Political Science Research and Methods (2025), **13**, 1031–1041
doi:10.1017/psrm.2024.38

RESEARCH NOTE

Facial finetuning: using pretrained image classification models to predict politicians' success

Asbjørn Lindholm[1], Christian Hjorth[1] and Julian Schuessler[2]

[1]Independent Researcher and [2]Department of Political Science and Centre for the Experimental-Philosophical Study of Discrimination (CEPDISC), Aarhus University, Aarhus, Denmark
Corresponding author: Julian Schuessler; Email: julians@ps.au.dk

(Received 30 October 2023; revised 13 February 2024; accepted 15 March 2024)

Abstract

There is a long-standing interest in how the visual appearance of politicians predicts their success. Usually, the scope of such studies is limited by the need for human-rated facial features. We instead fine-tune pretrained image classification models based on convolutional neural networks to predict facial features of 7,080 Danish politicians. Attractiveness and trustworthiness scores correlate positively and robustly with both ballot paper placement (proxying for intra-party success) and the number of votes gained in local and national elections, while dominance scores correlate inconsistently. Effect sizes are at times substantial. We find no moderation by politician gender or election type. However, dominance scores correlate significantly with outcomes for conservative politicians. We discuss possible causal mechanisms behind our results.

Keywords: political psychology; quantitiative methods

1. Introduction

A robust finding in psychology is that humans very quickly make inferences about social and personality characteristics based on others' visual appearance (Todorov *et al.*, 2015). The inferences also appear to contain substantial information: Based on photos alone, humans can predict a wide variety of outcomes, including political ones such as winning elections (Todorov *et al.*, 2005; Laustsen and Petersen, 2016), intraparty success (Laustsen and Petersen, 2017), and diversion of public funds (Casey, 2022). It appears that the human brain subconsciously processes detailed information that is imprinted in visual appearance and that through the process of natural selection was learned to be associated with human traits that are predictive of social outcomes even in the modern world. Furthermore, the visual appearance of humans may also have causal effects on others' behavior, thereby directly affecting outcomes.

The scope of such studies is usually restricted by the need to have humans rate photos. Furthermore, humans may recognize politicians from pictures and accordingly use additional information to make inferences about outcomes, complicating the interpretation of empirical relationships. However, modern machine learning opens up the possibility to mirror human inferences through complex statistical models, vastly increasing the scope of investigation and relying mechanically only on image data as the input. In this paper, we implement an approach based on transfer learning using fine-tuned pre-trained models. We take an existing convolutional neural network (CNN) trained for image classification (He *et al.*, 2016), fine-tune it to predict facial traits, using human ratings of 880 artificially generated faces (Peterson *et al.*, 2022), and then use it to predict central traits identified by the prior literature (attractiveness,

© The Author(s), 2024. Published by Cambridge University Press on behalf of EPS Academic Ltd. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psrm.2024.38 Published online by Cambridge University Press

trustworthiness, and dominance) for images of more than 7,000 Danish politicians. Predictive performance on a test set of artificially generated faces is good, with correlations of predicted and annotated ratings of up to 0.79. We are also able to replicate inter-correlations of traits reported in the prior literature.

Next, we evaluate how these traits correlate with real-world outcomes. We investigate both ballot paper placement, proxying for intra-party success (Laustsen and Petersen, 2017), as well as personal votes obtained in Danish local and national elections. We find that predicted attractiveness and trustworthiness correlate positively with both outcomes, in the case of votes substantially: A one standard deviation increase in the attractiveness score is associated with an 18 percent higher vote share, close to the estimate reported by Berggren *et al.* (2010) based on human ratings. These effects are mostly robust to controlling for candidates' gender, age, education, ethnicity, party, and incumbency. The picture is more complicated when it comes to dominance scores: They correlate weakly negatively with ballot paper placement and weakly positively with personal votes; however, this is sensitive to additional controls. We do find consistently that they correlate more positively with outcomes for conservative candidates, as predicted by psychological theories (Kleppestø *et al.*, 2019). We find no evidence that facial trait correlations vary by candidate gender or election type. Finally, we show that our results can plausibly be explained in quantitative terms by omitted variable bias (Cinelli and Hazlett, 2020). We regard facial traits to be both genetically and environmentally determined, and candidates' visual appearance on photos specifically to be a choice variable. Therefore, we cannot rule out that the facial trait predictions are proxies for deeper variables such as competence and charisma that explain the former's associations with electoral outcomes.

Overall, our findings are consistent with the prior literature, but expand the scope of inquiry substantially. Our machine learning approach allows other researchers to scale up future similar analyses quite easily. Furthermore, using sensitivity analysis, we cast some light on the causal interpretation of our findings.

