Gender differences in face-based trait perception and social decision making

Zoe W. He (wah016@ucsd.edu)
Department of Cognitive Science, University of California, San Diego
La Jolla, CA 92093 USA

Angela J. Yu (ajyu@ucsd.edu)
Department of Cognitive Science & Halıcıoglu Data Science Institute, University of California, San Diego
La Jolla, CA 92093 USA

Abstract

Despite recent progress in promoting gender equality, gender bias continues to post challenges to women's career advancement. Here, we use a statistically grounded framework to investigate how face-based social perception may contribute to gender biases in political and job application settings. By analyzing a large face dataset and performing a novel behavioral experiment, we find that: 1) female faces exhibit stronger anti-correlation between perceived trustworthiness and dominance, 2) this anti-correlation is due to distinct sets of facial features humans utilize to assess female and male faces for trustworthiness and dominance, 3) perceived dominance positively contributes to social decision preferences for female faces, contrary to prior suggestions that perceived dominance affects female candidates negatively, and 4) the anti-correlated perception of trustworthiness and dominance put females at a disadvantage in competitive environments. More generally, our findings reveal the important role of face-based trait perceptions underlying gender biases in social decision making.

Keywords: face perception, facial features, gender, dominance, trustworthiness, social decision making

Introduction

The year of 2021 marks a significant year for women in politics in the U.S., with the first female Vice President elected and at least 142 women holding seats in the current Congress. Despite an increase in the number of female politicians in the office, gender bias continues to pose challenges to women's career advancement in politics and other workplace settings. Women are still underrepresented in leadership positions and facing higher obstacles than men. Female candidates are often held to a different standard from male candidates: they have been reported to receive systematically more negative media coverage than their male counterparts (Meco, 2019), and are considered as less suitable despite similar credentials (Paul & Smith, 2008).

Interestingly, gender differences appear to reflect subjective biases rather than factual differences. Reports show that in the 2016 U.S. presidential election, Democratic nominee Hillary Clinton's was most criticized for being "too ambitious", "overly dominant", and more "untrustworthy" and "deceitful" than her male opponent, Donald Trump, despite a lack of factual evidence of their honesty levels (Foran, 2016). A number of psychological studies suggests that only female candidates are penalized when perceived as dominant or power-seeking (Williams & Tiedens, 2016; Okimoto & Brescoll, 2010), a social dilemma known as the "backlash effects" (Rudman, Moss-Racusin, Phelan, & Nauts, 2012) for females who violate prescriptive stereotypes of femininity. While a number of studies have examined social impressions and perceptual judgments as important factors underlying the gender bias in politics and workplace, few have tried to explain the phenomenon by considering the explicit role of face perception (Oh, Buck, & Todorov, 2019), a natural process humans use to form impressions about strangers since ancient times (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015).

Humans form rapid social impressions about other people's characters, such as trustworthiness, dominance, attractiveness, and competence, based on face appearance within as short as 100 milliseconds of exposure (Willis & Todorov, 2006). Although the accuracy of face-based social trait impressions is hotly debated (Todorov et al., 2015), humans heavily rely on these superficial judgments in many important social decisions (Todorov et al., 2015). For example, in politics and business, perceived competence, dominance, sociability, and trustworthiness are shown to be reliable predictors for an individual's electoral and employment success (Todorov et al., 2015; Olivola, Funk, & Todorov, 2014; Olivola & Todorov, 2010). It is also found that trustworthy-looking faces are more likely to attract investments and are less likely to be convicted on criminal trials (Olivola et al., 2014). Altogether, the social significance of face-based perception has the potential to provide novel insights into the nature of gender biases in social decisions.

Previous research has looked at the gender differentiation in perceived trustworthiness and dominance and how they affect the general likability of a face (Sutherland, Young, Mootz, & Oldmeadow, 2015). Trustworthiness and dominance facial dimensions are believed to be parallel to the warmth and competence dimensions, which are the basis for gender stereotyping (Cuddy, Fiske, & Glick, 2008). It has been shown that dominant-looking female faces are less preferred than both the less dominant female faces and the dominant-looking male faces (Cuddy et al., 2008). Furthermore, trustworthy-looking male faces are evaluated more positively than trustworthy-looking female faces. While this perceptual bias against female faces is consistent with the "backlash effects" towards counter-stereotypical women in the political scenario studies, a formal explanation of why only females suffer from violating stereotypical traits in important social scenarios is needed.

