Amsterdam University Press

COMPUTATIONAL COMMUNICATION RESEARCH 4.1 (2022) 254-274
HTTPS://DOI.ORG/10.5117/CCR2022.1.007.NEUM

ARTICLE

Body Language and Gender Stereotypes in Campaign Video

Markus Neumann
Wesleyan University

Erika Franklin Fowler
Wesleyan University

Travis N. Ridout
Washington State University

Abstract
We examine the impact of candidates' gender on the body language that they employ in their political advertisements. Using data on over 1,600 candidates appearing in almost 5,400 political ads that aired in the U.S. between 2017 and 2020, we employ automatic pose detection to trace the movement of their hands. We find, consistent with gender stereotypes, that male candidates use more assertive hand movements than female candidates. We also find evidence of more assertiveness among Democratic candidates and among candidates running for U.S. House, U.S. Senate, and governor.

Introduction

Gender matters in politics in a variety of ways. Gender matters not only for how voters view candidates but also for how candidates present themselves to voters. Research shows, for example, that men and women candidates are perceived to have different traits and levels of competence (Eagly & Karau, 2002; Schneider & Bos, 2014). The differences may occur, in part, because voters bring with them stereotypes about how men and women candidates ought to behave – stereotypes that may be (intentionally or unintentionally) activated by the campaign itself (Bauer, 2015; Cassese & Holman, 2018; Holman, Schneider, & Pondel, 2015). In this analysis, we focus on the relationship between a candidate's gender and the body language the candidate employs, specifically, the extent to which a candidate's body language is calm and restrained, on the one hand, or energetic and assertive, on the other hand. This is measured through the amount of vertical wrist movement, which serves as a proxy for up and down hand gestures, which voters associate with power and dominance (Everitt, Best, & Gaudet, 2016). We ask: Are stereotypes with respect to how men and women are expected to behave reflected in the body language of candidates in their political advertising? And does that candidate behavior depend on the characteristics of the candidate, including the party, the office sought, and incumbency?

To answer these questions, we employ an innovative approach, using automated pose detection to evaluate the body language of candidates that appear in televised political advertisements collected by the Wesleyan Media Project. By tracking the movement of candidates' wrists, we can evaluate the extent to which each candidate's gestures display assertiveness or power (as measured through their vertical wrist movement), which is typically seen as a masculine trait. This study goes beyond past research in several ways. Most importantly, we expand on the number of politicians analyzed, examining 1,658 different candidates who appear in 5,388 different ads across two election cycles. Because of this large sample size, we can examine not only gender as a correlate of body language but the impact of other characteristics, including incumbency status and the office sought. We examine candidates running for offices up and down the ballot, ranging from state legislature to U.S. Senate and governor.[1]

Our research, in addition to its contribution to the literature on gender in advertising, also illustrates how computational methods, here automatic pose detection, could be used to analyze politicians' gestures in political debates, interviews or speeches in addition to examining non-elite persons in political settings, such as actors in campaign ads or citizens attending rallies and protests. Ultimately, we find that, consistent with gender stereotypes, the body language of male candidates is generally more energetic and assertive, though the size of the effect is fairly small. Moreover, not only does body language vary by candidate gender, but the candidate's party and the office sought also influence how candidates employ body language in their political ads, with Democrats and candidates for governor, U.S. House and U.S. Senate exhibiting more assertive behavior.

Gender Stereotypes

A variety of research has affirmed that citizens hold stereotypes of men and women politicians. For instance, voters tend to ascribe greater overall competence, strength and leadership skills to men (Eagly & Karau, 2002; Hayes, 2005), but greater warmth and compassion to women (Hayes, 2005; Huddy & Terkildsen, 1993b; Koch, 1999). To some extent, gendered traits are also ascribed to candidates of particular parties, with Republicans more likely to be associated with masculine traits and Democrats with more feminine traits (Hayes, 2011; Winter, 2010, but see Schneider & Bos 2016) even though there is variation in how women from each party are evaluated (Dolan, 2018; Sanbonmatsu & Dolan, 2009).

Ultimately, the importance of these stereotypes is twofold. Not only might they influence people's evaluations of candidates and their vote choices, but stereotypes also may affect candidates' strategic behavior, playing to or attempting to counter gendered stereotypes. While some argue that playing to gendered stereotypes can benefit women candidates as traits such as caring and willingness to compromise may be viewed as strengths (Herrnson, Lay, & Stokes, 2003), others argue that the association of assertiveness and leadership with masculine stereotypes make it difficult for women to succeed in electoral politics (Huddy & Terkildsen, 1993a) because of a double-bind. Namely, if women work to counter gendered stereotypes, they may be penalized for not displaying the feminine traits expected of women, and if they play up their gender, they may be less likely to be seen as a leader (Carpinella & Bauer, 2021; Jamieson, 1995; Schneider & Bos, 2014). One study found, for example, that voters punished German Chancellor Angela Merkel for displays of anger (Boussalis, Coan, Holman, & Müller, 2021), yet there is also evidence that voters react similarly to gender stereotypical emotional displays (crying and anger) from candidates regardless of the candidate's gender (Brooks, 2011, 2013).