We regard the fact that the inferred features correlate robustly with real-world outcomes as a successful external validation (Ghassemi *et al.*, 2021) of our CNN. This may come as a surprise given that there are serious doubts about the ability of CNNs to infer complex traits from images. For example, Torres and Cantú (2022, p.125) write that "abstract concepts or latent traits, such as the emotions that images trigger or evoke, offer a hard case" for machine learning. Therefore, our contribution should motivate scholars to use CNNs and other neural network architectures to further analyze subtle elements of image and potentially audio and video data (Rittmann *et al.*, 2020; Nyhuis *et al.*, 2021).

A related paper is Joo *et al.* (2015), who use a support vector machine trained on a custom dataset of human ratings of US politicians' images to predict election results. They obtain inconsistent results on the role of facial features across election types, possibly due to the combination of a small sample (N = 222–650) and their use of a model trained from scratch. Rasmussen *et al.* (2023), in independent work, use proprietary and open algorithms as well as a fine-tuned CNN to infer facial features of Danish politicians, which are then used to predict their (binary) ideology. Our approach is fully open-source, has richer outcome variables, and we replicate existing analyses from the political psychology literature, specifically also including the dominance trait.

We note that the features we are modeling are not inherent attributes of an individual's facial appearance. Instead, they are perceptions that may vary widely among different observers and cultural contexts, and to some degree reflect conscious choices of the candidates. Consequently, the predictions made by our model should be understood as approximations of societal attitudes and interpretations rather than definitive or universal truths about an individual's character or personality.

2. Facial features and political success

Social attributions based on facial traits are not only pervasive but also predictive of outcomes within a range of contexts such as in dating (Olivola *et al.*, 2014), prison sentencing (Blair

https://doi.org/10.1017/psrm.2024.38 Published online by Cambridge University Press

et al., 2004) as well as employment and career advancements (Fruhen *et al.*, 2015). Of key concern for our endeavor, the literature has found this to also be the case in politics when choosing among political candidates (Ballew and Todorov, 2007; Antonakis and Dalgas, 2009; Laustsen and Petersen, 2015, 2017; Casey, 2022). There is likely a deep biological mechanism behind such inferences. Parts of the visual cortex appear to have evolved to subserve the function of facial processing (Kanwisher *et al.*, 1997). Moreover, scholars have found a clear consistency in how adults and children assess the character of a person based on their facial appearances (Cogsdill, 2014) and that these assessments happen within milliseconds (Olivola and Todorov, 2010); one study even estimated the minimal time exposure after which people start discriminating between different categories of faces to be as little as 33 milliseconds (Bar *et al.*, 2006).

In general, we suggest to differentiate between two different causal mechanisms that would explain the association of facial traits with real-world political outcomes. First, upon seeing a picture of a candidate, voters may infer personality traits from it, and then base their decision on whether to vote for the politician on these inferred traits. This would constitute a causal effect of facial features on others' behavior.

Note that while Danish ballots do not embed photographs, Danish political candidates have exceptionally high usage rates of social media, contributing to their visibility.[1] Furthermore, even in the lower-stakes local elections, up to 60 percent of candidates print ads in newspapers and display campaign posters (Hansen and Hoff, 2010). However, there are no studies of the exposure of Danish citizens to candidates' visual appearances that quantify this visibility.

Second, traits inferred from pictures may represent underlying qualities that drive political success. For example, politicians' appearance of competence might stem from genuine competence, influenced by genetic common causes or deliberate investments into their appearances on an individual or party level. Accordingly, facial features would be proxies for factors of relevance for vote outcomes (Casey, 2022, p.710). Therefore, if we could somehow statistically adjust for the underlying common cause of appearance and outcome, the association between facial features and outcomes should vanish. We approach this to some extent by measuring demographic variables and by asking how strong, in quantitative terms, the omitted variables would need to be in order to explain away our findings Cinelli and Hazlett (2020).

What specific traits should be relevant for political outcomes? Sutherland *et al.* (2013) identify three dimensions, *attractiveness*, *trustworthiness*, and *dominance* from which social inferences about the character of person are made. In line with these dimensions, Berggren *et al.* (2010) demonstrate that a candidate's attractiveness correlates with their electoral success. Laustsen and Petersen (2017) indicate a heightened preference for leaders with dominant facial features during times of social conflict. Finally, Casey (2022) reveals that respondents can predict politicians' misuse of public funds from headshots, underscoring the close relationship to the idea of trustworthiness.

Hypothesis 1: Inferred trustworthiness, dominance, and attractiveness scores correlate positively with political success.

The role of attractiveness may differ sharply depending on the politicians' gender. On one hand, attractive female candidates may suffer from negative gender stereotyping (Sigelman *et al.*, 1987), which could lead to attractiveness having a stronger effect among male candidates. On the other hand, society may emphasize the importance of attractiveness to a greater extent among female candidates relative to male candidates (Carroll and Fox, 2018).

Hypothesis 2: The correlation of attractiveness scores and outcomes varies by candidate gender.