In this paper, we assess the nature of gender biases in pol-

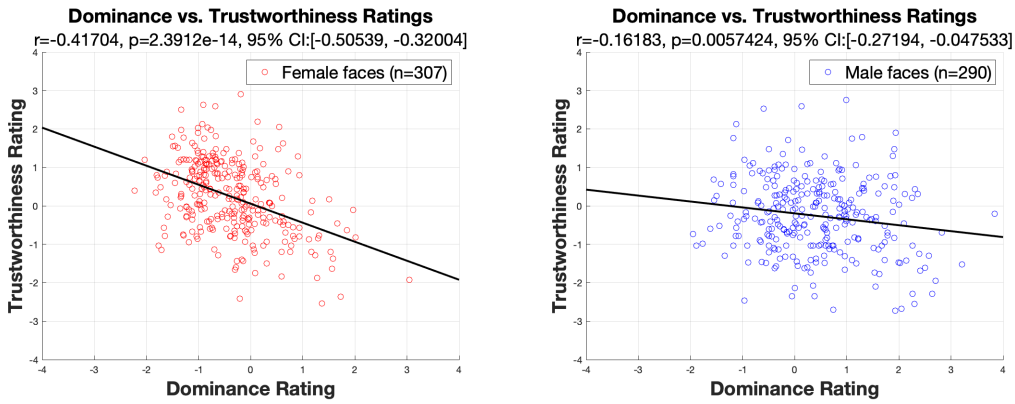

Figure 1: Correlations between dominance and trustworthiness ratings for 307 female faces (left) and 290 male faces (right) from the Chicago Face Database (CFD).

itics and competitive workplace, by investigating which parts of the process of social perception of faces are responsible for creating the discrepancy. In particular, we use a formal computational model to examine how social perception of faces may contribute to gender biases in social decision making. Here, we take a statistical approach to model faces as points in a vector space, the "face space", using a computer vision algorithm known as Active Appearance Model (Cootes, Edwards, & Taylor, 2001) (AAM, see Methods), and we use a regression model to find the linear combination of facial features that best predict human judgments for individual social traits (Guan, Ryali, & Yu, 2018). Assuming that faced-based judgment of both trustworthiness and dominance are contributing factors for relative preference in political and workplace scenarios, we propose that the overall likability of a face is linearly dependent on the predictive perceived trustworthiness and dominance of the face, with two coefficients for each term. The predictive trustworthiness and dominance are assumed to each be a linear transformation of the face vector. Based on this framework, we propose two hypotheses at the level of face processing to explain why only preferences for female faces suffer from an increased perception of dominance.

The first hypothesis (H1) assumes that both trustworthiness and dominance contribute positively to political and job preference for female faces, but female faces are less likely to be perceived as both trustworthy and dominant compared to male faces. Under this hypothesis, two competing sub-hypotheses are proposed to explain female disadvantages.

The first sub-hypothesis (H1a) assumes that humans use a common criteria for judging dominance and trustworthiness in both male and female faces, but this process somehow creates a representation that is naturally disadvantageous for female faces, because e.g. the female face distribution is such that it is less likely to be assessed as both trustworthy and dominant, compared to the male face distribution.

In contrast, the second sub-hypothesis (H1b) states that humans use different criteria (i.e. different sets of facial features) for male and female faces when making inferences about the two social traits. Formally, this means that the female faces have different face-to-trait mapping from male faces, in such a way that only female face projections are less likely to have high ratings for both traits (i.e. the social asssessment *criteria* are different from the two genders). An implication is that face-gender is immediately processed and then primes the subsequent face processing for social perception.

A second hypothesis (H2) is that perceived trustworthiness and dominance have opposite contributions to face preference for female faces; while trustworthiness is a positive factor for female faces in election and job hiring process, dominance contributes negatively for female faces in the same scenarios. As a result, dominant-looking female faces are less preferred. This hypothesis is consistent with suggestions in the existing literature about the *backlash effect*, whereby perceived assertiveness and power seeking in females cause moral outrage, than such perceived traits in males, because of the violation of gender stereotypes (Rudman et al., 2012).

We investigate these hypotheses using the proposed statistical models with two publicly available datasets containing face images and human ratings on various social traits (Bainbridge, Isola, & Oliva, 2013; Ma, Correll, & Wittenbrink, 2015). Our analysis indicates that (1) female faces exhibit significantly stronger negative correlation between perceived trustworthiness and dominance compared to the male faces (supporting H1), (2) distinct criteria (sets of facial features) are used for assessing trustworthiness and dominance for female and male faces (supporting H1b), (3) perceived dominance is a positive contributing factor for female faces (contrary to H2), and (4) the trait perception anti-correlation in female faces put them at a particular disadvantage, compared to males, in especially competitive electoral and job application settings (novel prediction). Altogether, our results support hypothesis 1b as the best explanation, and point out several directions for fruitful future inquiries.