Politicians' Body Language

In addition to a wealth of research examining gendered stereotypes and candidate strategy with respect to campaign content and issue focus, there is also literature in non-verbal communication on how a candidate's gender might influence body language and how body language can convey gendered messages about candidates. For example, research on the 2016 election suggests that Hillary Clinton showed more "mixed" body language cues (suggestive of behavior to counter gender stereotypes) while Donald Trump showed more hostile body language consistent with male stereotypes (Wasike, 2019). Body language may also convey social information, including reliable cues about whether the person is a man or woman (Pollick, Kay, Heim, & Stringer, 2005). Indeed, viewers' assumptions about the gendered nature of body language even leads them to ascribe different traits to non-gendered stick figures whose movement is modeled after that of men and women (Koppensteiner & Grammer, 2011). The expansiveness of body language movements, including the range over which candidates move their hands vertically, has been associated with perceptions of dominance and such expansive body language is most prominent among men politicians challenging the status quo (Koppensteiner, Stephan, & Jäschke, 2016).

Expectations

While the vast literature on gender in politics suggests mixed findings about whether women should play to or counter gendered stereotypes in their messaging, we have several reasons why we believe that body language may be more likely to conform to traditional gender stereotypes, with men displaying more assertive and energetic hand gestures than women. First, existing findings on hand movements suggest that gender conforming movements may be most beneficial for both men and women (Everitt et al., 2016). Second, although much campaign behavior—and many elements of political advertising—is strategically crafted, we believe that body language is lower down on the list of things that candidates may alter when they do try to push against gendered stereotypes. More specifically, the choice of issue topic, setting, language, and clothing in an ad seem easier to change than training oneself to alter one's body language. Therefore, we hypothesize that male candidates will demonstrate more energetic hand gestures than will female candidates.

Having said that, we also acknowledge that the effect of gender can also vary by party, by office and by incumbency status. Thus, we also explore the relationship between these other characteristics and candidates' body language to ensure that we are not conflating the effect of gender with these other characteristics. First, a candidate's party may have a strong influence not only on people's expectations and stereotypes of a candidate, but it may influence a candidate's use of body language as well. Voters associate Democrats with more feminine traits and associate Republicans with more masculine traits (Hayes, 2005; Winter, 2010). If candidates embrace these partisan stereotypes, then Republicans might use more assertive gestures than Democrats.

Second, the office held by the politician could influence the use of body language. More specifically, the degree of control that politicians exert over the presentation of themselves likely depends on the degree to which they have made politics their career. This, in turn, is directly related to the level of legislative professionalism of the institution in which they serve (Berry, Berkman, & Schneiderman, 2000; Carsey, Winburn, & Berry, 2017; Squire, 1992). Higher offices provide office-holders with greater resources and public visibility, which provides them with more control over their image in the eye of the public, and this is especially true in political advertising, which is expensive to produce. Politicians who have "bought into the system" and appear higher on the ballot might act more in accordance with its norms. Hence, we conjecture that Senators, governors and members of Congress might show more assertive body language than politicians in state legislatures and other down-ballot offices.

A third factor that might influence the use of body gestures is the candidate's status as an incumbent or challenger. Assertive gestures, for instance, may be a favored tool of politicians who aim to challenge the status quo (Bucy & Grabe, 2007; Everitt et al., 2016; Koppensteiner et al., 2016). Translated to the political realm as a whole, the implication is that challengers should be more likely to use energetic body language than incumbents.

Methods & Data

Face Detection & Face Recognition

In order to track a candidate's body language, we first need to identify the candidates in their videos. Given that most campaign videos feature many people other than the candidate, such as family members and constituents giving "testimonials," this is not a trivial task. Doing so by hand would be an excessively labor-intensive process: For example, a typical 30-second video shot at 24 frames per second (fps) consists of 30*24=720 images. For a dataset with thousands of videos, this would require the hand-coding of millions of images, which is simply not feasible.

Instead, we rely on machine learning to 1) detect all faces in each image and 2) classify whether any one face in each image belongs to the candidate or not. Given that face detection and recognition are among the staple methods of computer vision, with many well-tested and widely-used implementations, we use pre-trained models for this purpose. Given that the faces in our dataset and the faces these models were trained on are human and can thus be assumed to originate from the same data-generating process, we can assume that these models generalize to our data (and in fact, the model we use was trained on public figures, including politicians (Cao, Shen, Xie, Parkhi, & Zisserman, 2018)).

To this end, we use the MTCNN architecture (Zhang, Zhang, Li, & Qiao, 2016) for face detection and the FaceNet architecture (Schroff, Kalenichenko, & Philbin, 2015) for face recognition, both implemented in PyTorch by Esler

(2019). Using the MTCNN, we process each image of every video and apply bounding boxes to all detected faces. These faces are then cropped out and normalized, in preparation for the face recognition.

While detecting *any* humans in an image can be done with a model trained on different people, a face recognition model requires training data for the specific faces we want to be able to recognize. Given that machine learning methods, and in particular neural networks, generally work best with tens, if not hundreds or thousands of training samples per class, this would require an inordinate amount of hand-labeling in order to identify 1,658 politicians. Consequently, we do not rely on a traditional supervised approach but rather on a model that creates embeddings for faces. This allows their comparison through the use of distance scores. The FaceNet architecture is a model that has, effectively, been trained to recognize a set of faces in a training dataset, and in the process, it has learned how to tell humans apart from one another. The embeddings it creates for each face can then be compared to other embeddings, where faces of the same person will yield a lower distance score.[2] We classify two faces as a match when their distance score is lower than 1. It can then be used to recognize a different set of faces with only a single training image per class.

