Face-to-face: Three facial features that may turn the scale in close electoral races☆

Jan R. Landwehr [a,*], Michaela Wänke [b]

[a] Goethe University Frankfurt, Faculty of Economics and Business, Theodor-W.-Adorno Platz 4, 60629 Frankfurt am Main, Germany
[b] University of Mannheim, Professorship of Consumer and Economic Psychology, A,5,6, 68159 Mannheim, Germany

ARTICLE INFO

ABSTRACT

When voters in political elections intend to vote for a candidate who represents their political interests, they cannot avoid being influenced by the visual appearance of the candidates. In fact, many political campaigns are dominated by billboard advertising that mainly focuses on politicians' face portraits. While extant research already showed that a candidate's visual appearance has an impact on electoral outcomes, the present research goes a step further and investigates which specific facial features increase or decrease a politician's likelihood of winning an election. Evaluating 17 facial features, we identified mouth width, the ratio of eye height/width, and eyebrow height as relevant predictors of electoral success. In particular, we find across a correlative study, an experimental study, and an analysis of German Federal Elections that a wide mouth and large eyes that extend vertically have a robust and coherent influence on the likelihood of getting votes. Thin eyebrows increase the likelihood of getting votes in our hypothetical voting scenarios but not in the analysis of real election data. Furthermore, we show that the effects of the mouth and the eyes are statistically mediated by perceived trustworthiness and perceived dominance. In terms of real-life relevance, our analysis suggests that these three facial features can make the difference in close races.

Looks are important in winning votes. Although unsettling that voting for a particular candidate is not entirely determined by that candidate's politics, attitudes, values, abilities, expertise, or character, numerous studies attest to the power of looks. Merely by looking at photos (Antonakis & Dalgas, 2009) or silent video clips (Benjamin & Shapiro, 2009) of competing candidates, viewers quite accurately guessed who finally won the election. Even children's choices of who should be captain of a boat predicted the winners of political elections with 71 % accuracy (Antonakis & Dalgas, 2009). Voters' inability to ignore facial cues in their voting decision is also evident from the finding that "Republican voters […] are more likely to vote for a candidate (even a Democrat) the more that person has a stereotypically Republican-looking face" (Olivola, Tingley, & Todorov, 2018, p. 1157). Most intriguingly, Little, Burris, Jones, and Roberts (2007) computed a vector that represented the difference between the faces of the winner and the loser of a real election. They then built two new faces by shifting a neutral face along this vector in the direction of the winner or the loser. Mock elections between these newly created faces showed high agreement with the actual election outcome. Apparently, winners got the

winning look.

But what is the winning look? What are the facial features that win elections? Surprisingly, very few studies looked at specific facial features. While many studies provide evidence for the effects of candidate attractiveness (e.g., Berggren, Jordahl, & Poutvaara, 2010; Franklin & Zebrowitz, 2016), perceived competence (e.g., Antonakis & Dalgas, 2009; Olivola & Todorov, 2010; Todorov, Mandisodza, Goren, & Hall, 2005), perceived dominance (Little et al., 2007; Rule et al., 2010), or perceived trustworthiness (Chen, Jing, & Lee, 2014; Franklin & Zebrowitz, 2016), the role of objective and concrete facial features has been relatively neglected (for a notable exception see Re and Rule (2016), who showed a positive effect of mouth width on electoral success). One might expect that facial features related to subjective perceptions of beneficial traits such as trustworthiness or dominance should also be crucial in winning elections, but a closer look at the literature reveals that this is not so simple.

First, there is generally a lack of evidence which concrete facial features correspond with subjective ratings of a person's attributes. Most prior studies relate facial trait inferences such as facial competence to

---

holistic facial configurations such as facial masculinity or facial maturity (e.g., Olivola & Todorov, 2010) that are ultimately, however, reflections of simultaneous variations in several concrete facial features (such as mouth width or eye size). In this regard, the probably best-studied holistic facial configuration in person perception is baby-faceness, which is also not a single feature but the subjective impression of face maturity denoted by a combination of features that differentiate small children from adults (Berry & McArthur, 1985; Montepare & Zebrowitz, 1998). Thus, it is unclear whether large eyes or small noses or small chins, all characteristics of baby-faceness among others, or combinations of these determine perceived qualities. Besides, although baby-faceness corresponds to some personality attributions in some studies, there is mixed evidence for its role in predicting election results, with some studies finding a positive impact (Chang, Lee, & Cheng, 2017; Poutvaara, Jordahl, & Berggren, 2009), others finding a negative impact or no impact (Franklin & Zebrowitz, 2016; Olivola & Todorov, 2010).

Second, not only is the literature on linking specific facial features to perceived traits scarce, it also shows partly contradictory findings. For example, Vernon, Sutherland, Young, and Hartley (2014) found a negative correlation between perceived trustworthiness and eye height (vertical diameter). This finding contradicts Dotsch and Todorov's (2012) observation that a trustworthy face involves open eyes.

Third, one particular characteristic may have opposite implications for different relevant attributes. For example, whereas baby-faceness is negatively correlated with perceived competence and perceived dominance (Montepare & Zebrowitz, 1998; Poutvaara et al., 2009), it is positively correlated with trustworthiness (Montepare & Zebrowitz, 1998 for a review).[1]

Due to this lack of clear theoretical predictions, we chose to begin with a data-based approach to identify facial features that may enhance the likelihood of winning elections. In the first study, participants rated forty faces according to how likely they would vote for the person. Correlations with seventeen face measures identified three features that might make a face more electable: eyebrow height, mouth width, and eye ratio (ratio of eye height and eye width). In study 2, we validated these three features by systematically varying them in an experimental design. Moreover, based on the model of Oosterhof and Todorov (2008), we also assessed perceived trustworthiness and dominance as possible mediators. Finally, an analysis of the 2009 German Parliamentary elections brought confirmatory evidence that candidates with a wider mouth and a larger eye ratio have better chances of winning elections. However, the positive effect of thin eyebrows on hypothetical voting decisions found in studies 1 and 2 did not replicate in the real election data.

1. Transparency and data accessibility statement

Data collection for all reported studies followed APA's ethical principles in the conduct of research with human participants (American Psychological Association, 2017). None of the studies reported in this article was formally preregistered. All measures, manipulations, and exclusions in the studies are disclosed, and data analysis started only after data collection was finished. All materials are based on publically available stimulus databases. All data and code used in the analyses are available on OSF: https://osf.io/h9ygw/

2. Study 1: Identification of facial features

The aim of Study 1 was to identify those facial features that are predictive of people's stated likelihood of voting for a political candidate. To this end, we showed participants 20 female and 20 male unknown facial photos and asked them to indicate for each face how likely they would vote for this person in an important election (i.e., electability). Referring to Cunningham (1986), we measured 17 objective facial features (e.g., length of nose, width of mouth, etc.) in the digital facial photos using image processing software. Based on this data, we computed bivariate correlations between objective facial features and rated electability to identify the most important facial features to be used in our experimental study (Study 2) and our analysis of real election data (Study 3). We used R 4.0.2 (R Core Team, 2020) for the analyses of all three studies.

2.1. Method and procedure

A sample of 27 participants (78% female, 22% male, $M_{age} = 29.46$) was recruited for voluntary participation in an online study. Sample size corresponds to usual standards in the field of facial perception when ratings of face databases are collected (e.g., Re & Rule, 2016; Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013). We selected 20 standardized female and 20 standardized male photos of Caucasian[2] people from the AR Face Database (Martinez & Benavente, 1998). All facial pictures were standardized to a resolution of 400 × 420 pixels. Participants saw all faces[3] in random order and answered the following item for each face on a 7-point scale: "Would you elect this person to an important political office?" (1 = "I could hardly imagine it", 7 = "I could imagine it well"). Stimulus presentation duration and rating were self-paced. Afterward, participants completed an exit questionnaire, were debriefed, and thanked for participating.

Using a customized script for the statistical software R, we measured 17 distances/ratios (measurement scale: number of pixels) in the 40 faces following common approaches in research on facial perception and attractiveness (e.g., Cunningham, 1986). Fig. 1 shows the 17 facial metrics we use in the present research. Importantly, following Cunningham (1986), all horizontal (vertical) distances are divided by the total facial width (height) to obtain relative instead of absolute facial metrics.