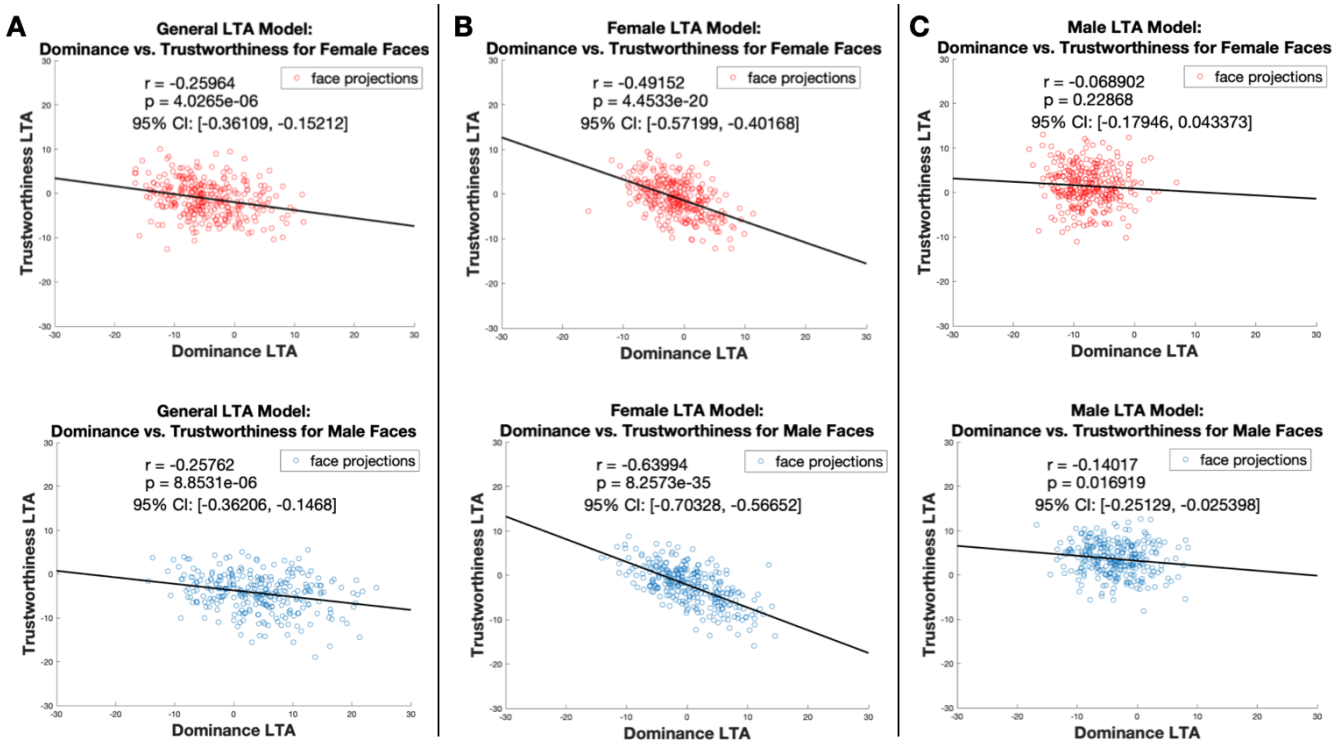

Figure 2: Projections of female (red) and male (blue) CFD faces onto the dominance-trustworthiness subspace from (A) a general gender-neutral model, (B) a model trained using female faces only, and (C) a model trained using male faces only. LTA (Linear Trait Axis): the linear combination of facial features for the target trait.

Results

Face Gender and Social Traits

To examine whether social trait judgments of faces differ by gender, we utilize the Chicago Face Database (CFD) (Ma et al., 2015), a publicly available dataset. It contains 597 demographically balanced face images (307 female, 290 male) with subjective ratings for nine social traits including trustworthiness and dominance. We find that female faces exhibit a highly negative correlation between dominance and trustworthiness ratings (Figure 1 left, $r = -0.42, p < 0.001$), while male faces elicit only mild negative correlation between dominance and trustworthiness ratings (Figure 1 right, $r = -0.16, p = 0.005$). This result indicates that it is less likely (due to the high negative correlation) to find a female face that is perceived as both highly dominant and trustworthy. Therefore, if both perceived dominance and perceived trustworthiness contribute positively to "likability" in election and hiring processes, then females would be naturally at a disadvantage.