To this end, we create a dataset of reference images by scraping the headshots of politicians from their Ballotpedia pages. Given that the large majority of national and even state-level politicians have a Ballotpedia page with a photo, this, in combination with the embedding-based approach to face recognition and the known sponsorship information from each ad, allows us to computationally recognize the faces of these candidates without any hand-labeling. There are also computational benefits to this approach, as it does not require a large model with the ability to recognize every person of interest in the dataset. For example, when trying to find the face of Alaska gubernatorial candidate Mark Begich in his videos, the model simply compares every face in them to Begich's Ballotpedia headshot, and it has no knowledge of, say, the face of Begich's opponent, Mike Dunleavy, even though they are both in our dataset. We recommend this approach to face recognition to other computer vision practitioners in political science, as single labeled photos for politicians of interest are readily available in various online databases, and it requires no human coding for supervised labels in large training datasets. (Of course, this method works best for candidate-sponsored ads).

The performance of this face recognition method has been validated by the authors of the method themselves (Schroff et al., 2015). They report an error rate of under 1% on the Labeled Faces in the Wild dataset, which also includes a number of politicians. Since the videos we use have already been coded by the Wesleyan Media Project for whether the candidate was pictured (albeit not for every frame - only for the video as a whole), we use this as an additional form of validation for how well the face recognition works on our particular dataset. There are 5,524 videos the WMP has coded in this way to which we have applied our face recognition model. Among these, there are 57 (about 1%) for which we do not detect the candidate in any of the ad frames. This is a low error rate, albeit with the caveat that it is a fairly easy test. Therefore we conduct another, more difficult test. The WMP codes for whether a candidate is pictured in any part of the ad *except* the oral approval, since federal candidates are bound by law to appear here. If a candidate is coded as not pictured in the ad, this means that they only appear in the oral approval, which is a very short time span. Therefore, the face recognition only has a few seconds, rather than potentially the full 30-60 seconds of the ad, to detect the candidate at least once. In this more difficult test, the error rate is higher but still only 7%. That being said, this is still only an approximation, compared to a validation on a (hypothetical) fully labeled video dataset. Finally, any error in the face recognition will either a) result in the video being removed from our dataset entirely if the candidate is not detected for at least 24 contiguous frames, meaning that it won't affect the results, or b) be reflected in the error rate of the final measure, which we test for explicitly below. We consider that measure to be the main validation for this paper's pipeline.

Pose Detection

After identifying which (if any) of the faces in a frame belong to the relevant candidate, we measure their body language by tracking the position of a set of landmarks across their body. We accomplish this using the pose detection framework OpenPose (Cao, Hidalgo Martinez, Simon, Wei, & Sheikh, 2019; Simon, Joo, Matthews, & Sheikh, 2017). This model consists of a convolutional neural network that first detects all body parts in an image, and then associates each with a specific person. This process produces 25 landmarks for every person, denoting the position of the nose, eyes, ears, throat, shoulders, elbows, wrists, pelvis, hips, knees, ankles, heels, big toes and small toes.[3] Of these, the position of the wrists and nose are of particular interest in this research. We rely on a model pre-trained on the COCO keypoint challenge dataset (Lin et al., 2014), which Cao et al. (2019) modified with additional foot keypoint annotations.

Figure 1 shows an application of the method to two images in our dataset from a campaign video of Jane Dittmar, the Democratic candidate for Virginia's 5th Congressional district in 2016. The lines connect various

Figure 1 Results of pose estimation to two frames of a campaign video of Jane Dittmar, the Democratic candidate for Virginia's 5th Congressional district in 2016.

positions on the body, including the wrists, which we can use to track vertical hand movement.

Finally, we combine the results of pose detection and face recognition. We do so by checking whether the bounding box for the identified candidate, produced by the face recognition, overlaps with any face detected by the pose detection. Specifically, we check whether any face recognition bounding boxes contain nose landmarks from the pose detection. This is done for every frame in the video, yielding the body landmarks for only the candidate and no one else.

On occasion, this pipeline yields false negatives, in that the candidate's face is only identified in some frames, but not those in between. This is because a) at a resolution of 480x320, the videos are fairly low quality and b) videos are harder to classify than still images because the figures in them are moving, which causes some frames to be of worse quality than others. To counteract this problem, we make use of the continuous motion of videos. We iterate through the frames of a video, and if a frame does not contain the picture of the candidate, we check whether the frame before or after does. If so, we check whether there is a person whose landmarks are within some small margin of error[4] from the candidate's pose in the frame in which they were successfully identified. If so, we conclude that this is the same person and therefore also classify it as the candidate. We continue this process iteratively until no more additional instances of the candidate can be identified. To ensure that our movement measures (described below) are based on a sufficient sample, we only include videos in which candidates and their hands are detected for at least one second (i.e., 24 consecutive frames).

