change, Closing statements, Entitlements, Financial crisis and bailouts, Hillary Clinton, Immigration, Income inequality, Media portrayal of candidates, Minimum wage, National debt and federal spending, National security and foreign affairs, Regulations on businesses, Tax loopholes and businesses leaving, Taxes, Technology impact on jobs, and the Trans-Pacific Partnership (TPP). These topics were derived from an initial list of topics generated by Ballotpedia (see `https://bit.ly/2VHYvv4`)

[8] Political and demographic covariates are measured as follows: *Republican* (1 = Republican; 0 = Independents), *age dummies* (18-24, 25-29, 30-39, 40-49, 50-65, 65+), *race dummies* (African-American, Asian, Hispanic, Other, White), and *female* (1 = female, 0 = male).

and determine the lag length via information criteria; this analysis suggests an "optimal" lag length of roughly 4 seconds. Second, we estimate the lag weights for 4 seconds, assuming a quadratic lag distribution.[9] Note that this specification is not only consistent with past empirical on non-verbal communication, but offers a good deal of flexibility regarding the dynamic relationship between candidate facial expressions and participant support.

5 Results

The key empirical findings are provided in Figure 4 and Table **??** in the appendix. The figure displays the estimated odds ratios of agreement with candidate statements for anger and happiness displays as well as candidate utterance negative sentiment, providing the estimated lag distribution for four seconds. It is important to consider the underlying dynamics when interpreting the results in Figure 4. For instance, consider the estimates for Donald Trump and expressions of anger. When Trump displays anger, the "instantaneous" (t_0) effect is an increase of 1.03 in the odds ratio of clicking "agree." In the next second (t_1), the estimated increase in the odds ratio of clicking agree is even higher (OR = 1.05). However, by the 4^{th} second, the effect has dropped considerably (OR = 1.007) and is statistically indistinguishable from 1 (p = 0.586). This inverted-U shaped temporal effect is

[9] In principle, the lag distribution may be estimated directly by incorporating a separate regressor for each second (1 to 4). In practice, however, multicollinearity makes direct estimation difficult (if not impossible) and thus it is common practice to place constrain the lag distribution to promote smoothness. Following past scholarship, we assume a quadratic lag function:

$$\beta_s = \lambda_0 + \lambda_1 s + \lambda_2 s^2, s = 0, ..., 4, \tag{1}$$

which we can then plug into the (ordered logistic) regression equation:

$$y_t = \alpha + \sum_{s=0}^{4} (\lambda_0 + \lambda_1 s + \lambda_2 s^2) x_{t-s} + \epsilon_t, \tag{2}$$

which one can re-write as:

$$y_t = \alpha + \lambda_0 z_t^0 + \lambda_1 z_t^1 + \lambda_2 z_t^2 + \epsilon_t, \tag{3}$$

where

$$z_t^0 = \sum_{s=0}^{4} x_{t-s}, z_t^1 = \sum_{s=0}^{4} s x_{t-s}, z_t^2 = \sum_{s=0}^{4} s^2 x_{t-s}. \tag{4}$$

We can then estimate (4) by standard models and recover the lag weights using (1). Uncertainty of the estimates can be assessed by examining the linear combination of coefficients expressed in (1).