2.2. Results and discussion

In the first step, we checked for any gender effects. Male participants provided higher electability judgments than female participants, and female faces were rated higher on electability than male faces. Importantly, only these two significant main effects occurred, but no interaction ($p > .71$). To remove these absolute level effects from further analyses, we z-standardized the electability judgments within each of the four possible combinations [2 (rater gender: female vs. male) x 2 (face gender: female vs. male)] to obtain a normalized dependent variable. As we aim to analyze our data on the level of the individual faces (aggregated across participants), we tested whether aggregating the data is justified. We found that 81% of the variance in the electability judgments was due to differences between the 40 faces, and only the remaining 19% was due to heterogeneity in participants' judgments, which justifies aggregation. Accordingly, we aggregated the normalized

---

[1] Although some studies found no relationship (Franklin & Zebrowitz, 2016; Poutvaara et al., 2009).

[2] We focused on Caucasian faces because most politicians we analyze in our third study are Caucasians and this first study is intended to lay the foundation for the later studies. Because perceptions of faces and corresponding trait inferences may vary conditional on ethnicity, we decided to focus on just one ethnicity to ensure the internal validity of our findings and to avoid another layer of complexity in our data. Accordingly, we explicitly note that this focus on just one ethnicity is a limitation of the current study in terms of generalizability.

[3] We also showed three additional faces: A morph of all male faces, a morph of all female faces, and a morph of all 40 faces. Furthermore, participants also rated all faces in a second round on another inferred trait characteristic. Since the scope of this research does neither extend to facial morphing nor to trait characteristics other than electability, this additional data is not used for the present research.
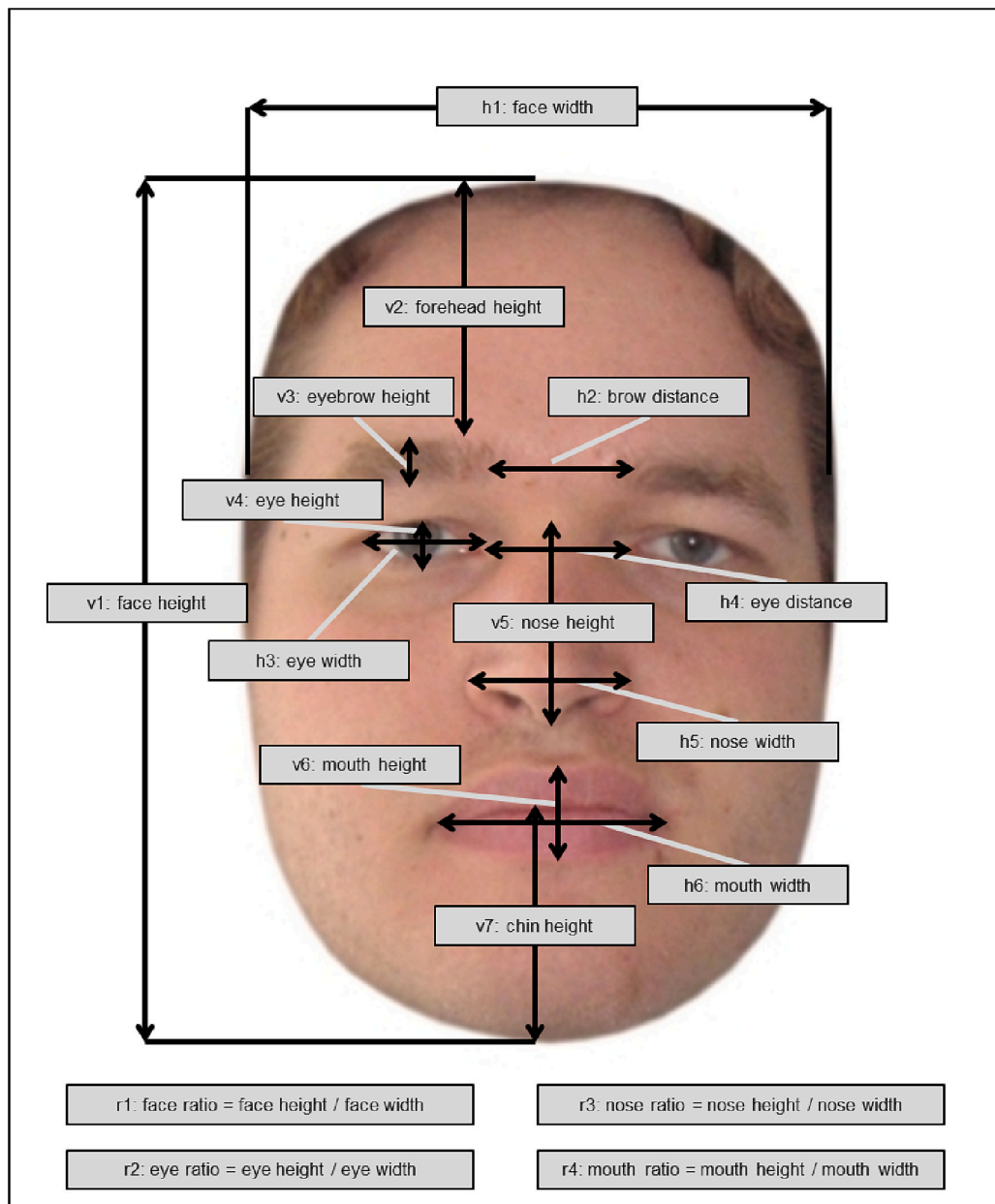
Fig. 1. Overview of all 17 facial features measured in the faces of Study 1.

electability judgments per face. In order to also remove any absolute differences in the measured facial metrics between female and male faces, we also z-standardized all 17 facial metrics within the female and male faces.

Based on this preprocessed data, we computed bivariate correlations between the dependent variable electability and each of the 17 facial metrics. We also computed 95%-confidence intervals for the correlations based on a bootstrapping with 5000 random draws. The results of this analysis are provided in Table 1 and show that only eyebrow height (v3; $r = -0.43$; 95%-CI: [-0.66; -0.13]), eye height (v4; $r = 0.35$; 95%-CI: [0.04; 0.63]), mouth width (h6; $r = 0.46$; 95%-CI: [0.22; 0.66]), and the ratio of eye height and eye width (r2: eye ratio; $r = 0.30$; 95%-CI: [0.03; 0.56]) are significantly correlated with electability. A post hoc sensitivity power analysis using G*Power 3.1.9.2 shows that the employed sample size of $N = 40$ allows detecting correlations of $r > 0.26$ with a power of 0.80. The four significant correlations are well above this critical threshold.

Out of these four significant measures, only eye height (v4) and eye

ratio (r2) are correlated with each other ($r(38) = 0.866$; $p < .001$; all other $r(38) < |0.193|$ and $p > .233$), which makes perfect sense since eye height is the nominator when computing the eye ratio. In line with Berry and McArthur (1985), we focus on the eye ratio in the subsequent analyses and studies because this measure captures the perceptually more important global gestalt of the eyes instead of just one perceptual feature. Interestingly, of all the features related to baby-faceness only large eyes had a significant effect, but chin, nose, and forehead size did not. Possibly, this explains the mixed findings in previous studies.

To examine the joint impact of the significant facial features, we regressed electability on eyebrow height ($b = -0.12$ [-0.22, -0.002]), mouth width ($b = 0.16$ [0.09, 0.25]), and eye ratio ($b = 0.09$ [0.001, 0.18]). All three predictors remain significant in this joint analysis, as indicated by their 95%-confidence intervals based on 5000 bootstraps. A post hoc sensitivity power analysis using G*Power 3.1.9.2 shows that the employed sample size allows detecting a minimum effect size of $f^2 = 0.161$, which is usually considered a medium-sized effect (Cohen, 1992). As further robustness checks, we estimated three additional regression

Table 1

Bivariate correlations of facial features with electability (Study 1).