[1]In 2015, 68 percent of Danish parliamentary candidates had a Twitter account (Burkal *et al.*, 2019), while in 2021, only 4.5 percent of German candidates had one (Schmidt, 2021).

https://doi.org/10.1017/psrm.2024.38 Published online by Cambridge University Press

On the voter level, variations in social dominance orientation (SDO) are indicative of preferences for hierarchical structures. Given their preference for hierarchies, voters with higher SDO scores may employ dominant facial traits as heuristics to infer a politician's ability to facilitate such preferred hierarchies. This led Laustsen and Petersen (2016) to predict:

Hypothesis 3: The correlation of dominance scores and outcomes is more positive for candidates from right-wing parties.

Lastly, there may be differences in correlations across elections. In high salience elections, voters may increasingly be exposed to the facial cues of political candidates through increased media coverage and increased campaigning.

Hypothesis 4: Correlations of scores and outcomes are larger in the national than in the local election.

3. Data

We scraped pictures and information on ballot paper placement, number of personal votes as well as background information of the candidates that were electable at the Danish Local Election in 2021 and at the Danish General Election in 2022. Specifically, we utilize that the Danish public broadcasting networks offer pages to each candidate where they have the option of uploading a photo as well as listing additional information such as their birth date, gender, and educational background. Around 90 percent of candidates choose to upload a picture. We then merged this data with election data provided by Kasper Møller Hansen and Valdemar Østed. This data contained information about each candidate's ballot paper placement, incumbency, number of personal votes, and total number of personal votes cast for each party, allowing us to calculate an alternative measure of electoral success. For the local election, we used data from the Danish Election Authority (KMD, 2021).

For fine-tuning the CNN, we utilize the ratings obtained in the "One Million Impressions" dataset (Peterson *et al.*, 2022), in which just over 1000 faces were generated based on a database of 70,000 faces using generative adversarial networks. The synthetically generated faces were then rated by at least 30 participants per attribute. The raters mostly self-identified as White. Therefore, their ratings may be appropriate for investigating facial appearances in a country in Denmark, but possibly not in ethnically more heterogeneous countries. However, facial inferences appear to be quite similar across world regions (Jones *et al.*, 2021). We utilize the 880 pictures of adult faces.

We take the number of personal votes obtained by a candidate as our first outcome measure. The absolute number of votes differ greatly between the local and general elections and we therefore z-transform them within each election. After this, we take the natural log due to the heavy right skew and to facilitate interpretation of effect sizes. Our second outcome is the ballot position of a candidate, measured as a proportion of candidates ranked below them, adjusted to exclude the candidate's own position. We similarly z-transform this variable within elections. We discard a small number of candidates (25) who ran unopposed.

For capturing the ideology of political candidates, we use a dichotomous measurement (used by Laustsen and Petersen (2017)) based on party and *bloc* affiliation (i.e., either the liberal and left-leaning political block or the conservative and right leaning political bloc).[2] Gender is measured as a dichotomous variable. Age is measured as the age of the candidate at the time of election. The education variable is recoded to fit with the Danish ISCED levels.

[2]We treat the new party *Moderaterne* as a conservative party.

https://doi.org/10.1017/psrm.2024.38 Published online by Cambridge University Press

As the scraped candidate photos are captured at varying distances, we crop and center each photo to the candidate's face in order to improve their resemblance to the training images. We furthermore discard images not suitable for model inference. These include poor quality images, cartoon images and images without the candidate's face. In addition, some candidates have not uploaded an image. An overview of the collected and missing data is provided in the appendix. Overall, we use image data for 78 percent of all candidates. Candidates with a missing photo on average received very few votes. Therefore, we regard our sample as broadly representative.

4. Methods and results

To measure the facial traits, we fine-tune a pre-trained convolutional neural network. Previous applications of computer vision within social science have mostly relied on model architectures from the VGG family (Williams *et al.*, 2020; Rasmussen *et al.*, 2023) and the ResNet family (Rittmann *et al.*, 2020). Following this, we initially experimented with using pre-trained VGG-16 and ResNet-50 models as our convolutional base. Since we obtained the best initial results using ResNet-50, we chose this as our base, and fine-tuned it to predict the facial traits using the "One Million Impressions" dataset (Peterson *et al.*, 2022).