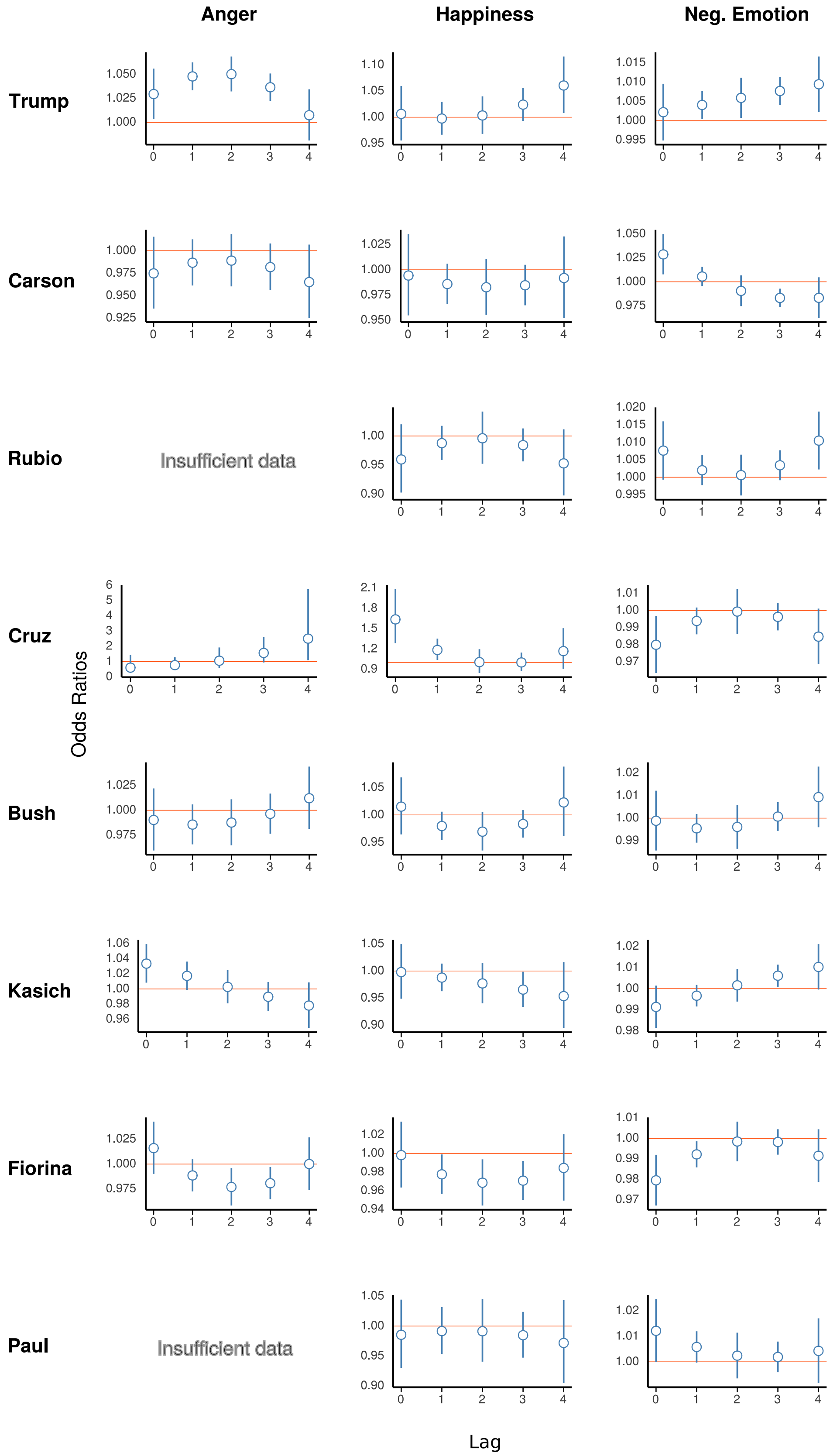

Figure 4: *Estimating the influence of facial displays of emotion on immediate impressions.* This figure displays the estimated lag distribution ($t_0 \dots t_4$) for each candidate and each emotion variable (anger/happiness facial display and negative emotion words). It provides the estimated odds ratios from a mixed effects ordered logistic regression (circles) and 95% confidence intervals (blue line), assuming a quadratic lag structure and a random intercept for each focus group participant. All models include the following covariates: party identification, age, race, and gender. The models also include topic-level fixed effects. Note that these plots are on different scales.

consistent with the assumed relationship put forth by Nagel et al. (2012).