The face detection, face recognition, and pose detection were all computed using Google Colab, on a variety of GPUs (Tesla K80, P4, P100 and V100).[5,6]

Measuring powerful behavior

Pose estimation provides us with a set of keypoints across the subject's body at any one point in time. Given that most gesticulation occurs through the hands, and that the theory of Everitt et al. (2016) on gestures signalling power also revolves around them, we construct a measure for this concept by tracking the position of the wrists. Given that Everitt et al. (2016) postulate that powerful behavior is associated specifically with vertical movement, we focus on hand movement on the y-axis. The model outputs a set of two-dimensional coordinates for the wrist (and other) keypoints, and we measure how much they move over the course of a video.

First, we group together contiguous sections of keypoints. This means we measure different scenes in the ad separately. We only keep any one section if the candidate's hands are shown for at least 24 contiguous frames (i.e., 1 second). Because many videos either don't show the candidate's hands at all, or only very briefly,[7] our original dataset of over 7,000 videos gets reduced to the 5,388 we analyze here.[8]

We measure the candidate's use of vertical movement through the space they cross with their hands. For a video section, we compute the 2D kernel density of all the keypoints of the wrist. For candidates who hardly move their hands, the kernel density will be very concentrated in one area, meaning that the overall space will be small. Conversely, when candidates gesticulate more actively, their wrists will cross a larger space, so the kernel density will be more spread out. In order to remove outliers and capture the space that the candidate's wrists traverse most frequently rather than spurious movement, we discard the parts of the kernel density that are less than 25% of its overall maximum value. Then we take the outline of the resultant shape and measure its height at continuous points from left to right. We then average these measures, which gives us the average height of the space the candidate crosses with their hands. This process is illustrated in Figure 2. We repeat this procedure for every section of video in which the candidate appears, and for both wrists.

To aggregate our measures across the video, we multiply each section's measure by the section's length (in frames), sum across the sections, and then divide by the total length. This way, we weight measures from longer sections more heavily, given that they both have a greater effect on the viewer, and are more reliable to measure.[9] The measures from the hands are then summed together, so that a candidate who gesticulates with both hands rather than just one receives a higher score. This is the final movement score for the video, and constitutes the dependent variable in the models described below. It ranges from 0 to 79 (higher values correspond to greater movement), with mean 10 and standard deviation 8.

Figure 2 The figure illustrates how the candidate's amount of movement is measured. Panel (a): The orange points show the position of Steyer's right wrist over the course of about 5 seconds. Then, a 2D kernel density is applied to these points. We remove outliers by omitting values below 25% of the density's maximum value. The blue shaded area shows the result of this. Panel (b): Finally, we measure the height of this 'cloud' by calculating the distance between the highest and lowest point for each column of points in it. The average of these distances is the measure for the candidate's amount of right-hand movement during the scene. The sum of the right and left hand is the final movement score, used as our dependent variable.

To ensure that our measure conforms to human perception, we sampled a validation set of 127 [10] videos and hand coded them for the amount of vertical movement by the candidate on a scale of 1 to 5, with 5 being the highest. The correlation between the human coded data and our measure is 0.52. Since the movement measure sits at the end of a chain of procedures – face detection, face recognition and pose detection – any error in these procedures will also be reflected in the error of the measure. Considering that, a correlation of 0.52 is quite high, which we take to be a strong signal that our pipeline works.

Dataset

Our dataset consists of 5,388 videos of political ads from the Wesleyan Media Project's archive from the 2017-2018 and 2019-2020 election cycles. The videos were originally captured by Kantar/CMAG, a commercial firm that monitors ads placed on national cable, national broadcast and local broadcast television in all 210 media markets in the United States. Kantar/CMAG provides meta-data about each ad, such as who paid for the ad, where it aired and an estimate of the cost, and the Wesleyan Media Project's staff does additional coding on several attributes of each ad, including the favored and targeted candidate, ad tone and the issues mentioned.

These videos all have a resolution of 480x320 and consist of a total of 4,199,000 frames, making up over 46 hours of footage. The ads were sponsored by 1,658 candidates (counting the same person separately when running in different races, but the same when running for the same office in different years) who ran at both the federal and state levels in the 2018 and 2020 election cycles (including a few elections that took place in 2017 and 2019). Ads sponsored by both general election and primary election candidates are included in our data. Party-sponsored and group-sponsored ads are not included since we are focused on self-presentation. We also excluded candidate-sponsored attack ads because, by definition, they focus on an opponent, not the sponsoring candidate.[11]

The gender of all politicians in the dataset was hand-coded by Wesleyan Media Project staff, with women coded 1 and men coded 0.[12] There are 1,147 men and 511 women in the dataset. We also coded the party of each candidate, with Republicans coded 1 and Democrats coded 0. Given that partisanship is a major predictor of political behavior, we test whether it also influences body language. Candidates not belonging to either the Democratic or Republican party were omitted. We were left with 832 Democrats and 826 Republicans.

We also take into account the type of office for which the candidate is running. Here, we combined all state-level offices below the governorship (State Representative, State Senate, Delegate, Assembly, State Supreme Court Judge, and Attorney General) into one "down ballot" category. Candidates for governorships, the U.S. House of Representatives, the U.S. Senate and the presidency each constitute separate categories, with down-ballot races being the reference category. See Table 1 for the number of ads by office that we examined in each of the two election cycles.

We also code for whether a candidate is an incumbent, challenger or running for an open seat. This variable was produced from data compiled by OpenSecrets, formerly known as the Center for Responsive Politics (CRP)[13] and from joint work by the Center for American Women and Politics (CAWP)[14] and the National Institute of Money in Politics (NIMP).[15] The reference category is challenger.