For training the algorithms, we fine-tune the following hyperparameters: Number of fully connected layers, number of parameters in each layer, batch size, dropout rate, lambda value for L2 regularization, and initial learning rate. Specifically, we divide our training data into a train, validation, and a holdout test set using a 60/20/20 split and a batch size of 16. As the ResNet-50 convolutional base has been trained on the Imagenet data, it takes as input zero-centered 224×224 images. We therefore resize all our images to 224×224, and zero-center each color channel. For each of the three algorithms, the final model consists of a ResNet-50 convolutional base, a fully connected layer with 4096 parameters, and a prediction layer with a linear activation function, as we are predicting continuous values. We use an ADAM optimizer with a mean absolute error loss function typically employed for regression tasks (Rittmann *et al.*, 2020; Rasmussen *et al.*, 2023). As the size of the training dataset is fairly small, to prevent overfitting, we employ data augmentation techniques where we randomly zoom in, adjust brightness and contrast, and horizontally flip the images. Additionally, we add early-stopping callback. Lastly, we decrease the learning rate by a factor of 10 when the accuracy does not improve over three epochs.

We evaluate each of the algorithms on the holdout test set, using mean average error (MAE) and Pearson's correlation coefficient as evaluation metrics. To establish a naïve baseline, we compare predictions to mean and random guessing (Rittmann *et al.*, 2020). The variances of each of the outcomes are similar (see appendix). Each of the models are able to predict the annotated scores in the test data with MAEs ranging from 0.576 to 0.732, or approximately 0.5 standard deviations. Pearson's correlation of annotated and predicted scores are relatively high, with values ranging from 0.75 to 0.79 (see Appendix). In each case, the algorithm outperforms the naïve baseline significantly. Figure 1 illustrates this. For each of the facial traits, the plots show a tight, almost perfectly linear relationship between the annotated and predicted scores, with only slight deviations at the tail ends, most likely due to the low occurrence of extreme values in the training data. In the Appendix, we further show that the algorithms are able to reproduce gender and age patterns in the training data. We also show correlations between predicted facial treats, which largely match correlations found in previous studies (Oosterhof and Todorov, 2008; Joo *et al.*, 2015; Peterson *et al.*, 2022).

We also fine-tuned our models with the Chicago Face Database (Ma *et al.*, 2015), but obtained significantly worse results (e.g., the predictive correlation for trustworthiness was 0.1). This highlights that the choice of fine-tuning data can be highly consequential. We furthermore detail descriptive differences between the two datasets in the Appendix. The "One Million Impressions" dataset is based on a sample that is predominantly White, in contrast to the

https://doi.org/10.1017/psrm.2024.38 Published online by Cambridge University Press

https://doi.org/10.1017/psrm.2024.38 Published online by Cambridge University Press

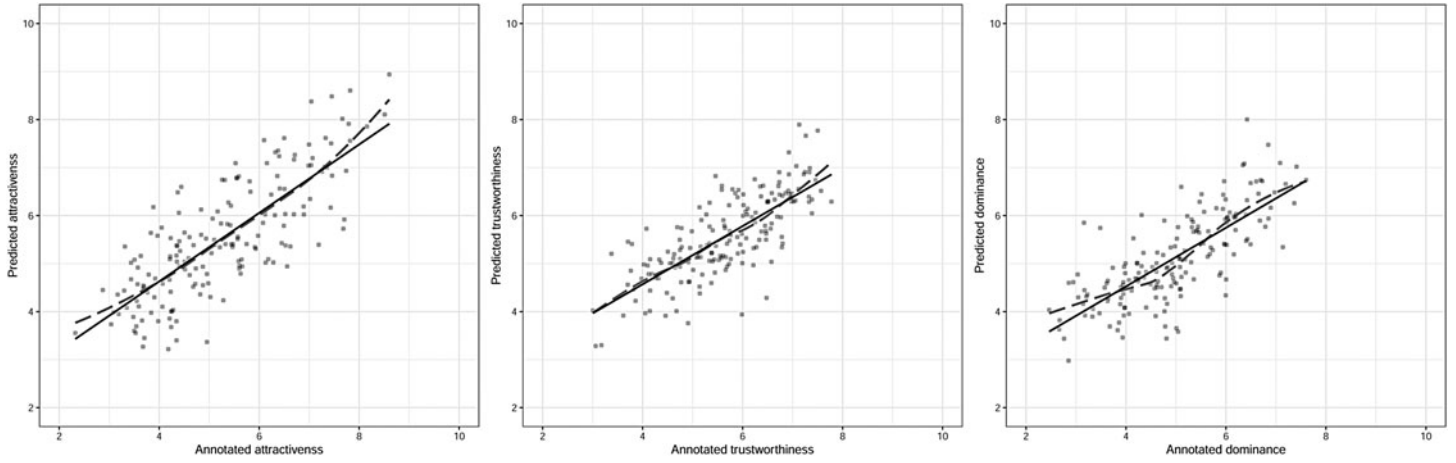


Figure 1. Out-of-sample evaluation of our fine-tuned model. Human-annotated scores are on the x-axis. Model predictions are on the y-axis. Solid line is a linear, dashed line a LOWESS fit. Left: Attractiveness. Middle: Trustworthiness. Right: Dominance.

more diverse Chicago Face Database. Therefore, it should *a priori* be more suited to our task that contains almost exclusively White Danish candidates.