Aside from the anger effect of Donald Trump, which is the clearest and most consistent finding among our models, a number of other anger display effects are observed. Viewers also responded positively to Kasich anger displays, although the dynamics are different to those of Donald Trump. A statistically significant and positive instantaneous effect is observed (OR = 1.03), but the effect is indistinguishable from 1 in the next and following time lags. The effect for Kasich dissipates, therefore, over time, while that of Trump follows a curvilinear path. Interestingly, we find a negative effect of anger on support for Fiorina. The effect of a Fiorina anger display is negative at t_2 (OR = 0.98) and t_3 (OR = 0.98).

Turning to the effects of hedonic signals, we find that viewers were more likely to reward some candidates while punishing others when exposed to a happiness display. Our models indicate a statistically significant positive effect of happiness facial displays for Ted Cruz t_0 (OR = 1.64) and t_1 (OR = 1.19), after which the effect becomes indistinguishable from 1. Our results also show a positive effect of a happiness display for Trump at t_4 (OR = 1.06). Similar to her agonistic displays, our results suggest that Carly Fiorina was also penalized for her displays of happiness. The model results indicate a negative effect at t_1 (OR = 0.98), t_2 (OR = 0.97), and t_3 (OR = 0.97).

Figure 4 also displays the results for an important candidate-level control variable: the negative sentiment of candidate utterances. Here, again, we find conflicting results whereby some candidates benefited from the use of negative emotive verbal communication, while others were more likely to lose support by using such language. Donald Trump's use of words with negative sentiment is positively associated with viewer approval, starting from t_1 (OR = 1.004), and increasing in strength in t_2 (OR = 1.006), t_3 (OR = 1.008), and t_4 (OR = 1.009). Similarly, Marco Rubio's negative tone also had a positive effect on viewer support, but later on at t_4 (OR = 1.01). On the other hand, Carly Fiorina, Ted Cruz, and Ben Carson were less likely to receive viewer support following a negative utterance. Specifically, negative sentiment of speech is negatively related with support for Fiorina at t_0 (OR = 0.98) and to a less extent at t_1 (OR = 0.99). The results for Cruz also suggest a negative effect of negative sentiment, however this is limited to an instantaneous effect (OR = 0.98), with all other time lags being indistinguishable from 1. Lastly, we find a shifting downward effect for Carson; at t_0 negative sentiment is positively related with support (OR = 1.03) but this effect becomes negative at t_3 (OR = 0.98).

With respect to time-invariant focus group participant demographic effects, we find that females were more likely to agree with Jeb Bush (log-odds = 0.26; [relative risk ratio = 1.3]) and Marco Rubio (0.24; [1.27]). Relative to Independents, self-identified Republicans were more likely to agree with Bush (0.38; [1.47) and Ben Carson (0.44; [1.56]) and less likely to agree with John Kasich (-0.26; [0.77]) and Rand Paul (-0.21; [0.81], $p<0.10$). African-American respondents were less likely than Whites to agree with Donald Trump (-0.67; [0.51), Rubio (-0.44; [0.65], $p<0.10$), and Paul (-0.47; [0.63]). Lastly, generally speaking and relative to 30-39 year-olds, older respondents were more likely to agree with Rubio, Cruz, Carson and Trump.

6 Conclusion

Our results suggest that facial signals of emotion by participants in a televised political debate may influence how viewers evaluate candidate performance. To our knowledge, this study was the first to directly compare agonistic and hedonic facial displays with continuous response measurements of viewer reactions at the second-by-second level of analysis for the full duration of a major televised political debate. We accomplished this task by relying on theoretically informed, machine learning applications used to detect and classify facial emotive displays.

The main finding of the study is that anger/threat displays by Donald Trump and John Kasich had a positive effect on the likelihood of immediate support from focus group viewers. These results conform with other studies in the literature which have investigated a similar relationship. For instance, Tiedens (2001) found that experimental subjects who were exposed to an angry response by Bill Clinton to the Monica Lewinsky scandal, in contrast to a sad one, were more likely to directly support Clinton as well as indirectly support him via a higher attribution of competence.