Finally, our models include a control for the number of a video's frames in which the candidate's hands are visible (requiring that their face also be visible and recognized). A large majority of the ads in our dataset are 30 seconds long, but they range from 5 to 120 seconds. There is a lot of variation in whether the candidate is shown for only a second or two, or throughout a large portion of the video. Therefore, we account for the combined length of time for which the candidate and their hands are shown. The variable is referred to as "Control: Candidate frames" in the regression tables.

Table 1 Number of ads by office and election cycle, with all state-level offices below governor combined into 'Down-ballot'.

	2017-2018	2019-2020
Down-ballot	572	422
Governor	688	207
House	1093	1119
President	0	417
Senate	339	531

Results

We start by examining the distribution of vertical hand movement for men and women candidates. For women, the range is from 0.01 to 48.93, with a median of 8.35 and a standard deviation of 6.59. For men, the range is larger, 0.01 to 51.35, with a slightly larger median of 8.90 and a standard deviation of 6.55. The statistics suggest more vertical hand movement among men candidates than among women candidates, though the difference in medians is not particularly large. (The difference between the respective means is similarly small—9.55 for women and 9.96 for men.)

To get a truer picture of whether these differences between men and women are statistically significant, we regress the measure of the candidate's vertical hand movement throughout a video on the covariates described above. Figure 3 (see Table 2 in the Appendix for the full regression results) shows the estimates from a set of models, with the stepwise addition of the variables. To control for potential within-unit effects, we rely on a random effects model, clustered by candidate.[16,17,18]

In what is essentially a bivariate model of movement and gender (albeit still controlling for the number of frames featuring the candidate's hands), we find that there is a negative correlation between female candidates and the amount of vertical hand movement. In other words, women candidates demonstrate less vertical hand movement at $p < .05$, which is consistent with the notion of gender-stereotypical behavior.

When party is added, we observe a negative and statistically significant coefficient estimate for Republicans. That is, Republican candidates exhibit less vertical hand movement than Democratic candidates. This finding is contrary to our expectation that Republicans would exhibit more assertive behavior, an expectation rooted in the stereotypes of politicians of both parties. This finding, then, suggests that instead of playing to stereotypes, Republican and Democratic politicians may be trying to counteract people's

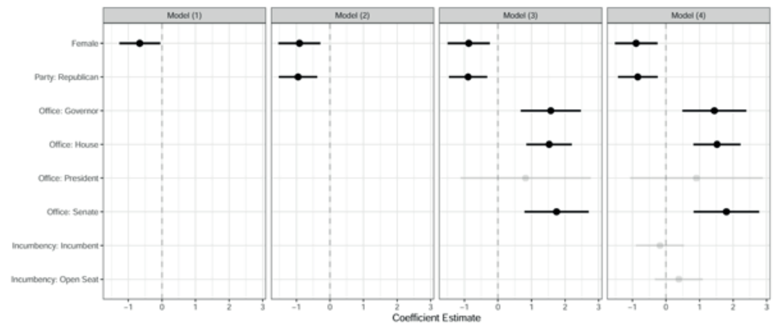

Figure 3 Vertical hand movement regressed on candidate-level covariates, with candidate-level random effects. Effects that are not statistically significant at the 5% level are shown with lower opacity. Full regression results in Table 2.

expectations of how they behave—and the subsequent inferences that voters make about their traits and policy positions. Importantly, though, the impact of gender remains in Model 2, with women candidates demonstrating less vertical hand movement at $p < .05$.

When we turn to the office level, we had expected to find that candidates for higher office, such as the presidency, U.S. Senate and the U.S. House, would exhibit greater vertical hand movement—our indicator of assertiveness—than candidates for state legislative positions. The estimates in Model 3 provide some qualified support for that expectation. All of the coefficients on the indicators of office type are positive, suggesting greater vertical hand movement among candidates for president, U.S. Senate, governor and U.S. House than among candidates in down-ballot races (the omitted category). But only the governor, U.S. Senate, and U.S. House estimates are statistically significant at $p < .05$ in all of the models. While the coefficient is positive for the presidential race indicator in all three models, there is a high degree of uncertainty given the relatively small number of presidential ads in our sample. In short, there is suggestive, but far from definitive, evidence that candidates running for higher-level offices are more assertive than candidates running for lower-level offices. The effect of gender on candidate hand movement remains strong, with women exhibiting less vertical hand movement $p < .05$.

With respect to incumbency, we had expected to find that non-incumbents would use more vertical hand movement than incumbents, but we find no statistically significant effects when we add these variables into Model 4. Incumbents do not appear to differ from challengers and candidates running for an open seat in their propensity to use assertive gestures. Again, the impact of gender remains at $p < .05$.

In Model 5 (see Table 2 in the Appendix), we explore whether the impact of the independent variables might be different depending on whether the candidate is male or female. To do this, we include several additional variables in the model, each an interaction of a female indicator with the other covariates. We find, however, that none of the coefficients on the interaction variables are significant predictors of the use of vertical hand movements. This suggests that while, overall, men are more likely to use such assertive gestures, that finding does not depend on being a Republican or Democrat, running for a particular office, or running as an incumbent.