After successful fine-tuning and evaluation, we applied the algorithm to predict the facial traits of the candidates. Finally, we correlated these trait scores with their personal votes and ballot paper placement. Figure 2 presents results on the main effects, featuring models that include and exclude age, education, and gender as covariates.

In line with Hypothesis 1, we see a positive effect of attractiveness and trustworthiness on both ballot paper placement and personal votes, whereas the effect of dominance is ambiguous.

Specifically, an increase of one standard deviation in attractiveness corresponds to an enhancement of 15 percent in personal votes with controls and 18 percent without controls. In a similar vein, a one standard deviation increase in trustworthiness leads to increases of 8 percent and 15 percent in personal votes with and without controls, respectively. While the effect of attractiveness on ballot paper placement remains stable after introducing controls, the effect of trustworthiness turns insignificant when including controls.

Regarding dominance, the results are highly sensitive to the particular model specification. For personal votes, the effect is positive as hypothesized, but when including controls, this turns insignificant. For ballot placement, the effect is consistently negative.

In the appendix, we show that these relationships are not artifacts of the linearity assumption, using kernel-regularized least squares regression that allows for nonlinearity (Hainmueller and Hazlett, 2014). Additional control for incumbency, which may be a post-treatment variable, does largely not alter results, even though incumbency correlates strongly with the outcomes. Finally, we also show that when using all three traits simultaneously to predict outcomes, the

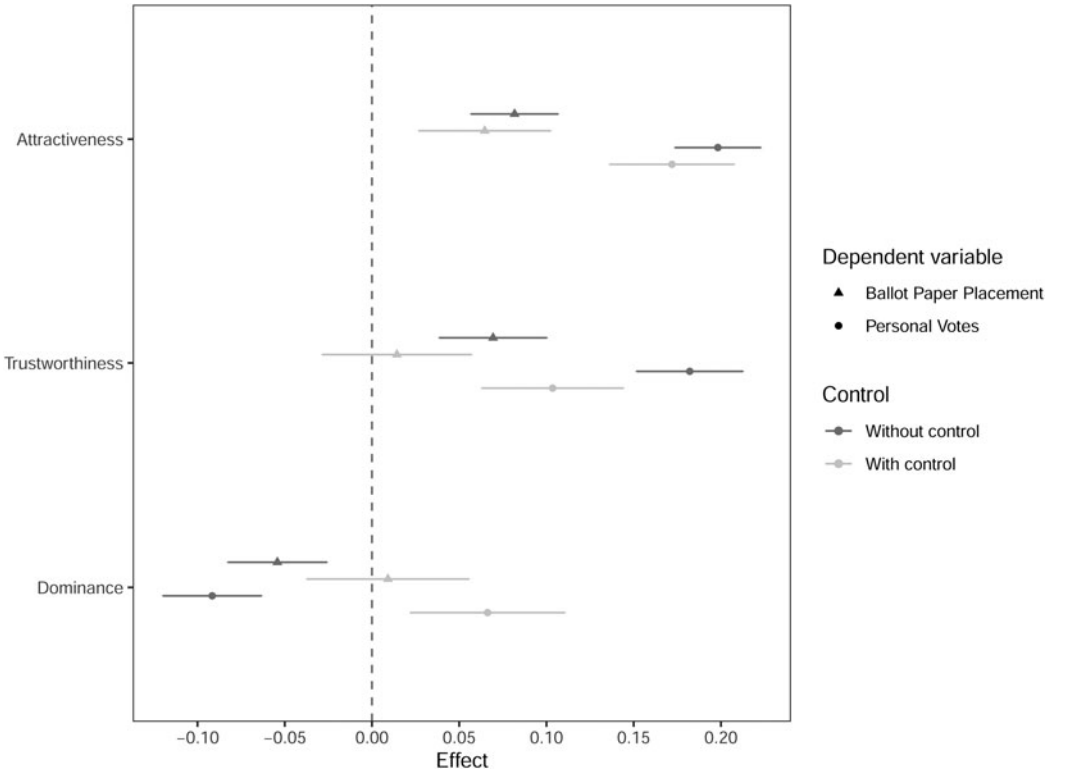


Figure 2. Main effects.

https://doi.org/10.1017/psrm.2024.38 Published online by Cambridge University Press

effect of attractiveness is robust, while the effects of trustworthiness and dominance become very small and insignificant. In the appendix, we show that the results are also robust to the inclusion of party fixed effects.

Figure 3 shows results on moderation by election type, gender, and ideology. In contrast to our hypotheses, the results only provide support for the effect of dominance being moderated by ideology, albeit the effect on ballot paper placement becomes insignificant with inclusion of controls. The fact that gender does not moderate the relationships may be due to the fact that the two theoretical mechanisms we described earlier (negative gender stereotyping versus positive discrimination of attractive females) vary across voters and cancel each other out.