In contrast, our results suggest that Fiorina's anger displays tend to decrease viewer support, which is relevant to an ongoing debate on whether counterstereotypical gender behavior among political elites leads to disproportionate penalties at the ballot box (see Hitchon et al. 1997, Brooks 2011; 2013, Bauer 2017, Fridkin et al. 2019, Cassese and Holman 2019). This tentative and objectively narrow result suggest that aggressive displays (or even any emotive signal more generally) by female candidates—at least among Republican contenders—might be counterproductive.

The weak effect of happiness/reassurance displays on viewer support, aside from that found for Ted Cruz, is generally inconsistent with past scholarship. To refresh, theories from human ethology and social psychology would lead us to believe that happiness/reassurance displays should increase positive affect toward to the displaying candidate, and this in turn should increase support. So why are we not observing a strong positive response to these signals in our sample? In their classic study, Sullivan and Masters (1988) found that happiness/reassurance displays by Ronald Reagan had a much stronger impact on post-exposure attitudes than those of his Democratic rival, Walter Mondale. This difference was attributed to the ability of Reagan to project a clearer display of emotion than his counterpart. In our study, our reliance on an automated system that ascribes a confidence score on a set of eight emotions, as well as our choice of collapsing a given frame's classification on the strongest confidence score, limits our ability to determine the effect of clear versus unclear signals on viewer responses. However, we relax some of these coding rules and focus on the raw confidence scores for analysis in the appendix. The results of these models are generally very similar to those reported above, with some notable differences. Carly Fiorina's anger effect is no longer statistically significant and we also find a positive effect of anger for Ben Carson and Rand Paul. As a further complication, Bucy and Grabe (2008) show how supporters of a candidate will decode facial displays of candidates differently than opponents; although, their study suggests that happiness/reassurance displays are able to neutralize negative feelings of critics (92). Unfortunately, our data do not allow for us to control for a given viewer's prior attitudes toward a given candidate.

Our study also provides a set of descriptive findings that speak to the extant biopolitics literature on emotive displays and campaigns. First, previous work showed how front-runners in political campaigns are more likely to display (or be shown to display by the media) hedonic rather than agonistic facial expressions (see Sullivan and Masters 1988, Bucy and Grabe 2008). In contrast, the descriptive findings of our study show how Donald Trump, who was slightly ahead of Ben Carson at the time of the debate, was among the least likely to display a hedonic display throughout the debate and also signaled the most anger, both verbally and nonverbally. Though, it is possible that this contrary finding is due to the fact that the observations gathered for analysis were from a relatively early period of the campaign. Bucy and Grabe (2008), for instance, finds that as Election Day approaches, hedonic displays by candidates are much more likely. This change of tune might be the result of a last minute strategy of cohesion building among supporters. Second, the Face API found almost no displays of fear for the candidates in the fourth GOP primary debate. While theory would suggest that these displays should have a detrimental effect on viewer impressions, since they signal submissiveness, we were not able to test this proposition due to extreme missingness. While surprising, a similar descriptive account was observed for televised coverage of speeches and interviews during the 1984 Democratic primary race. Specifically, Masters et al. (1987b) found that Walter Mondale expressed the most fear/evasion displays during his speeches (3% of his total displays) and only 1% of his emotive displays during interviews signalled fear/evasion. Gary Hart showed even less fear: no such displays during his speeches and only 1% of his displays in interviews showed fear. The third candidate under study, Jesse Jackson, showed no fear/evasion displays in any of his interviews and speeches. It seems, therefore, that candidates for the highest political office in the United States are (increasingly) more resistant to "leaking" fear displays to the viewing public when publicly competing for social dominance.