The finding so far is consistent with both hypothesis 1a and 1b. The two sub-hypotheses primarily differ in terms of whether there is a common mechanism to judge trustworthiness and dominance in male versus female faces. To compare the two sub-hypotheses, we use a model that maps face images to trait perception (Guan et al., 2018; Ryali, Wang, & Yu, 2020). We train linear regression models of trustworthiness and dominance as a function of AAM features (face image featuers). The models are trained with three training datasets: faces from both genders, female faces only, and male faces only. The female and male faces are then projected into each of the three dominance-trustworthiness subspaces (features to traits), as illustrated in Figure 2. Each projection indicates the predicted dominance and trustworthiness ratings for each face. The common linear models (Figure 2A, train on both male and female faces) predict a similar amount of negative correlation for female and male faces ($r = -0.26, p < 0.001$). It fails to capture the significant gender difference present in the actual trait ratings and shows insufficiency in predicting the correlation coefficient in either gender. The gender-specific models (Figure 2B & 2C, trained only on female faces or only on male faces), on the other hand, make more accurate predictions for the respective genders: the female linear models predict stronger anti-correlations for both genders, and the male linear models predict weaker anti-correlations for both genders. These results suggest that hypothesis 1b is a better explanation than hypothesis 1a: humans use different criteria for different genders for judging dominance and trustworthiness. This also implies that gender information is rapidly processed and then influences additional, finer-grained face-based processing. This is consistent with previous MEG studies showing

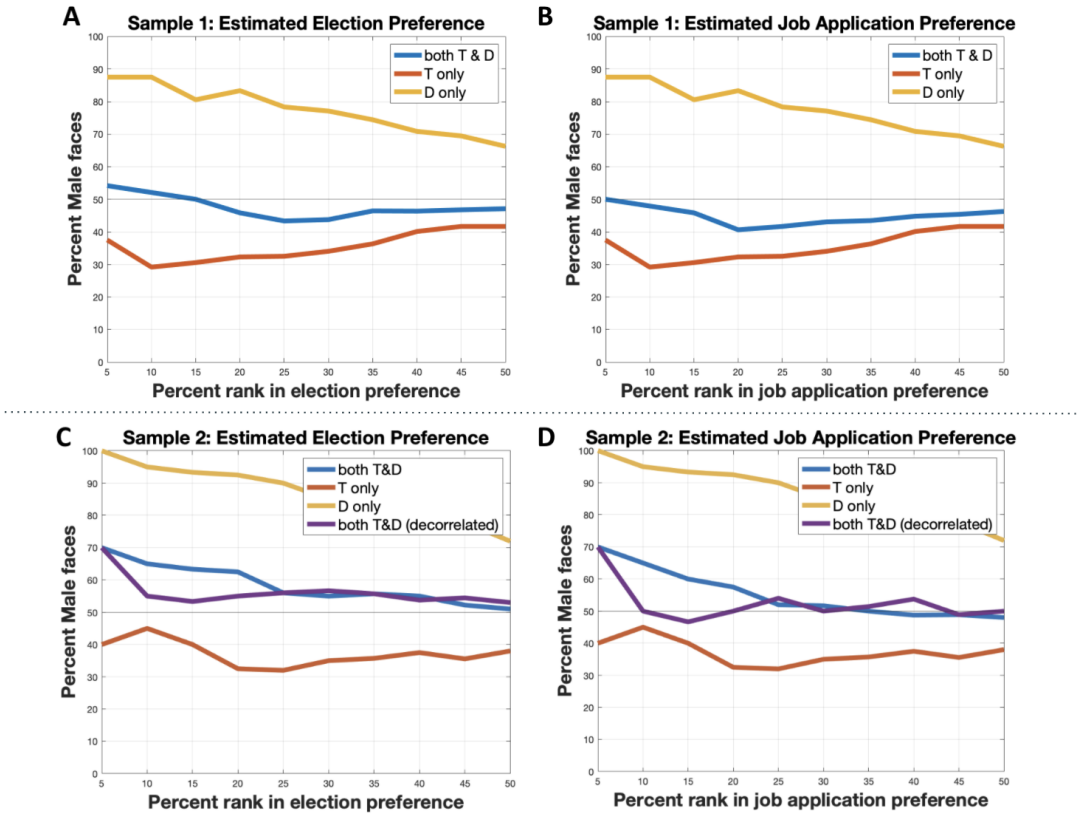

Figure 3: Predicted election and job application preference for CFD faces using the regression coefficients estimated from the experiment. (A)-(B): Predictions for 240 female and 240 male faces sampled from all CFD faces. (C)-(D): Predictions for 100 female and 100 male faces sampled from above-average trustworthiness ratings. T&D: Prediction using both trustworthiness and dominance ratings. T: Prediction using trustworthiness ratings only. D: Prediction using dominance ratings only.

that gender-based information is available in the neural signal much sooner than other face-based information, such as identity and familiarity (Dobs, Isik, Pantazis, & Kanwisher, 2019).