In sum, we find overall support for Hypothesis 1, with the exception of dominance scores, where the findings are somewhat inconsistent. This largely replicates existing findings. In contrast to Hypothesis 2 and 4, we find no moderation by gender or election type. However, in accordance with Hypothesis 3, we find that dominance scores are more predictive of success for conservative candidates, replicating Laustsen and Petersen (2016). In exploratory analyses reported in the Appendix, we also investigate moderation by candidate age and ethnicity. We find that older candidates profit more from attractiveness and less from dominance. However, the interaction effects are very small. One explanation for this result could be that since older candidates are generally rated as less attractive but more dominant (see Appendix), the variance in attractiveness is more consequential for voter behavior. That is, an attractive older politician (or dominant young politician) is relatively unusual, and therefore triggers stronger inferences among voters. We find no interaction with ethnicity, but this may due to low statistical power, as almost all candidates are classified as White.

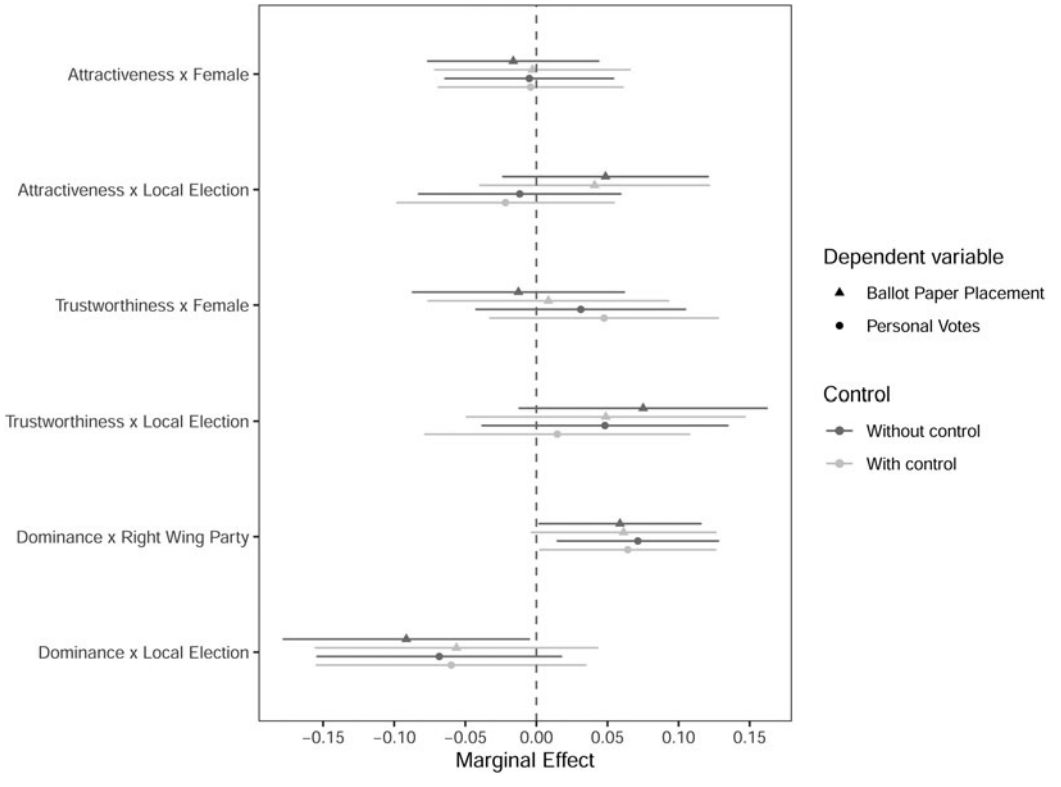


Figure 3. Interaction effects.

https://doi.org/10.1017/psrm.2024.38 Published online by Cambridge University Press

Finally, we ask: can our estimates be regarded as causal, or could they be plausibly explained away by omitted variables? We use the methodology developed by Cinelli and Hazlett (2020) to investigate this (see the appendix for details). We find that for all of our main estimates, an unobserved confounder as or twice as strong as the gender control variable (which is highly related to all of the facial traits, but only weakly to the outcome) could potentially push our estimates to statistical non-significance. More generally, confounder(s) that simultaneously explain about 2 to 4 percent of the variance in the facial trait and of the outcome would also explain away our results. Genetic common causes (Oskarsson *et al.*, 2018) and/or general political competence leading to more skillful visual self-presentation could plausibly constitute such confounders. If our estimates were much larger in size, this would increase the robustness on this metric and more plausibly suggest a causal role. However, with our research design and findings, we can not further elucidate this question.