Variable	r	95%-lower bound	95%-upper bound
v1: face height	−0.13	−0.42	0.17
v2: forehead height	0.04	−0.26	0.35
v3: eyebrow height	−0.43*	−0.66	−0.13
v4: eye height	0.35*	0.04	0.63
v5: nose height	0.10	−0.22	0.41
v6: mouth height	0.04	−0.26	0.32
v7: chin height	−0.07	−0.31	0.19
h1: face width	−0.18	−0.49	0.12
h2: brow distance	0.10	−0.20	0.36
h3: eye width	0.13	−0.26	0.48
h4: eye distance	−0.14	−0.46	0.19
h5: nose width	0.07	−0.27	0.42
h6: mouth width	0.46*	0.22	0.66
r1: face ratio	0.11	−0.16	0.39
r2: eye ratio	0.30*	0.03	0.56
r3: nose ratio	0.02	−0.28	0.31
r4: mouth ratio	−0.13	−0.42	0.17

Note. v1-v7: vertical measures; h1-h6: horizontal measures; r1-r4: ratios (height/width). Provided are correlation estimates for each facial feature with subjective electability and 95%-confidence intervals based on 5000 bootstraps; *$p < .05$.

models, where (1) all possible interactions are also included; (2) quadratic effects of the predictors are included; (3) interactions between facial features and gender of the face are included. In all three alternative regression models, the three main effects remain at least marginally significant ($p < .11$), and none of the interactions or quadratic effects becomes significant (quadratic eyebrow effect: $p = .06$; all other $p > .34$). Hence, for the remaining studies, we focus on the main effects of the three facial features identified as significant predictors of electability in the present study: eyebrow height, mouth width, and eye ratio.

3. Study 2: Experimental evidence

Study 1 identified three facial characteristics that are related to political electability based on a correlative dataset. The aim of Study 2 was to cross-validate these results by experimentally manipulating the three facial characteristics to test the causality of their effects. Moreover, a second aim was to examine the underlying psychological mechanism that makes these features enhance the likelihood of winning elections. To this aim, we are drawing on the work of Oosterhof and Todorov (2008), who proposed that judgments of trustworthiness and dominance are the two basic dimensions underlying all face evaluations. Based on this work, we assume that these two dimensions potentially mediate the impact of the three facial features on electability. Therefore, we also measured subjective perceptions of trustworthiness and dominance.[4] In addition to a continuous electability rating like in Study 1, Study 2 also employs more realistic choice scenarios typical of political elections. More specifically, we provided participants with a hypothetical political race in which four candidates competed, and participants had to vote for one of the four candidates.

3.1. Stimulus material

We built two choice sets: one consisting of four male faces and one consisting of four female faces. We decided not to mix gender within one choice set to control for a potential main effect of gender.[5] Because we intend to identify the effects of manipulated facial features, we decided to use faces that are, per se, neutral concerning electability, which should enable the facial features to have a stronger impact on perceived electability. To this end, we selected the four female and four male faces closest to the average electability evaluation in Study 1 ($M = 3.42$ on a scale from 1 to 7). The selected four female and four male faces ($M = 3.46$; range: 3.41-3.52) were very close to the average electability evaluation and can, thus, a priori be regarded as being neutral with respect to electability. Next, we built 2 (eyebrow height: thin vs. thick) × 2 (mouth width: narrow vs. wide) × 2 (eye ratio[6]: small vs. large) = 8 versions of each face (see Fig. 2).

In particular, we determined the standard deviation of each of the three facial features in the first study's 20 female and 20 male faces (separated by face gender) and reduced or enlarged each facial feature by one standard deviation using image morphing software (gtkmorph). An example of the resulting eight versions of one face is provided in Fig. 2.

3.2. Method and procedure

A sample of 396 participants (45% female, 55% male, $M_{age} = 34.97$) was recruited on Amazon MTurk for an online study that took, on average, approximately 5 min. This sample size aligns with common recommendations for Multinomial Logistic Models that suggest an optimal predictive performance for samples larger than 50 events per predictor variable (de Jong et al., 2019). Our most complex Multinomial Logistic Model contains seven predictor variables such that a sample larger than 350 participants can be considered adequate. The study consisted of five parts. First, participants saw the four male faces. For each face, one of the 2 (eyebrow height: thin vs. thick) × 2 (mouth width: narrow vs. wide) × 2 (eye ratio: small vs. large) = 8 versions was randomly selected. Participants were asked for which of the four faces they would most likely vote merely based on facial appearance. In the second phase, the procedure was repeated with female faces. In the third phase, all eight faces (four female + four male faces) were presented again in random order, and participants were asked to judge each face on the following three continuous visual analog slider bars: (1) "I could not imagine voting for this politician" - "I could well imagine voting for this politician"; (2) "not at all trustworthy" - "extremely trustworthy"; (3) "not at all dominant" – "extremely dominant". The ratings were internally recorded with numeric integers between 1 and 101. In a fourth phase, participants saw for each of the eight faces (four female and four male faces) all eight (2x2x2) versions of the face on a single screen and were asked which of the eight versions of one face they would most likely elect merely based on facial appearance. Thus, in this fourth task, the manipulation of facial features is revealed to participants, and we examined whether a within-subjects manipulation of facial features would also show the predicted effects. Importantly, the manipulation of the facial features is only revealed in this fourth task. Hence, the manipulation of facial features can be regarded as a between-subjects manipulation in the three earlier tasks. Finally, participants completed

---

[4] Prior research has shown that subjective facial competence judgments are the best predictor of electoral success (Olivola & Todorov, 2010). Nevertheless we decided to focus on trustworthiness and dominance as putative mediators because Oosterhof and Todorov (2008) found that "valence and dominance […] are sufficient to describe face evaluation and show that these dimensions can be approximated by judgments of trustworthiness and dominance." (p. 11087). Moreover, Olivola and Todorov (2010) found that facial competence judgments are highly correlated with facial trustworthiness judgments ($r = 0.55$) and we consider trustworthiness the more fundamental judgmental dimension based on Oosterhof and Todorov (2008).

[5] A limitation of choice sets that keep candidates' gender constant is that we cannot assess whether our results also hold for political races in which candidates of different genders compete. Recent research has, however, shown that voting behavior may differ between political races where only participants of the same vs. of different genders compete (Jäckle, Metz, Wenzelburger, & König, 2020).

[6] We actually manipulated the height of the eyes. Thus, eye ratio was changed by changing the nominator of the ratio.
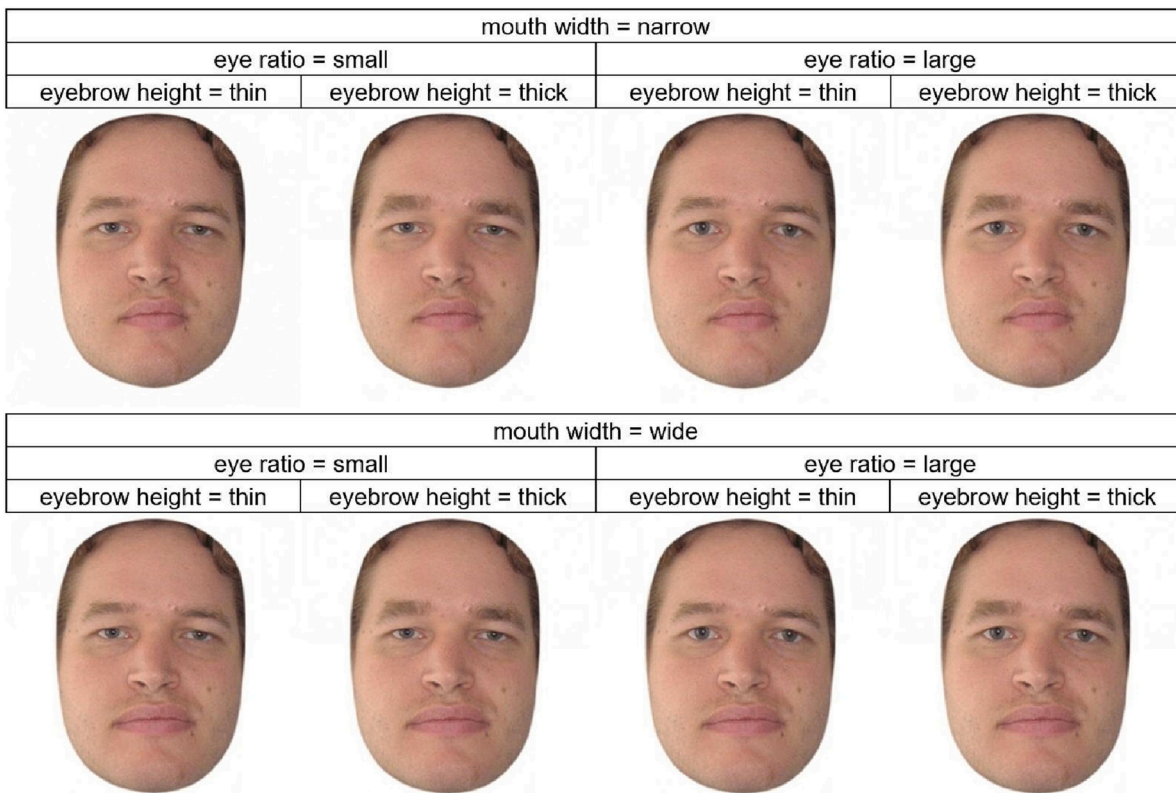
Fig. 2. Exemplary manipulation of facial features in Study 2.
Note. The third facial configuration in the second row (i.e., wide mouth, large eye ratio, and thin eyebrows) is expected to receive the highest share of hypothetical votes (see Fig. 3).