Discussion

Our analysis of body language – specifically, energetic hand gestures – among politicians who appear in political advertisements reveals that such movement does not happen at random. Rather, it is associated with the characteristics of the political candidates who appear in those ads. Specifically, we found an association between the gender of the candidate and the use of vertical hand movements such that women candidates exhibited less wide-ranging body language than men candidates. Although the effect sizes were small, these findings point to another subtle way by which gender influences candidate behavior.

In addition to an impact of candidate gender on body language, we also found an effect of party. Republican candidates exhibit less vertical hand movement (again, our indicator of assertiveness) than Democratic candidates, a finding that contradicts citizens' stereotypes of the parties. The type of office the candidate was running for also appeared to influence vertical hand movement. Candidates running for governor, U.S. House and U.S. Senate, in particular, exhibited greater vertical hand movement than candidates in down-ballot races. In sum, we have suggestive evidence that candidates running for higher offices are exhibiting more assertive behavior than candidates running for lower offices.

One potential objection to our main finding is that women, who are on average shorter than men, may have shorter arms, and as such are more constrained in the extent to which they can move their hands. While perhaps true, most hand movement that we observed in political ads did not use the full range of the arm's reach. Moreover, the portrayal of body language shown in political ads is something that is, to some degree, strategic. And even if candidates are not consciously thinking about how expansive they should make their hand movements when they are being filmed for a political ad, ad makers and campaign consultants can exert some control over the message conveyed. Namely, if ad makers want to showcase a more or less energetic candidate, they can do so through their choice of camera angles and fields of vision. That is, the camera person, the ad's director or the ad's editor can make a short candidate look very assertive (or not) through the choices they make in filming and editing.

Why does all this matter? Most importantly, the body language that candidates use in their self-presentations could have small, but real, impacts on how voters perceive them. According to Mehrabian (1972), the impact of a speech is 7% content, 38% tone of voice and 55% body language. And yet the majority of quantitative research on how political campaigns communicate has focused on that 7%, largely ignoring nonverbal communication. A candidate's body language may not be a more important influence than, say, partisanship in how citizens perceive a politician, but a host of research has shown the impact of non-verbal communication, including body language.

That men and woman present themselves differently in their campaign ads – one of the central findings of this research – is, on the one hand, not surprising. But it also highlights that differences in how women and men campaign have not disappeared, in spite of recent research that has portrayed those differences as declining (Sapiro, Cramer Walsh, Strach, & Hennings, 2011). Our research also raises the question of whether those less conspicuous differences in candidate presentation across gender are all intentional or not. If a candidate decides to mention the issue of abortion in a political ad, we can safely assume that decision was intentional and perhaps even based on survey evidence and focus group analysis. But whether a candidate displays more assertive or restrained gestures in a political ad has many potential explanations. First, it may be intentional on the part of the candidate, that is, the candidate wants to appear more or less energetic. Second, it could be unintentional, stemming from the personality of the individual and what the person has learned by being in a particular political context and his or her lived experience of being male or female in society. Third, it may be strategic on the part of ad makers, who film or edit an ad in a particular way so as to highlight particular gestures.

Our research leaves open some avenues for future research. For one, future work might employ the techniques used here cross-nationally to determine how universal gendered patterns in the use of hand gestures are and whether they vary with the political system or degree of sexism in a society. Moreover, the methods developed here might be useful for social scientists in analyzing gestures of other people featured in political ads;

politicians in political debates, interviews, speeches or rallies; and in settings beyond politics to analyze the movements of people featured in videos. Our research could also be expanded to examine how candidates' presentations of themselves vary depending on the gender of their opponents—and, in the case of attack ads, how campaigns choose to present their opponents. In short, is body language different when candidates appear by themselves in positive ads versus in negative ads in which they use video of an opponent?

In the end, we find that body language, including something as simple as a hand gesture, depends on a candidate's gender. These gestures may send subtle cues to voters about the candidate's traits, potentially influencing voters' evaluations of candidates and their choices at the ballot box.

Appendix

Table 2 Movement regressed on candidate level covariates, with candidate level random effects. The table shows that female politicians move less, as do Republicans as well as federal and gubernatorial candidates. Interaction effects between covariates and gender are not statistically significant.

	Dependent variable:				
	Vertical hand movement				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Female	−0.659 **	−0.909 ***	−0.870 ***	−0.889 ***	−1.159
	(0.311)	(0.319)	(0.319)	(0.323)	(0.829)
Party: Republican		−0.949 ***	−0.889 ***	−0.839 ***	−1.094 ***
		(0.290)	(0.291)	(0.300)	(0.353)
Office: Governor			1.573 ***	1.442 ***	1.652 ***
			(0.457)	(0.485)	(0.557)
Office: House			1.524 ***	1.520 ***	1.510 ***
			(0.346)	(0.357)	(0.434)
Office: President			0.824	0.906	1.197
			(0.987)	(1.007)	(1.160)
Office: Senate			1.745 ***	1.800 ***	1.969 ***
			(0.487)	(0.497)	(0.596)
Incumbent				−0.178	−0.100
				(0.363)	(0.447)
Open Seat				0.383	0.303
				(0.363)	(0.455)
Control: Candidate frames	0.002 ***	0.002 ***	0.002 ***	0.002 ***	0.002 ***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)