5. Conclusion

The contribution of our research lies in using a CNN to investigate the previously proposed relationship between facial traits and political outcomes. Crucially, our application constitutes a case of "transfer learning," suggesting that the original image classification CNN learned general features of pictures. With minimal fine-tuning, these general features could be successfully transformed into predictions of human facial features as interpreted by humans. Despite reservations about the value of CNNs (Torres and Cantú, 2022), our findings should inspire scholars to embrace CNNs and other neural network architectures when using image, audio, and video data (Rittmann *et al.*, 2020; Nyhuis *et al.*, 2021). One working assumption for such an approach is that the human raters used for generating the original data behave similarly to the humans under study in the target data (in our case, Danish voters). The technical challenge then is to fit models flexibly enough to pick up on the image features that predict ratings, which can be evaluated empirically using test data.

Unlike previous studies that required human raters, our innovative use of CNN offers a scalable and reliable method for extracting perceived facial attributes. Furthermore, using CNN grants the advantage of impartiality; unlike human raters, the CNN remains uninfluenced by potential knowledge of the candidate whose face is assessed.

The congruence of our findings with established literature—especially the positive correlation of attractiveness and trustworthiness as well as the moderating effect of ideology in the case of dominance (Berggren *et al.*, 2010; Laustsen and Petersen, 2018; Casey, 2022)—underscores the efficacy of our machine-learning approach. While we have not ventured into "explainable AI" tools such as heatmaps to investigate our CNN, given concerns over performance and robustness (Ghassemi *et al.*, 2021), our overall methodology can be viewed as a successful external validation of our CNN. A shortcoming of our analysis is that the data used for fine-tuning do not contain competence ratings, which have been established to predict vote choice (Klofstad, 2017).

While our models offer precision assessing facial traits and effect sizes are both substantial and significant, we remain ambiguous about whether our results constitute causal effects or whether they partially derive from confounding common causes such as competence. The fact that both of these mechanisms are plausible may also explain the relatively large effects, as these can be thought of as the sum of causal effects and confounding. Finally, it is possible that the causal effects *per se* are sizable and homogeneous, insofar as the bio-psychological mechanisms explaining it are deeply ingrained and relatively stable across individuals, in contrast to more variable political preferences.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psrm.2024.38. To obtain replication material for this article, https://doi.org/10.7910/DVN/CH9AXM

Acknowledgements. We would like to thank participants at CEPDISC 2023 for valuable feedback.

https://doi.org/10.1017/psrm.2024.38 Published online by Cambridge University Press

Author contributions. AL: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – Original Draft, Writing – Review & Editing, Visualization. CH: Conceptualization, Methodology, Writing – Original Draft, Writing – Review & Editing, Formal Analysis, Literature Review. JS: Methodology, Writing – Original Draft, Writing – Review & Editing, Supervision, Project administration.

Competing interest. The authors declare none.

References

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

Ballew CC and Todorov A (2007) Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences* **104**, 17948–17953.

Bar M, Neta M and Linz H (2006) Very first impressions. *Emotion* **6**, 269–278.

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

Blair IV, Judd CM and Chapleau KM (2004) The influence of afrocentric facial features in criminal sentencing. *Psychological Science* **15**, 674–679.

Burkal R, Kæmsgaard Eberholst M and Mark B-O (2019) From hybrid media system to hybrid-media politicians: Danish politicians and their cross-media presence in the 2015 national election campaign. In Lilleker DG, Koc-Michalska K, Negrine R, Gibson R, Vedel T, Strudel S (eds), *Social Media Campaigning in Europe*. Abingdon, UK: Routledge, p. 14.

Carroll SJ and Fox RL (2018) *Gender and Elections: Shaping the Future of American Politics*, 4th Edn. Cambridge, UK: Cambridge University Press.

Casey K (2022) Snap judgments: predicting politician competence from photos. *The Journal of Politics* **84**, 698–713.

Cinelli C and Hazlett C (2020) Making sense of sensitivity: extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **82**, 39–67.

Cogsdill EJEA (2014) Inferring character from faces: a developmental study. *Psychological Science* **25**, 1132–1139.

Fruhen LS, Watkins CD and Jones BC (2015) Perceptions of facial dominance, trustworthiness and attractiveness predict managerial pay awards in experimental tasks. *The Leadership Quarterly* **26**, 1005–1016.

Ghassemi M, Oakden-Rayner L and Beam AL (2021) The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health* **3**, e745–e750.

Hainmueller J and Hazlett C (2014) Kernel regularized least squares: reducing misspecification bias with a flexible and interpretable machine learning approach. *Political Analysis* **22**, 143–168.

Hansen KM and Hoff J (2010) Den kommunale valgkampagne anno 2009. En kommenteret tabelrapport.

He K, Zhang X, Ren S and Sun J (2016) Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, pp. 770–778.

Jones BC, DeBruine LM, Flake JK, Liuzza MT, Antfolk J, Arinze NC, Ndukaihe IL, Bloxsom NG, Lewis SC and Foroni F (2021) To which world regions does the valence–dominance model of social perception apply?. *Nature Human Behaviour* **5**, 159–169.