The current study is certainly not without its limitations. While indicative of increased support, we should stress that our results suggest small effect sizes and that based on the data available to us, we cannot make claims about the persistence of these effects beyond the duration of the debate. Aside from measurement issues discussed above, future work would benefit from additional control variables. Vocal characteristics of candidate utterances (e.g., pitch, tone, volume), for instance, should be included, as there is evidence that these factors resonate in politics (see Nagel et al. 2012, Dietrich et al. 2017). Further, environmental effects on viewer perceptions, e.g., studio audience laughter, booing, and applause (Fein et al. 2007, Stewart et al. 2018), should also be brought in to future studies.

We believe that this study offers a novel first step in increasing our ability to observe and compare candidate appearance effects on voter impressions. Further, the automated systems employed in this project can be applied to other politically relevant visual environments, from nightly news reports to joint press conferences of world leaders to parliamentary debates. Our study, therefore, hopes to show the way for other related studies in varying contexts. Lastly, we hope that our study will not only engage with the extant literature on nonverbal political communication which covers a range of disparate academic disciplines (political science, psychology, communications), but also will help inform political practitioners such as political campaign managers and press officers.

References

Ahler, D., Citrin, J., Dougal, M., and Lenz, G. (2017). Face value? experimental evidence that candidate appearance influences electoral choice. *Political Behavior*, 39(1):77–102.

Anastasopoulos, L. J., Badani, D., Lee, C., Ginosar, S., and Williams, J. (2016). Photographic home styles in congress: a computer vision approach. *arXiv preprint arXiv:1611.09942*.

Antonakis, J. and Dalgas, O. (2009). Predicting elections: Child's play! *Science*, 323(5918):1183–1183.

Antonakis, J. and Eubanks, D. (2017). Looking leadership in the face. *Current Directions in Psychological Science*, 26(3):270–275.

Argyle, M., Alkema, F., and Gilmour, R. (1971). The communication of friendly and hostile attitudes by verbal and nonverbal signals. *European Journal of Social Psychology*, 1(3):385–402.

Argyle, M., Salter, V., Nicholson, H., Williams, M., and Burgess, P. (1970). The communication of inferior and superior attitudes by verbal and nonverbal signals. *British journal of social and clinical psychology*, 9(3):222–231.

Bargal, S., Barsoum, E., Ferrer, C., and Zhang, C. (2016-10). Emotion recognition in the wild from videos using images. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, page 433–436.

Bauer, N. M. (2017). The effects of counterstereotypic gender strategies on candidate evaluations. *Political Psychology*, 38(2):279–295.

Benoit, W., Hansen, G., and Verser, R. (2003). A meta-analysis of the effects of viewing us presidential debates. *Communication Monographs*, 70(4):335–350.

Birdwhistell, R. (1955). Background to kinesics. etc: A review of general semantics.

Boehm, C. (1999). Hierarchy in the forest: Egalitarianism and the evolution of human altruism.

Boydstun, A., Glazier, R., Pietryka, M., and Resnik, P. (2014). Real-time reactions to a 2012 presidential debate: A method for understanding which messages matter. *Public Opinion Quarterly*, 78(1):330–343.

Brooks, D. J. (2011). Testing the double standard for candidate emotionality: Voter reactions to the tears and anger of male and female politicians. *The Journal of Politics*, 73(2):597–615.

Brooks, D. J. (2013). *He runs, she runs: Why gender stereotypes do not harm women candidates*. Princeton University Press.

Bruschke, J. and Divine, L. (2017). Debunking nixon's radio victory in the 1960 election: Re-analyzing the historical record and considering currently unexamined polling data. *The Social Science Journal*, 54(1):67–75.

Bucy, E. (2016). The look of losing, then and now: Nixon, obama, and nonverbal indicators of opportunity lost. *American Behavioral Scientist*, 60(14):1772–1798.

Bucy, E. and Gong, Z. (2018). In/appropriate aggression in presidential debate: How trump's nonverbal displays intensified verbal norm violations. In *2016. In The Facial Displays of Leaders*, page 73–95. Palgrave Macmillan, Cham.