Female* Party: Republican					0.943 (0.675)
Female* Office: Governor					−0.918 (1.173)
Female* Office: House					0.080 (0.767)
Female* Office: President					−1.342 (2.394)
Female* Office: Senate					−0.518 (1.110)
Female* Incumbent					−0.155 (0.787)
Female* Open Seat					0.293 (0.762)
Constant	9.294 *** (0.223)	9.820 *** (0.274)	8.623 *** (0.373)	8.563 *** (0.455)	8.646 *** (0.533)
Observations	5,388	5,388	5,388	5,208	5,208
R^2	0.052	0.054	0.057	0.058	0.059
Adjusted R^2	0.052	0.053	0.056	0.056	0.056
F Statistic	45.880 ***	56.628 ***	81.356 ***	79.905 ***	83.664 ***

* p<0.1;
** p<0.05;
*** p<0.01

Notes

1. Replication materials are located at https://github.com/markusneumann/BodyLanguage.
2. The score is the vector norm of the distance between the embeddings of two faces.
3. In addition to OpenPose, which produces 2D landmarks, we also considered models that create 3D representations of the human body. However, we found that these approaches frequently struggled with videos in which only parts of the entire body were visible. We also experimented with the AlphaPose framework, (Fang, Xie, Tai, & Lu, 2017; Li et al., 2018; Xiu, Li, Wang, Fang, & Lu, 2018), which works similarly to OpenPose. However, due to the greater speed of OpenPose, as well as the fact that it has a Windows executable, which makes it easy to use for those who are not familiar with Linux systems, we opted for OpenPose.
4. Through iteration, we decided that a 5 pixel average for all visible landmarks works best for this. This is a fairly conservative value, as different

persons are usually much further away. Even if, say, the hands of two people were fairly close together, their other body parts would still be far enough apart to make them easily distinguishable with this method.

5. Which GPU is provided depends on what is available at the time, as well as whether Google Colab or Colab Pro is used. Since we used both, and processed the data in batches (due to time limits on Colab), we got a lot of different GPUs.
6. One of the downsides of neural models in general, especially when they are computed on varying hardware, is that they cannot be 100% replicable; see https://pytorch.org/docs/stable/notes/randomness.html.
7. It is fairly common for videos to show the candidate's torso, cutting off at the wrists, so that their hands flicker in and out of the pose detection.
8. Another reason for this loss of data is that even though our face recognition model is quite accurate, it will occasionally miss the candidate. If we assume an error rate of 7% (see above), the candidate will be missed every 14 frames – almost twice per second. While we interpolate the candidate if their face is detected in both adjacent frames and the pose 'skeleton' differs little to these frames, this process still involves a considerable amount of data 'wastage'.
9. Our measure is different from Koppensteiner et al. (2016), who use the distances between maxima and minima in a time series of movement measures. We do not follow this approach because campaign ads frequently consist of a series of quickly changing shots, not all of which feature the candidate. Consequently, this time series would be constantly interrupted. Our density-based approach captures the vertical space in a way that is practically different, but conceptually similar, and does so in a way that is much more robust for campaign ads. We also constructed a much simpler measure by calculating the Euclidean distance between a keypoint in any two consecutive frames and then averaging over the video. In ideal situations, where the candidate is viewed from the front and talks for an extended period of time, this simpler measure works fine. However, we found our density-based measure to produce much more reliable results when the conditions were not ideal, and it also correlates much more strongly with our human-coded validation measures.
10. We originally sampled 100, but subsequently raised our standards for when the face recognition as well as the video quality (originally, our dataset contained some videos that were effectively just a series of still images) were good enough, which eliminated several videos from our dataset, so we sampled and coded 40 more.
11. Contrast ads, those that mention both an opponent and a favored candidate, were also excluded.
12. We didn't detect the gender of the candidates automatically because we already had the hand-coded data. Hand-coding gender of people whose names and images are available is a very quick procedure. We processed hundreds of candidates in a few minutes this way. It is true that this could

have been done automatically, but there is an error rate to automated coding, and so we saw no reason to add additional complexity.

13. See opensecrets.org.
14. See https://cawp.rutgers.edu/.
15. See http://followthemoney.org.
16. Because all of our variables of interest – gender, party, office and incumbency status – are static within units, we cannot rely on a fixed effects model.
17. A candidate who runs for a different office in 2020 than in 2018 is counted as a separate unit.
18. An important limitation shared by all of these models is the low R-squared. For the models from Figure 3, this ranges from 0.052 to 0.058. This indicates that our independent variables explain very little variation in candidates' vertical hand movement. This is unsurprising: we are attempting to explain a two-dimensional representation of a physical phenomenon through social factors. It is very likely that there are a number of other influential factors which we cannot account for—and we can only speculate what they might be. Based on our observations, the physical location is one such factor, as candidates sitting at a table have much less space to move their hands around than candidates standing freely. Similarly, the camera angle likely also plays a role that we cannot account for here. It is likely that there are a number of other factors that are particular to a specific video. Ergo, we acknowledge that omitted variable bias is a limitation of our paper.

References

Bauer, N. M. (2015). Emotional, sensitive, and unfit for office? Gender stereotype activation and support female candidates. *Political Psychology*, 36(6), 691-708.

Berry, W. D., Berkman, M. B., & Schneiderman, S. (2000). Legislative professionalism and incumbent reelection: the development of institutional boundaries. *The American Political Science Review*, 94 (4), 859–874.