an exit questionnaire, were debriefed, and thanked for participating.

3.3. Results

3.3.1. Male and female faces choice tasks

Fig. 3 depicts the vote share for each facial feature combination aggregated over the male and female voting scenarios (parts 1 and 2 of the study). With eight combinations, the share that can be expected by pure chance is 12.5%. As can be seen, only the combination of a wide mouth, large eye ratio, and thin eyebrows leads to a choice probability significantly above the chance level.

As a more formal test of the hypothesis that mouth width and eye ratio contribute positively to electability and eyebrow height contributes negatively, we applied a Multinomial Mixed Panel Data Logit Model with random parameters using the mlogit()-function (Version 0.3-0) of the statistical software R (Croissant, 2013).

In accordance with the results of Study 1, we find that the likelihood of "voting" for a particular face is positively influenced by mouth width ($b = 1.03$; $SE = 0.28$; $z = 3.69$; $p < .001$) and eye ratio ($b = 0.46$; $SE = 0.23$; $z = 2.04$; $p = .041$) and negatively influenced by eyebrow height ($b = -0.46$; $SE = 0.21$; $z = -2.20$; $p = .028$). As a robustness check, we also estimated a model that additionally includes the three two-way interactions and the three-way interaction between the facial features. However, adding these four additional parameters does not significantly improve model fit ($\chi^2(4) = 8.47$; $p = .076$), which suggests that the "main effects only"-model is an adequate representation of the empirical data.

3.3.2. Mediation analysis based on continuous evaluations

To better understand the underlying mechanism, we examined whether the three facial features influence perceptions of trustworthiness and dominance and whether these perceptions influence evaluations of electability and, ultimately, the likelihood of being chosen, as observed in the previous analysis of the choice tasks. To account for the repeated measures structure of the data, we used the lme ()-function of the nlme-library of the statistical software R (Pinheiro et al., 2015) to analyze the effects on trustworthiness, dominance, and electability. In order to analyze the effects on voting behavior, the mlogit function described earlier was used. The significance of the indirect effects is evaluated using a bootstrapping procedure with 5000 random draws. Fig. 4 provides an overview of the conceptual mediation model and the estimated path coefficients for the indirect effects of theoretical interest.

For mouth width, we find positive effects on trustworthiness ($b_{11} = 0.13$; $SE = 0.03$; $t(2769) = 3.87$; $p < .001$) and dominance ($b_{21} = 0.08$; $SE = 0.03$; $t(2769) = 2.17$; $p = .030$) and positive indirect effects on voting through trustworthiness and electability ($b_{11}*b_{34}*b_{46} = 0.07$ [0.03, 0.12]) and through dominance and electability ($b_{21}*b_{35}*b_{46} = 0.01$ [0.001, 0.03]).

For eye ratio, we find a positive effect on trustworthiness ($b_{12} = 0.17$; $SE = 0.03$; $t(2769) = 5.33$; $p < .001$) and a negative effect on dominance ($b_{22} = -0.08$; $SE = 0.03$; $t(2769) = -2.25$; $p = .024$) and a positive indirect effect on voting through trustworthiness and electability ($b_{12}*b_{34}*b_{46} = 0.10$ [0.06, 0.15]) and a negative indirect effect through dominance and electability ($b_{22}*b_{35}*b_{46} = -0.01$ [-0.03, -0.002]).

For eyebrow height, we do not find an effect on trustworthiness ($b_{13} = -0.02$; $SE = 0.03$; $t(2769) = -0.56$; $p = .576$) and only a marginally significant positive effect on dominance ($b_{23} = 0.06$; $SE = 0.03$; $t(2769) = 1.86$; $p = .063$). The indirect effects are not significant for the eyebrows (indirect effect on voting through trustworthiness and electability ($b_{13}*b_{34}*b_{46} = -0.01$ [-0.05, 0.03]); indirect effect on voting through dominance and electability ($b_{23}*b_{35}*b_{46} = 0.01$ [-0.002, 0.03])).

In sum, the results of the mediation analysis are consistent with the interpretation that mouth width has a positive effect on voting behavior
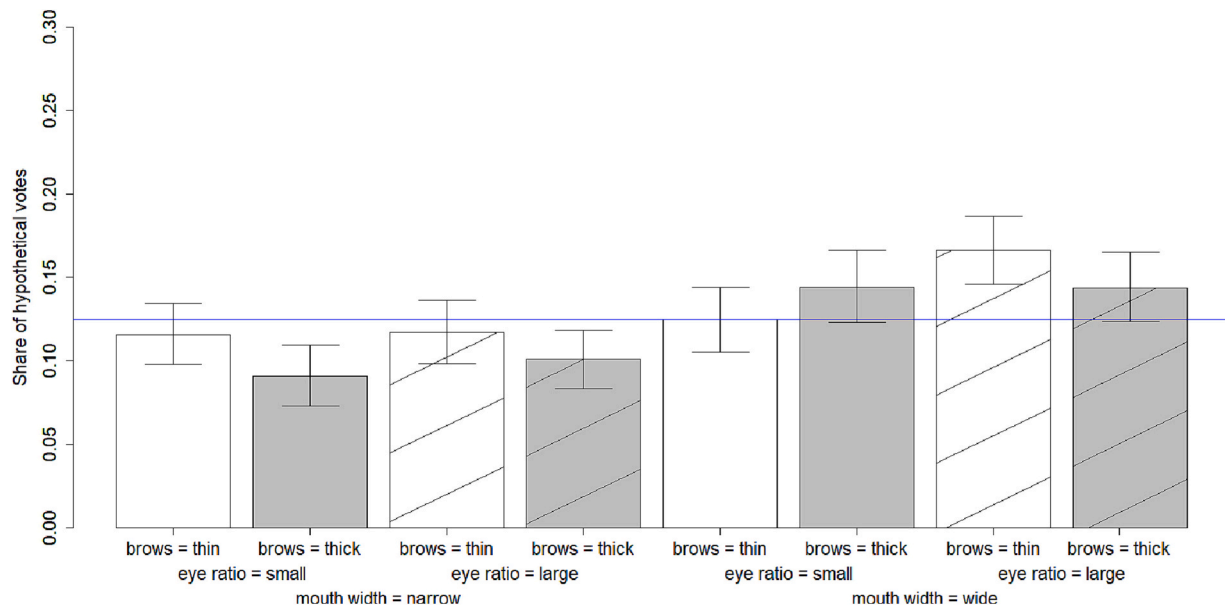
Fig. 3. Observed choice probabilities for the eight facial configurations of Study 2.
Note. The figure shows aggregated shares of hypothetical votes across male and female faces. Because there are eight ($2 \times 2 \times 2$) possible facial configurations, each has an a priori choice probability of 1/8 (12.5%), indicated by the horizontal line. 95%-confidence intervals based on 5000 bootstraps indicate that only the combination of a wide mouth, large eye ratio, and thin eyebrows leads to a choice probability significantly above the chance level.
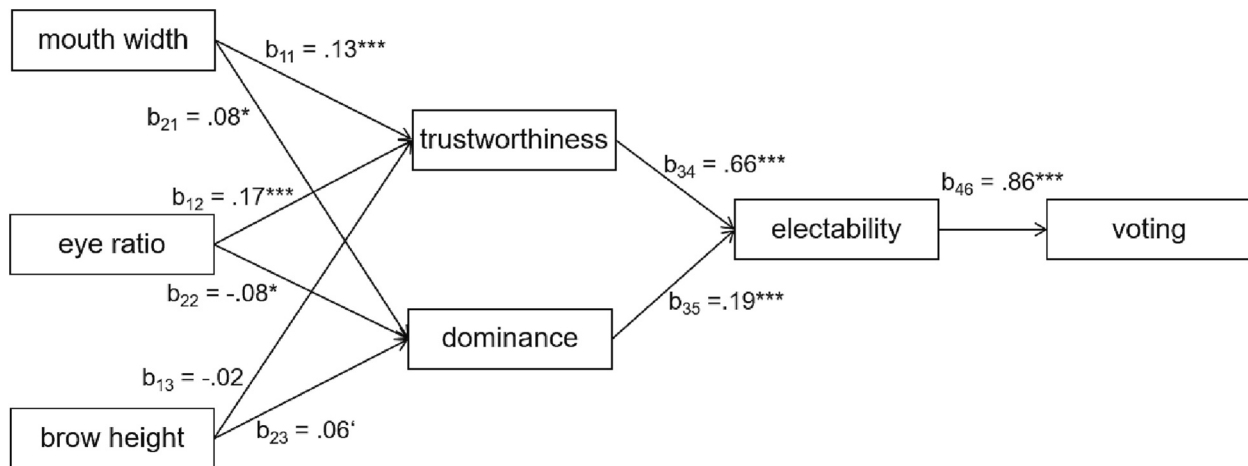