Bucy, E. and Newhagen, J. (1999). The emotional appropriateness heuristic: Processing televised presidential reactions to the news. *Journal of Communication*, 49(4):59–79.

Bucy, E. and Stewart, P. (2018). The personalization of campaigns: Nonverbal cues in presidential debates. *Oxford Research Encyclopedias Series.*

Bucy, E. P. and 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(1):78–98.

Budesheim, T. and DePaola, S. (1994). Beauty or the beast? the effects of appearance, personality, and issue information on evaluations of political candidates. *Personality and Social Psychology Bulletin*, 20(4):339–348.

Burgoon, J. K., Guerrero, L. K., and Manusov, V. (2011). Nonverbal signals. *The SAGE handbook of interpersonal communication*, pages 239–280.

Bush, L. K., Barr, C. L., McHugo, G. J., and Lanzetta, J. T. (1989). The effects of facial control and facial mimicry on subjective reactions to comedy routines. *Motivation and emotion*, 13(1):31–52.

Campbell, R., Benson, P., Wallace, S., Doesbergh, S., and Coleman, M. (1999). More about brows: How poses that change brow position affect perceptions of gender. *Perception*, 28(4):489–504.

Camras, L. (1980). Children's understanding of facial expressions used during conflict encounters. *Child Development*, page 879–885.

Cantú, F. (2019). The fingerprints of fraud: Evidence from mexico's 1988 presidential election. *American Political Science Review*, 113(3):710–726.

Carpini, M. and Keeter, S. (1996). *What Americans know about politics and why it matters.* Yale University Press.

Casas, A. and Williams, N. W. (2019). Images that matter: Online protests and the mobilizing role of pictures. *Political Research Quarterly*, 72(2):360–375.

Cassese, E. C. and Holman, M. R. (2019). Playing the woman card: Ambivalent sexism in the 2016 us presidential race. *Political Psychology*, 40(1):55–74.

Chaiken, S. and Trope, Y., editors (1999). *Dual-process theories in social psychology.* Guilford Press.

Chance, M. (1976). The organization of attention in groups. methods of inference from animal to human behavior.

Cohn, J. and Ekman, P. (2005). Measuring facial action. the new handbook of methods in nonverbal behavior research.

Coleman, S. (2000). Televised election debates. international perspectives.

Darwin, C. (1872). The expression of the emotions in man and animals.

De Waal, F. and Waal, F. B. (2007). *Chimpanzee politics: Power and sex among apes*. JHU Press.

Dietrich, B. J., Hayes, M., and O'Brien, D. Z. (2017). Pitch perfect: Vocal pitch and the emotional intensity of congressional speech. *American Political Science Review*, pages 1–22.

Druckman, J. (2003). The power of television images: The first kennedy nixon debate revisited. *Journal of Politics*, 65(2):559–571.

Efrain, M. and Patterson, E. (1974). Voters vote beautiful: the effect of physical appearance on a national election. canadian journal of behavioural science/revue canadienne des sciences du.

Eibl-Eibesfeldt, I. (1979). *Ritual and ritualization from a biological perspective. In Human ethology: Claims and limits of a new discipline*. University Press.

Ekman, P. and Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*. Ishk.

Fein, S., Goethals, G. R., and Kugler, M. B. (2007). Social influence on political judgments: The case of presidential debates. *Political Psychology*, 28(2):165–192.

Fridkin, K. L., Gershon, S. A., Courey, J., and LaPlant, K. (2019). Gender differences in emotional reactions to the first 2016 presidential debate. *Political Behavior*, pages 1–31.

Friedenberg, R. (1997). Patterns and trends in national political debates: 1960-1992. rhetorical studies of. In *National Political Debates—1996, Wesport CT, Praeger*, page 61–90.

Hass, H. (1970). *The human animal*. Delta, New York.

Hellweg, S. A., Pfau, M., and Brydon, S. R. (1992). *Televised presidential debates: Advocacy in contemporary America*. Praeger Publishers.