Boussalis, C., Coan, T. G., Holman, M. R., & Müller, S. (2021). Gender, candidate emotional expression, and voter reactions during televised debates. *American Political Science Review*, 115(4), 1242-1257. http://doi:10.1017/S0003055421000666

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*. Princeton University Press.

Bucy, E. P., & Grabe, M. E. (2007). Taking television seriously: A sound and image bite analysis of presidential campaign coverage, 1992–2004. *Journal of Communication*, 57(4), 652–675.

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2018). Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition* (*FG 2018*) (pp. 67–74).

Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., & Sheikh, Y. A. (2019). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Carpinella, C., & Bauer, N. M. (2021). A visual analysis of gender stereotypes in campaign advertising. *Politics, Groups, and Identities*, 9(2), 369–386.

Carsey, T. M., Winburn, J., & Berry, W. D. (2017). Rethinking the normal vote, the personal vote, and the impact of legislative professionalism in U.S. state legislative elections. *State Politics and Policy Quarterly*, 17(4), 465–488.

Cassese, E. C., & Holman, M. R. (2018). Party and gender stereotypes in campaign attacks. *Political Behavior*, 40(3), 785–807.

Dolan, K. (2018). *Voting for women: How the public evaluates women candidates*. Routledge.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573.

Esler, T. (2019). *Face recognition using pytorch*. Retrieved from https://github.com/timesler/facenet-pytorch

Everitt, J., Best, L. A., & Gaudet, D. (2016). Candidate gender, behavioral style, and willingness to vote: support for female candidates depends on conformity to gender norms. *American Behavioral Scientist*, 60(14), 1737–1755.

Fang, H.-S., Xie, S., Tai, Y.-W., & Lu, C. (2017). RMPE: Regional multi-person pose estimation. In *Iccv*.

Hayes, D. (2005). Candidate qualities through a partisan lens: A theory of trait ownership. *American Journal of Political Science*, 49(4), 908–923.

Herrnson, P. S., Lay, J. C., & Stokes, A. K. (2003). Women running "as women": Candidate gender, campaign issues, and voter-targeting strategies. *Journal of Politics*, 65(1), 244–255.

Holman, M. R., Schneider, M. C., & Pondel, K. (2015). Gender targeting in political advertisements. *Political Research Quarterly*, 68(4), 816–829.

Huddy, L., & Terkildsen, N. (1993a). The consequences of gender stereotypes for women candidates at different levels and types of office. *Political Research Quarterly*, 46(3), 503–525.

Huddy, L., & Terkildsen, N. (1993b). Gender stereotypes and the perception of male and female candidates. *American Journal of Political Science*, 119–147.

Jamieson, K. H. (1995). *Beyond the double bind: Women and leadership*. Oxford University Press.

Koch, J. W. (1999). Candidate gender and assessments of senate candidates. *Social Science Quarterly*, 80(1), 84–96.

Koppensteiner, M., & Grammer, K. (2011). Body movements of male and female speakers and their influence on perceptions of personality. *Personality and Individual Differences*, 51(6), 743–747.

Koppensteiner, M., Stephan, P., & Jäschke, J. P. M. (2016). Moving speeches: Dominance, trustworthiness and competence in body motion. *Personality and Individual Differences*, 94, 101–106.

Li, J., Wang, C., Zhu, H., Mao, Y., Fang, H.-S., & Lu, C. (2018). Crowdpose: Efficient crowded scenes pose estimation and a new benchmark. *arXiv preprint arXiv:1812.00324*.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., … Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In *European conference on computer vision* (pp. 740–755).

Mehrabian, A. (1972). *Nonverbal Communication*. New Brunswick: Aldine Transaction.

Pollick, F. E., Kay, J. W., Heim, K., & Stringer, R. (2005). Gender recognition from point-light walkers. *Journal of Experimental Psychology: Human Perception and Performance*, 31(6), 1247.

Sanbonmatsu, K., & Dolan, K. (2009). Do gender stereotypes transcend party? *Political Research Quarterly*, 62(3), 485–494.

Sapiro, V., Cramer Walsh, K., Strach, P., & Hennings, V. (2011). Gender, context, and television advertising: A comprehensive analysis of 2000 and 2002 House races. *Political Research Quarterly*, 64(1), 107–119.

Schneider, M. C., & Bos, A. L. (2014). Measuring stereotypes of female politicians. *Political Psychology*, 35(2), 245–266.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815–823).

Simon, T., Joo, H., Matthews, I., & Sheikh, Y. (2017). Hand keypoint detection in single images using multiview bootstrapping. In *CVPR*.

Squire, P. (1992). The theory of legislative institutionalization and the california assembly. *Journal of Politics*, 54(4), 1026–1054.

Wasike, B. (2019). Gender, nonverbal communication, and televised debates: a case study analysis of Clinton and Trump's nonverbal language during the 2016 town hall debate. *International Journal of Communication*, 13, 26.

Winter, N. J. (2010). Masculine Republicans and feminine Democrats: Gender and Americans' explicit and implicit images of the political parties. *Political Behavior*, 32(4), 587–618.

Xiu, Y., Li, J., Wang, H., Fang, Y., & Lu, C. (2018). Pose Flow: Efficient online pose tracking. In *Bmvc*.

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10), 1499–1503.