Fig. 4. Mediation model estimated for Study 2.
Note. To enhance the readability, the figure only shows the coefficients relevant for determining the indirect effects of theoretical interest. Of course, the statistical models also include the direct effects of the facial features on electability and voting and the direct effects of trustworthiness and dominance on voting. ***$p < .001$; **$p < .01$; *$p < .05$; '$p < .10$.

due to its positive effects on trustworthiness and dominance. The eye ratio seems to positively affect voting behavior due to its positive effect on trustworthiness, which is mitigated by a contrary, negative effect on dominance. However, because trustworthiness was identified as a much stronger predictor of electability and voting behavior than dominance, the total effect of eye ratio on electability and voting behavior is positive. Finally, the negative total effect of eyebrow height on voting behavior cannot be explained by trustworthiness and dominance. Thus, other mediators are likely to be at work for this facial characteristic.

In line with recent recommendations regarding the interpretation of mediation analyses (Fiedler, Harris, & Schott, 2018; Otter, Pachali, Mayer, & Landwehr, 2018), we intentionally used cautious wording when describing our results, and we would like to explicitly note that our proposed mediation model is only one of several possible conceptual structures that are compatible with the observed data. This cautionary

note is essential in the present case because we propose a serial mediation model with correlative paths between the two layers of mediators and between the mediators and the dependent variable. Yet, for theoretical reasons, we consider the proposed sequence plausible because trustworthiness and dominance are more fundamental perceptions than electability, and electability is conceptually a precursor of voting behavior.

3.3.3. Replication analysis of the within-subjects choice task

In the final choice task of the study, participants saw all eight manipulated versions of each of the eight faces and had to pick the version of each face they would most likely vote for. As we observed eight repeated choices, we estimated a Multinomial Mixed Panel Data Logit Model with random parameters using the mlogit()-function (Version 0.3-0) of the statistical software R (Croissant, 2013). In

accordance with the preceding analyses, we find that mouth width ($b = 0.26$; $SE = 0.04$; $z = 6.93$; $p < .001$) and eye ratio ($b = 0.34$; $SE = 0.04$; $z = 9.09$; $p < .001$) have positive effects and eyebrow height ($b = -0.11$; $SE = 0.04$; $z = -2.94$; $p = .003$) has a negative effect on choice probability. Thus, the effects also replicate in a within-subjects scenario, where participants become aware of the experimental manipulation of facial features, which attests to the robustness of the observed effects.

3.4. Discussion

Study 2 replicates the explorative finding of Study 1 in a well-controlled experimental setting. Across three different ways of measuring an increase in a face's likelihood of getting votes (i.e., choice scenario with a between-subjects manipulation of facial features, rating on a continuous scale, choice scenario with a within-subjects manipulation of facial features), we find converging evidence that mouth width and eye ratio increase the likelihood of getting votes. In contrast, eyebrow height decreases the likelihood of getting votes. Moreover, we find that the positive effect of mouth width is positively mediated by perceived trustworthiness and perceived dominance, and the positive effect of eye ratio is positively mediated only by perceived trustworthiness. For eyebrow height, we were not able to identify a meaningful mediating mechanism.

A crucial limitation of Studies 1 and 2 is that they only assessed hypothetical voting decisions. To determine whether these laboratory results also have validity in actual elections where usually more information than just a facial portrait is available, we analyzed real election data in Study 3.

4. Study 3: Analysis of real election data

The aim of Study 3 is to complement the evidence for an internally valid mechanism observed in Study 2 with external validity based on a real-life dataset. To this end, we examine the outcome of the 2009 German Federal election. For this election, Germany is divided into 299 electoral districts. In each district, candidates from the major parties compete for votes, and the candidate with the relative majority of votes will represent the district in the German Parliament. Importantly, in each district, each party's direct candidate is heavily advertised with billboard posters usually showing mainly the face of the candidate. Since we are interested in the effects of facial features on electoral success, this practice makes these direct votes an ideal dependent variable for our purpose.

4.1. Data description

We accessed the publically available official outcome of the 2009 German Federal election (www.bundeswahlleiter.de). For each electoral district, this dataset includes, among other things, information on the name of the direct candidates and the share of votes for each direct vote candidate. We only analyzed the data for the five parties eventually represented in Parliament (i.e., CDU/CSU, SPD, Green Party, FDP, and the left-wing party LINKE). In five of the 299 electoral districts, only four of the five parties nominated a direct candidate (either the green party or the left-wing party refrained from nominating a candidate in these districts). Based on this dataset, we conducted a web search for portrait photos of these 1490 candidates.[7] We found adequate photos for 1458 candidates (~98% of the candidates; 28% female; 72% male). The 32

candidates without a photo are excluded from any further analyses. Using the same customized script for the statistical software R already used in Study 1, we measured the three facial features (mouth width, eye ratio, and eyebrow height) identified as important voting behavior predictors in the two previous studies. Again, we z-standardized the facial features within the male and female faces, respectively, to remove any absolute level differences in the features between the two genders.

In line with previous research (Jäckle & Metz, 2017), we also collected data for several control variables that are either appearance-related or related to a candidate's political position/situation. In particular, we considered gender (0 = male; 1 = female; available from the official election report), age (continuous; based on a web search for the candidates), and the following appearance-related variables that we coded based on the collected portrait photos (0 = not present; 1 = present): suit/blazer, tie, glasses,[8] beard, and bald head. In terms of political control variables, we coded based on the previous German Federal election in 2005 whether a candidate has an advantage of incumbency (0 = no; 1 = yes), how contested an electoral district was (inspired by Jäckle and Metz (2017) we computed this variable as the difference between the vote share of the winning candidate and the vote share of the runner-up within an electoral district and multiplied this difference by -1 such that larger values encode a more intense contest), and each candidate's party affiliation[9] (using four dummy-variables with the overall winning party CDU as the base category; the data was available from the official election report).

4.2. Analysis and results

We report two different ways of modeling the data. First, we analyzed the percentage of direct votes each candidate attracted depending on the three facial features. To this end, we used the lme ()-function of the nlme library of the statistical software R (Pinheiro et al., 2015) to estimate a Linear Mixed Model, including a random intercept conditional on the electoral district to account for any random variation between districts. Second, we coded which of the five candidates of each electoral district won the district. Then, we applied a standard Multinomial Logit model using the mlogit()-function (Version 0.3-0) of the statistical software R (Croissant, 2013) to analyze the impact of the three facial features on the likelihood of winning an electoral district. Both modeling approaches included gender, age, suit/blazer, tie, glasses, beard, bald head, advantage of incumbency, contested electoral district,[10] and party affiliation as control variables. Table 2 provides the results of both statistical models.