Hitchon, J. C., Chang, C., and Harris, R. (1997). Should women emote? perceptual bias and opinion change in response to political ads for candidates of different genders. *Political Communication*, 14(1):49–69.

Hughes, S. and Bucy, E. (2016). Moments of partisan divergence in presidential debates: Indicators of verbal and nonverbal influence. In Schill, D., Kirk, R., and Jasperson, A., editors, *Political communication in real time: Theoretical and applied research approaches*, page 249–274. Routledge, New York, NY.

Jarman, J. (2005). Political affiliation and presidential debates: A real-time analysis of the effect of the arguments used in the presidential debates. *American Behavioral Scientist*, 49(2):229–242.

Joo, J., Bucy, E. P., and Seidel, C. (2019). Computational communication science— automated coding of televised leader displays: Detecting nonverbal political behavior with computer vision and deep learning. *International Journal of Communication*, 13:23.

Joo, J. and Steinert-Threlkeld, Z. (2018). Image as data: Automated visual content analysis for political science. arxiv p.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, 58(9):697.

Kahneman, D. (2011). Thinking, fast and slow.

Kirk, R. and Schill, D. (2014). Cnn's dial testing of the presidential debates. In *Techno Politics in Presidential Campaigning: New Voices, New Technologies, and New Voters, 74.*

Kraus, S. (1996). Winners of the first 1960 televised presidential debate between kennedy and nixon. *Journal of Communication*, 46(4):78–96.

Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems(pp*, page 1097–1105.

Lanoue, D. and Schrott, P. (1989). *Voters' reactions to televised presidential debates: Measurement of the source and magnitude of opinion change.* Political Psychology.

Lanzetta, J., Sullivan, D., Masters, R., and McHugo, G. (1985). Emotional and cognitive responses to televised images of political leaders. mass media and political thought: An information- processing approach.

Lanzetta, J. T. and Orr, S. P. (1980). Influence of facial expressions on the classical conditioning of fear. *Journal of Personality and Social Psychology*, 39(6):1081.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, volume 86, page 2278–2324.

Little, A., Burriss, R., Jones, B., and Roberts, S. (2007). Facial appearance affects voting decisions. *Evolution and Human Behavior*, 28(1):18–27.

Little, A., Roberts, S., Jones, B., and DeBruine, L. (2012). The perception of attractiveness and trustworthiness in male faces affects hypothetical voting decisions differently in wartime and peacetime scenarios. *The Quarterly Journal of Experimental Psychology*, 65(10):2018–2032.

Lorenz, K. (1966). *On aggression.* Routledge.

Masters, R. (1989). *The nature of politics.* Yale University Press.

Masters, R., Sullivan, D., Feola, A., and McHugo, G. (1987a). Television coverage of candidates' display behavior during the 1984 democratic primaries in the united states. *International Political Science Review*, 8(2):121–130.

Masters, R. D., Sullivan, D. G., Feola, A., and McHugo, G. J. (1987b). Television coverage of candidates' display behavior during the 1984 democratic primaries in the united states. *International Political Science Review*, 8(2):121–130.

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

McArthur, L. and Baron, R. (1983). Toward an ecological theory of social perception. *Psychological review*, 90(3):215.

McNeil, D. (1970). The acquisition of language: The study of developmental psycholinguistics.

Mehrabian, A. and Wiener, M. (1967). Decoding of inconsistent communications. journal of personality and social. *psychology*, 6(1):109.

Messaris, P. and Abraham, L. (2001). The role of images in framing news stories. In *Framing public life*, pages 231–242. Routledge.

Montepare, J. and Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal behavior*, 27(4):237–254.

Nagel, F., Maurer, M., and 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(5):833–850.

Olivola, C. and Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of nonverbal behavior*, 34(2):83–110.

Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001.

Reinemann, C. and Maurer, M. (2005). Unifying or polarizing? short-term effects and postdebate consequences of different rhetorical strategies in televised debates. *Journal of Communication*, 55(4):775–794.