Joo J, Steen FF and Zhu S-C (2015) Automated facial trait judgment and election outcome prediction: social dimensions of face. In *2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile: IEEE, pp. 3712–3720.

Kanwisher N, McDermott J and Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience* **17**, 4302–4311.

Kleppestø TH, Czajkowski NO, Vassend O, Røysamb E, Eftedal NH, Sheehy-Skeffington J, Kunst JR and Thomsen L (2019) Correlations between social dominance orientation and political attitudes reflect common genetic underpinnings. *Proceedings of the National Academy of Sciences* **116**, 17741–17746.

Klofstad CA (2017) Looks and sounds like a winner: perceptions of competence in candidates' faces and voices influences vote choice. *Journal of Experimental Political Science* **4**, 229–240.

KMD (2021) Kandidatliste for kommunalvalg 2021.

Laustsen L and Petersen MB (2015) Does a competent leader make a good friend? Conflict, ideology and the psychologies of friendship and followership. *Evolution and Human Behavior* **36**, 286–293.

Laustsen L and Petersen MB (2016) Winning faces vary by ideology: how nonverbal source cues influence election and communication success in politics. *Political Communication* **33**, 188–211.

Laustsen L and Petersen MB (2017) Perceived conflict and leader dominance: individual and contextual factors behind preferences for dominant leaders: perceived conflict and leader dominance. *Political Psychology* **38**, 1083–1101.

Laustsen L and Petersen MB (2018) When the party decides: the effects of facial competence and dominance on internal nominations of political candidates. *Evolutionary Psychology* **16**, 147470491773200.

Ma DS, Correll J and Wittenbrink B (2015) The Chicago face database: a free stimulus set of faces and norming data. *Behavior Research Methods* **47**, 1122–1135.

https://doi.org/10.1017/psrm.2024.38 Published online by Cambridge University Press

Nyhuis D, Ringwald T, Rittmann O, Gschwend T and Stiefelhagen R (2021) Automated video analysis for social science research. In Engel U, Quan-Haase A, Xun Liu S, Lyberg LE (eds), *Handbook of Computational Social Science*, vol. **2**. Abingdon, UK: Routledge, pp. 386–398.

Olivola C, Eastwick P, Finkel E, Ariely D and Todorov A (2014) A picture is worth a thousand inferences: appearance-based first impressions predict leader selection and mate choice. *NA – Advances in Consumer Research Volume* **39**, 276–277.

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

Oosterhof NN and Todorov A (2008) The functional basis of face evaluation. *Proceedings of the National Academy of Sciences* **105**, 11087–11092.

Oskarsson S, Dawes CT and Lindgren K-O (2018) It runs in the family. *Political Behavior* **40**, 883–908.

Peterson JC, Uddenberg S, Griffiths TL, Todorov A and Suchow JW (2022) Deep models of superficial face judgments. *Proceedings of the National Academy of Sciences* **119**, e2115228119.

Rasmussen SHR, Ludeke SG and Klemmensen R (2023) Using deep learning to predict ideology from facial photographs: expressions, beauty, and extra-facial information. *Scientific Reports* **13**, 5257.

Rittmann O, Ringwald T and Nyhuis D (2020) Pay attention to this! explaining emphasis in legislative speech using automated video analysis in the US house of representatives.

Schmidt J-H (2021) Facebook-und twitter-nutzung der kandidierenden zur bundestagswahl 2021. *Media Perspektiven* **12**, 639–653.

Sigelman L, Sigelman CK and Fowler C (1987) A bird of a different feather? an experimental investigation of physical attractiveness and the electability of female candidates. *Social Psychology Quarterly* **50**, 32.

Sutherland CA, Oldmeadow JA, Santos IM, Towler J, Michael Burt D and Young AW (2013) Social inferences from faces: ambient images generate a three-dimensional model. *Cognition* **127**, 105–118.

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

Todorov A, Olivola CY, Dotsch R and Mende-Siedlecki P (2015) Social attributions from faces: determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology* **66**, 519–545.

Torres M and Cantú F (2022) Learning to see: convolutional neural networks for the analysis of social science data. *Political Analysis* **30**, 113–131.

Williams NW, Casas A and Wilkerson JD (2020) *Images as Data for Social Science Research: An Introduction to Convolutional Neural Nets for Image Classification*, 1st Edn. Cambridge, UK: Cambridge University Press.

Cite this article: Lindholm A, Hjorth C and Schuessler J (2025) Facial finetuning: using pretrained image classification models to predict politicians' success. *Political Science Research and Methods* **13**, 1031–1041. https://doi.org/10.1017/psrm.2024.38

https://doi.org/10.1017/psrm.2024.38 Published online by Cambridge University Press