For the first model, we find that mouth width ($b = 0.53$; $SE = 0.17$; $t(1144) = 3.05$; $p = .002$) and eye ratio ($b = 0.34$; $SE = 0.17$; $t(1144) = 2.03$; $p = .043$) significantly increase the likelihood of getting direct votes. The effect of eyebrow height ($b = -0.06$; $SE = 0.16$; $t(1144) = -0.39$; $p = .699$) is not significant but directionally consistent with studies 1 and 2. Concerning the control variables, we find effects of party membership (CDU/CSU serves as the reference category; the other four parties are dummy-coded: Green: $b = -24.16$; $SE = 0.62$; $t(1144) =$

---

[7] We conducted the web search in 2012, which is within the four-year term of office of the German parliament. In line with Jäckle and Metz (2017), we mostly used photos from the candidates' website, from the website of the German parliament, or from the websites of the candidates' party. It is important to note that most photos we collected in 2012 were actually photographed earlier (close to the actual election in 2009).

[8] see Fleischmann, Lammers, Stoker, and Garretsen (2019) for an in-depth discussion of the importance of glasses for electoral success.
[9] Because previous research has established that voters can infer political orientation and party affiliation from facial portraits (Herrmann & Shikano, 2016; Ivanov, Delmas, Muller, & Wänke, 2018; Olivola, Sussman, Tsetsos, Kang, & Todorov, 2012; Samochowiec, Wänke, & Fiedler, 2010; for a review see Wänke, 2015) and that in the U.S. Republican voters are more likely to vote for candidates with a stereotypically Republican-looking face (Olivola et al., 2012, 2018), we consider it especially important to control for party affiliation in our statistical analyses to account for a potential relationship between facial features and political orientation.
[10] This control variable was only included in model 1 but not in model 2 because the Multinomial Logit model cannot handle a predictor variable that does not vary within a district.

Table 2
Results of Study 3.

	Model 1) Linear Mixed Models analysis of the percentage of direct votes	Model 2) Multinomial Logit Model analysis of winning an electoral district
mouth width	0.53**	0.47***
eye ratio	0.34*	0.24*
eyebrow height	−0.06	−0.12
gender (0 = male; 1 = female)	0.82'	0.70'
age	−0.01	−0.01
suit/blazer (0 = no; 1 = yes)	1.53***	0.48
tie (0 = no; 1 = yes)	1.56**	0.90*
glasses (0 = no; 1 = yes)	0.23	0.05
beard (0 = no; 1 = yes)	−1.15**	0.31
bald head (0 = no; 1 = yes)	0.62	0.14
incumbency (0 = no; 1 = yes)	10.63***	2.45***
contested electoral district	0.02	NA
party: Green	−24.16***	−4.17***
party: Left	−21.70***	−1.44***
party: FDP	−25.50***	−19.69
party: SPD	−10.57***	−1.94***

Note. ***$p < .001$; **$p < .01$; *$p < .05$; '$p < .10$. The multinomial logit model (model 2) cannot include a predictor variable that does not vary within a district. Hence, there is no estimate for "contested electoral district". The estimate for "party: FDP" is not significant in model 2 because the FDP did not win a single district such that the standard error of the estimate is inflated.

-38.75; $p < .001$; LINKE: $b = -21.70$; $SE = 0.62$; $t(1144) = -35.10$; $p < .001$; FDP: $b = -25.50$; $SE = 0.56$; $t(1144) = -45.85$; $p < .001$; SPD: $b = -10.57$; $SE = 0.52$; $t(1144) = -20.29$; $p < .001$), suit/blazer ($b = 1.53$; $SE = 0.43$; $t(1144) = 3.56$; $p < .001$), tie ($b = 1.56$; $SE = 0.48$; $t(1144) = 3.26$; $p = .001$), beard ($b = -1.15$; $SE = 0.43$; $t(1144) = -2.66$; $p = .008$), and advantage of incumbency ($b = 10.63$; $SE = 0.53$; $t(1144) = 20.16$; $p < .001$). The effect of candidates' gender was marginally significant ($b = 0.82$; $SE = 0.48$; $t(1144) = 1.72$; $p = .085$), and the effects of all other control variables were not significant ($p > .284$). A post hoc sensitivity power analysis using G*Power 3.1.9.2 that makes the simplifying assumption that the observations are not nested shows that the employed sample size allows detecting a minimum effect size of $f^2 = 0.004$, which is usually considered a small effect (Cohen, 1992).

For the second model, we find that mouth width ($b = 0.47$; $SE = 0.13$; $z = 3.74$; $p < .001$) and eye ratio ($b = 0.24$; $SE = 0.12$; $z = 2.00$; $p = .046$) have significant positive effects on the likelihood of winning a district, while the effect of eyebrow height is not significant but directional consistent with studies 1 and 2 ($b = -0.12$; $SE = 0.12$; $z = -1.03$; $p = .304$). Concerning the control variables, we find strong effects of party membership (CDU/CSU serves as the reference category; the other four parties are dummy-coded: Green: $b = -4.17$; $SE = 1.02$; $z = -4.08$; $p < .001$; LINKE: $b = -1.44$; $SE = 0.32$; $z = -4.49$; $p < .001$; FDP: $b = -19.69$; $SE = 1698.3$; $z = -0.01$; $p = .991$; SPD: $b = -1.94$; $SE = 0.27$; $z = -7.11$; $p < .001$), tie ($b = 0.90$; $SE = 0.37$; $z = 2.45$; $p = .014$), and advantage of incumbency ($b = 2.45$; $SE = 0.30$; $z = 8.04$; $p < .001$). The effect of candidates' gender was marginally significant ($b = 0.70$; $SE = 0.40$; $z = 1.77$; $p = .077$), and the effects of all other control variables were not significant ($p > .192$).

4.3. Discussion

Study 3 replicates the pattern of results for the effects of mouth width and eye ratio found in Studies 1 and 2 in a field setting using real election data. However, the effect of eyebrow height, albeit directional consistent with Studies 1 and 2, was insignificant and, hence, turned out to be less robust and replicable in a field setting. Nevertheless, the key results of Study 2's well-controlled experiment seem externally valid. However, we explicitly caution the reader against over-interpreting our results. Although our results indicate that facial features influence voting behavior, it is important to note that, by far, the strongest predictors for the direct vote in an electoral district are a candidate's party membership and the advantage of incumbency. Therefore, our results do not imply that facial features can overwrite political preferences and are a major factor in electoral outcomes. Rather the results imply that in a very close race between candidates, facial features can turn the scale in one candidate's favor. For instance, all else being equal, a candidate with an advantageous facial configuration featuring a mouth width and an eye ratio each one standard deviation above average, and eyebrow height one standard deviation below average, will receive approximately one additional percentage point of votes compared to an average facial configuration given the estimates of the first statistical model.

5. General discussion

A large amount of literature attests to the importance of candidates' looks in political elections. This begs the question of what exactly makes a face draw votes. Going beyond subjectively ascribed traits, the present research gives an answer by identifying specific and quantifiable facial features. Across three studies and six different statistical modeling approaches, we find a robust and coherent pattern: Mouth width and eye ratio have positive effects on different indicators of voting probability, and eyebrow height has a negative yet less robust effect that is significant in three out of six analyses (see Table 3 for a summary of the key findings).

This pattern emerged in a correlational study (Study 1), was replicated when experimentally manipulating the respective features (Study 2), and was further confirmed for the effects of mouth width and eye ratio in a field study that analyzed actual votes in the 2009 German federal election (Study 3). Hence, politicians with wide mouths and large eyes that extend vertically should be more likely to get votes based on their facial features—potentially further enhanced by thin eyebrows, as suggested by our lab studies.