Richell, R., Mitchell, D., Peschardt, K., Winston, J., Leonard, A., and Dolan, R. (2005). Trust and distrust: The perception of trustworthiness of faces in psychopathic and non- psychopathic offenders. *Personality and Individual Differences*, 38(8):1735–1744.

Salter, F. (1995). *Emotions in command: A naturalistic study of institutional dominance.* Oxford University Press.

Schill, D. and Kirk, R. (2014). Courting the swing voter: "real time" insights into the 2008 and 2012 us presidential debates. *American Behavioral Scientist*, 58(4):536–555.

Schill, D., Kirk, R., and Jasperson, A. E. (2016). *Political Communication in Real Time: Theoretical and Applied Research Approaches.* Taylor & Francis.

Schmidt, K. and Cohn, J. (2001). Human facial expressions as adaptations: Evolutionary questions in facial expression research. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, 116(33):3–24.

Senior, C., Phillips, M., Barnes, J., and David, A. (1999). An investigation into the perception of dominance from schematic faces: A study using the world-wide. *Web. Behavior Research Methods, Instruments, & Computers*, 31(2):341–346.

Somit, A. and Peterson, S. (1997). *Darwinism, dominance, and democracy: The biological bases of authoritarianism.* Praeger Publishers.

Stewart, P., Salter, F., and Mehu, M. (2011). *The face as a focus of political communication: Evolutionary perspectives and the ethological methods. The Sourcebook for Political Communication Research.* Routledge, New York.

Stewart, P. A., Eubanks, A. D., Dye, R. G., Gong, Z. H., Bucy, E. P., Wicks, R. H., and Eidelman, S. (2018). Candidate performance and observable audience response: Laughter and applause–cheering during the first 2016 clinton–trump presidential debate. *Frontiers in psychology*, 9.

Sullivan, D. and Masters, R. (1988). Happy warriors": Leaders' facial displays, viewers. *Emotions, and Political Support. American Journal of Political Science*, page 345–368.

Tiedens, L. Z. (2001). Anger and advancement versus sadness and subjugation: the effect of negative emotion expressions on social status conferral. *Journal of personality and social psychology*, 80(1):86.

Todorov, A., Mandisodza, A., Goren, A., and Hall, C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308(5728):1623–1626.

Todorov, A., Olivola, C., Dotsch, R., and Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual review of psychology*, 66:519–545.

Todorov, A., Said, C., Engell, A., and Oosterhof, N. (2008). Understanding evaluation of faces on social dimensions. In *Trends in cognitive sciences*, volume 12, page 455–460.

Torres, M. (2018). Give me the full picture: Using computer vision to understand visual frames and political communication. *URL: http://qssi. psu. edu/new-faces-papers-2018/torres-computer-vision-and-politicalcommunication.*

Trichas, S. and Schyns, B. (2012). The face of leadership: Perceiving leaders from facial expression. *The Leadership Quarterly*, 23(3):545–566.

Trichas, S., Schyns, B., Lord, R., and Hall, R. (2017). "facing" leaders: Facial expression and leadership perception. *The Leadership Quarterly*, 28(2):317–333.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131.

Van Hooff, J. (1967). The facial displays of the catarrhine monkeys and apes.

Vancil, D. L. and Pendell, S. D. (1987). The myth of viewer-listener disagreement in the first kennedy-nixon debate.

Way, B. M. and Masters, R. D. (1996). Political attitudes: Interactions of cognition and affect. *Motivation and Emotion*, 20(3):205–236.

Zhang, H. and Pan, J. (2019). Casm: A deep-learning approach for identifying collective action events with text and image data from social media. *Sociological Methodology*, 49(1):1–57.

Zhu, J., Milavsky, J., and Biswas, R. (1994). Do televised debates affect image perception more than issue knowledge? a study of the first. *Human Communication Research*, 20(3):302–333.