By identifying three concrete and objective facial cues as determinants of voting behavior, our research makes three important theoretical and practical contributions to the literature. First, we conceptually replicate and extend prior findings by Re and Rule (2016), who also showed evidence for a positive effect of mouth width on electoral success. Yet, we examine two additional facial features and how they relate to subjectively ascribed traits. We identified perceived trustworthiness and perceived dominance as potential mediators for the

Table 3
Summary of the estimated total effects across the analyses of all three studies.

Analysis	Effect of mouth width	Effect of eye ratio	Effect of eyebrow height
Study 1) Continuous rating of electability	$b = 0.16$*	$b = 0.09$*	$b = -0.12$*
Study 2) Between-subjects choice task	$b = 1.03$***	$b = 0.46$*	$b = -0.46$*
Study 2) Continuous rating of electability	$b = 0.17$***	$b = 0.17$***	$b = -0.02$
Study 2) Within-subjects choice task	$b = 0.26$***	$b = 0.34$***	$b = -0.11$**
Study 3) Linear mixed model	$b = 0.53$**	$b = 0.34$*	$b = -0.06$
Study 3) Multinomial logit model	$b = 0.47$***	$b = 0.24$*	$b = -0.12$

Note. ***$p < .001$; **$p < .01$; *$p < .05$; '$p < .10$.

effects of mouth width and eye ratio.[11] Importantly, mouth width increases both perceived trustworthiness and perceived dominance, which both have positive effects on electability. Eye ratio positively impacts perceived trustworthiness (i.e., probably due to the activation of a baby schema) but harms perceived dominance. However, the total effect of eye ratio is positive, which indicates that the path through trustworthiness outweighs the path through dominance. For the negative effect of eyebrow height on electability, we find that neither of the two potential mediators explains a sufficient amount of variance. Hence, the negative effect of eyebrow height needs to be explained by different perceptual inferences and calls for additional research on other mediator variables.

Second, our findings go beyond previous research that merely considered subjectively rated impressions such as perceived competence (e.g., Antonakis & Dalgas, 2009; Olivola & Todorov, 2010; Todorov et al., 2005), attractiveness (e.g., Berggren et al., 2010; Franklin & Zebrowitz, 2016), dominance (Little et al., 2007; Rule et al., 2010), trustworthiness (Chen et al., 2014; Franklin & Zebrowitz, 2016), or holistic facial configurations such as baby-faceness (e.g., Chang et al., 2017; Franklin & Zebrowitz, 2016; Olivola & Todorov, 2010; Poutvaara et al., 2009). Examining the impact of such subjective impressions on voting behavior imposes the challenge that some of these impressions are correlated and conceptually not clearly distinguishable. Accordingly, if researchers identify facial traits as relevant predictors of electoral success, their findings inherently depend on which traits they measured, how they labeled them, and how research participants interpreted them. For example, competence may be interpreted as intelligence, assertiveness, the ability to make decisions in times of crisis, or the ability to gather others behind a political position. Complicating matters further, attractiveness and trait inferences that drive voting decisions may, to some extent, be confounded (e.g., Banducci, Karp, Thrasher, & Rallings, 2008), making it difficult to disentangle the impact of specific factors. Thus, to answer our initial research question: "What makes a face win votes?", identifying objective facial features offers a cleaner and less ambiguous answer than merely relying on subjective impressions. However, we do not mean to dismiss the equally important research question "Why" do large eyes, a wide mouth, and perhaps thin eyebrows win votes? That is, what character inferences are drawn from these features?

By identifying objectively measurable facial features, we also respond to Fiedler (2014) who called for social psychological research that is not limited to intrapsychic phenomena but extends to environmental variables outside the organism to enable more powerful theory-building. Our research connects observable voting behavior to extrapsychic visual cues (i.e., concrete facial features) and uses subjective psychological assessments merely to connect the two (i.e., by using perceived trustworthiness and dominance as mediator variables). In this sense, our research approach and our empirical findings may inform fundamental theories of face perception, such as the emotional overgeneralization theory of perceived trustworthiness (Todorov, 2011; Zebrowitz & Montepare, 2008) that assumes that a wider mouth reminds people of the emotional expression of a smile, which translates into increased perceptions of trustworthiness. Similarly, the eyes have been found to convey information on complex mental states due to their optical adaption to specific perceptual functions and information-

processing needs (Lee & Anderson, 2017), which could be combined with our findings to explain the special importance of large eyes for judgments of trustworthiness and enhanced electability.

Third, in terms of practical implications, our research may hopefully be useful in boosting (Hertwig & Grüne-Yanoff, 2017) voters against the unwarranted influence of politicians' facial appearance by educating them which concrete facial features should be discounted in their judgments. In this regard, the identification of concrete facial features is more helpful than the identification of holistic facial configurations because only the former allows directing voters' attention to the specific perceptual input that needs to be discounted. As a practical example of this research's relevance, consider the 2017 German federal election case. During the election campaign, the right-wing party AFD posted edited pictures of the top candidate of the SPD, Martin Schulz, on Facebook where, besides other edits, the width of the mouth of Martin Schulz was reduced. According to our results, this unauthorized modification of the facial image of Martin Schulz should have reduced his perceived electability. Voters educated in discounting the impact of this facial cue could, on the one hand, better notice this modification in the first place and, on the other hand, discount the impact of this perceptual impression in their voting decision.

A potential limitation of our research and opportunity for future research accrues from our decision to focus on dominance and trustworthiness (in line with Oosterhof & Todorov, 2008) as potential mediators of our key effect, thereby ignoring other potentially important impressions, such as perceived competence. Nevertheless, we could predict electability very well by these two inferred traits alone (just these two inferred traits can explain 67% of the variance in electability), yet in our data, trustworthiness ($b = 0.66$) seemed more important than dominance ($b = 0.19$). In this respect, it should be noted that the weighting of these subjectively perceived traits is not universal and may depend on cultural preferences (Rule et al., 2010), socio-economic status (Unkelbach, Brütting, Schilling, & Wänke, n.d., in press), political ideology (Unkelbach et al., n.d., in press), and political conditions such as war or peace (Little et al., 2007). Thus, one might expect that in times of war, where power and dominance increase in weight, the advantage of large eyes may shrink. Hence, future studies may provide more data on cross-cultural and situational influences on what makes a face attract votes.

Another interesting line of future research would be to integrate our current findings on concrete facial features with findings on the effects of global facial typicality (e.g., Dotsch, Hassin, & Todorov, 2016; Sofer, Dotsch, Wigboldus, & Todorov, 2015). This related line of research showed that faces high in global typicality trigger more favorable evaluations and are perceived as more trustworthy. It would be interesting to examine the relative importance of global face typicality and the size of mouth, eyes, and eyebrows in shaping judgments of trustworthiness and overall evaluation.

In sum, the present research shows that three specific features in a politician's face can turn the scale in close electoral races. While politicians could be tempted to optimize their faces (at least their pictorial depictions) according to this finding, the present research should primarily motivate voters to keep a critical eye on their candidates and to potentially discount any effect of these three critical facial features.

Author contributions

J.L. and M.W. developed the research question and conjointly wrote the article. J.L. programmed the studies, collected the data, and conducted the statistical analyses. Both authors approved the final version of the manuscript for submission.

Declaration of Competing Interest

The authors declare no competing interests.

---

[11] On a cautionary note, we would like to stress that a significant mediation does not necessarily mean that we identified the real mediator, that other variables that we did not measure do not also play a role (or even a more prominent role), or that the impact of unobserved third variables on the relationship between mediator and dependent variable can be ruled out (Fiedler et al., 2018; Otter et al., 2018). In other words, while perceived trustworthiness and dominance had a positive impact on voting, it is not clear whether this relationship remains had we statistically controlled for additional perceived impressions.

References

American Psychological Association. (2017). Ethical principles of psychologists and code of conduct (2002, amended effective June 1, 2010, and January 1, 2017). http://www.apa.org/ethics/code/index.html.

Antonakis, J., & Dalgas, O. (2009). Predicting elections: Child's play! *Science, 323*(5918), 1183. https://doi.org/10.1126/science.1167748

Banducci, S. A., Karp, J. A., Thrasher, M., & Rallings, C. (2008). Ballot photographs as cues in low information elections. *Political Psychology, 29*(6), 903–917. https://doi.org/10.1111/j.1467-9221.2008.00672.x

Benjamin, D. J., & Shapiro, J. M. (2009). Thin-slice forecasts of gubernatorial elections. *Review of Economics and Statistics, 91*(3), 523–536. https://doi.org/10.1162/rest.91.3.523

Berggren, N., Jordahl, H., & Poutvaara, P. (2010). The looks of a winner: Beauty and electoral success. *Journal of Public Economics, 94*(1-2), 8–15.

Berry, D. S., & McArthur, L. Z. (1985). Some components and consequences of a babyface. *Journal of Personality and Social Psychology, 48*(2), 312–323.

Chang, C.-T., Lee, Y.-K., & Cheng, Z.-H. (2017). Baby face wins? Examining election success based on candidate election bulletin via multilevel modeling. *Asian Journal of Social Psychology, 20*(2), 97–112.

Chen, F. F., Jing, Y., & Lee, J. M. (2014). The looks of a leader: Competent and trustworthy, but not dominant. *Journal of Experimental Social Psychology, 51*, 27–33. https://doi.org/10.1016/j.jesp.2013.10.008

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Croissant, Y. (2013). mlogit: multinomial logit model. R package version 0.2-4. http://CRAN.R-project.org/package=mlogit.

Cunningham, M. R. (1986). Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty. *Journal of Personality and Social Psychology, 50*(5), 925–935.

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science, 3*(5), 562–571. https://doi.org/10.1177/1948550611430272

Dotsch, R., Hassin, R. R., & Todorov, A. (2016). Statistical learning shapes face evaluation. *Nature Human Behaviour, 1*, 1–6. https://doi.org/10.1038/s41562-016-0001

Fiedler, K. (2014). From intrapsychic to ecological theories in social psychology: Outlines of a functional theory approach. *European Journal of Social Psychology, 44*, 657–670. https://doi.org/10.1002/ejsp.2069

Fiedler, K., Harris, C., & Schott, M. (2018). Unwarranted inferences from statistical mediation tests – An analysis of articles published in 2015. *Journal of Experimental Social Psychology, 75*, 95–102. https://doi.org/10.1016/j.jesp.2017.11.008

Fleischmann, A., Lammers, J., Stoker, J. I., & Garretsen, H. (2019). You can leave your glasses on: Glasses can increase electoral success. *Social Psychology, 50*(1), 38–52. https://doi.org/10.1027/1864-9335/a000359

Franklin, R. G., & Zebrowitz, L. A. (2016). The influence of political candidates' facial appearance on older and younger adults' voting choices and actual electoral success. *Cogent Psychology, 3*(1), 1–12. https://doi.org/10.1080/23311908.2016.1151602

Herrmann, M., & Shikano, S. (2016). Attractiveness and facial competence bias face-based inferences of candidate ideology. *Political Psychology, 37*(3), 401–417. https://doi.org/10.1111/pops.12256

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science, 12*(6), 973–986. https://doi.org/10.1177/1745691617702496

Ivanov, I., Delmas, F., Muller, D., & Wänke, M. (2018). Interpersonal accuracy in a political context is moderated by the extremity of one's political attitudes. *Journal of Experimental Social Psychology, 79*, 95–106. https://doi.org/10.1016/j.jesp.2018.07.001

Jäckle, S., & Metz, T. (2017). Beauty contest revisited: The effects of perceived attractiveness, competence, and likability on the electoral success of German MPs. *Politics & Policy, 45*(4), 495–534. https://doi.org/10.1111/polp.12209

Jäckle, S., Metz, T., Wenzelburger, G., & König, P. D. (2020). A catwalk to congress? Appearance-based effects in the elections to the U.S. house of representatives 2016.

American Politics Research, 48(4), 427–441. https://doi.org/10.1177/1532673X19875710

de Jong, V. M. T., Eijkemans, M. J. C., van Calster, B., Timmerman, D., Moons, K. G. M., Steyerberg, E. W., & van Smeden, M. (2019). Sample size considerations and predictive performance of multinomial logistic prediction models. *Statistics in Medicine, 38*, 1601–1619.

Lee, D. H., & Anderson, A. K. (2017). Reading what the mind thinks from how the eye sees. *Psychological Science, 28*(4), 494–503. https://doi.org/10.1177/0956797616687364

Little, A. C., Burris, R. P., Jones, B. C., & Roberts, S. C. (2007). Facial appearance affects voting decisions. *Evolution and Human Behavior, 28*(1), 18–27. https://doi.org/10.1016/j.evolhumbehav.2006.09.002

Martinez, A. M., & Benavente, R. (1998). *The AR face database. CVC technical report #24 (June 1998)*.

Montepare, J. M., & Zebrowitz, L. A. (1998). Person perception comes of age: The salience and significance of age in social judgments. *Advances in Experimental Social Psychology, 30*, 93–161. https://doi.org/10.1016/S0065-2601(08)60383-4

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

Olivola, C. Y., Sussman, A. B., Tsetsos, K., Kang, O. E., & Todorov, A. (2012). Republicans prefer republican-looking leaders: Political facial stereotypes predict candidate electoral success among right-leaning voters. *Social Psychological and Personality Science, 3*(5), 605–613. https://doi.org/10.1177/1948550611432770

Olivola, C. Y., Tingley, D., & Todorov, A. (2018). Republican voters prefer candidates who have conservative-looking faces: New evidence from exit polls. *Political Psychology, 39*(5), 1157–1171. https://doi.org/10.1111/pops.12489

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Otter, T., Pachali, M. J., Mayer, S., & Landwehr, J. R. (2018). Causal inference using mediation analysis or instrumental variables – Full mediation in the absence of conditional independence. *Marketing ZFP, 40*, 41–57.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2015). nlme: Linear and nonlinear mixed effects models. R package version 3.1-120. URL: http://CRAN.R-project.org/package=nlme.

Poutvaara, P., Jordahl, H., & Berggren, N. (2009). Faces of politicians: Babyfacedness predicts inferred competence but not electoral success. *Journal of Experimental Social Psychology, 45*(5), 1132–1135. https://doi.org/10.1016/j.jesp.2009.06.007

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL https://www.R-project.org/ [Computer software].

Re, D. E., & Rule, N. O. (2016). The big man has a big mouth: Mouth width correlates with perceived leadership ability and actual leadership performance. *Journal of Experimental Social Psychology, 63*, 86–93. https://doi.org/10.1016/j.jesp.2015.12.005

Rule, N. O., Ambady, N., Adams, R. B., Jr., Ozono, H., Nakashima, S., Yoshikawa, S., & Watabe, M. (2010). Polling the face: Prediction and consensus across cultures. *Journal of Personality and Social Psychology, 98*(1), 1–15. https://doi.org/10.1037/a0017673

Samochowiec, J., Wänke, M., & Fiedler, K. (2010). Political ideology at face value. *Social Psychological and Personality Science, 1*(3), 206–213. https://doi.org/10.1177/1948550610372145

Sofer, C., Dotsch, R., Wigboldus, D. H. J., & Todorov, A. (2015). What is typical is good: The influence of face typicality on perceived trustworthiness. *Psychological Science, 26*(1), 39–47. https://doi.org/10.1177/0956797614554955

Todorov, A. (2011). Evaluating faces on social dimensions. In A. Todorov, S. T. Fiske, & D. Prentice (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind* (pp. 1–47). Oxford University Press.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*(5728), 1623–1626. https://doi.org/10.1126/science.1110589

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion, 13*(4), 724–738. https://doi.org/10.1037/a0032335

Unkelbach, F., Brütting, T., Schilling, N., & Wänke, M. (n.d.). Looking competent does not appeal to all voters equally: The role of social class and politicians' facial appearance for voting likelihood. Personality and Social Psychology Bulletin, (In press).

Vernon, R. J. W., Sutherland, C. A. M., Young, A. W., & Hartley, T. (2014). Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Science, 111*(32), E3353–E3361. https://doi.org/10.1073/pnas.1409860111

Wänke, M. (2015). It's all in the face: Facial appearance, political ideology and voters' perceptions. In J. P. Forgas, K. Fiedler, & W. D. Crano (Eds.), *Social psychology and politics* (pp. 143–162). Psychology Press.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2*(3), 1497–1517. https://doi.org/10.1111/j.1751-9004.2008.00109.x