Trait Perception and Social Decision Making

Next, we investigate the contributions of perceived trustworthiness and perceived dominance to preferences in election and job application scenarios. We collect human ratings of social traits (including dominance and trustworthiness) and preferences in social scenarios (including election and job application scenarios) in an experiment, using 36 pairs of female faces (see Methods) selected from the Bainbridge dataset (Bainbridge et al., 2013). We use step-wise regression to predict election and job application preferences based on perceived trustworthiness and dominance traits. We find that the best fitting model is multiple linear regression including both trustworthiness and dominance. The estimated weights for election preference are 0.678 ($p < 0.001$) for trustworthiness and 0.214 ($p < 0.01$) for dominance, and the weights for job application preference are 0.662 ($p < 0.001$) for trustworthiness and 0.170 ($p = 0.03$) for dominance. Therefore, we conclude that perceived trustworthiness and dominance are both positive factors for female candidates in election and job application. This finding is inconsistent with hypothesis 2 that the dominance trait itself contributes negatively to social preference for female faces.

Predicting gender biases in political and job preference

So far, we find a much stronger negative correlation between perceived trustworthiness and dominance for female faces, than for male faces. Our results indicate this difference arises from using different criteria to judge these traits in male and female faces. We also find that both trustworthiness and dominance contribute positively to preference for female candidates in both election and job application scenarios.

Next, we investigate whether the combination of these two properties are sufficient to yield an overall advantage for male candidates over female candidates using model simulations. With the assumption that the relative contributions of dominance and trustworthiness are consistent across genders, we use the estimated weights from the multiple regression models (in last section) to predict preference ratings for two ran-

dom samples of female and male faces from CFD. The first sample contains 240 random faces for each gender, while the the second sample contains 100 random faces that are sampled from faces that have above-average trustworthiness ratings for each gender. The predictive preference of those two samples are estimated using both trustworthiness and dominance ratings, using trustworthiness only, and using dominance only.

Figure 3 shows the proportion of male faces in the top 5% to top 50% of all faces for election and job application preference. When the preference ratings are estimated using only dominance (yellow curves) or only trustworthiness (red curves), there is a consistent male advantage (above the 50% threshold) or female advantage (below the 50% threshold), respectively, in both samples for both election and job application scenarios. However, when the preference ratings are estimated using both the trustworthiness and dominance traits (blue curves in Figure 3), a male advantage occurs only in the above-average trustworthiness sample (Figure 3B), with 70% of the 5% most preferred being male faces in both the election and the job application scenario. This implies our current model predicts the male advantage primarily emerges in situations where both male and female candidates are perceived to be trustworthy enough and the social environment is highly competitive (e.g. only selecting the top 5%-10%).

Besides the anti-correlation between trustworthiness and dominance, it is also possible that male faces benefit from baseline differences in trustworthiness and dominance ratings for female and male face. To examine this, we re-predict the political and job preference for both genders with decorrelated trustworthiness and dominance ratings (by randomly shuffling the pairing of trustworthiness and dominance ratings for each gender), and compare the shuffled (decorrelated) predictions against the previous predictions (see purple curves in Figure 3C-D). We find that the male advantage greatly diminishes in the above-average trustworthiness sample, indicating that this male advantage in election and job preference is primarily due to the negative correlation between trustworthiness and dominance, although baseline differences in male and female face ratings for the two traits also appear to play a minor role.

Discussion

In this paper, we proposed a statistically formulated theory for election and job application preferences, using perceived trustworthiness and dominance traits for face images. Using this framework, we examined various face-perception related factors underlying gender biases in political and job hiring preferences. We investigated the sources of gender biases in face-based social perception at two different levels: one level concerns the sets of facial features recruited for forming social trait judgments for trustworthiness and dominance, and the other level concerns the contribution of the perceived trustworthiness and dominance to the overall preference for a face in election and job application decisions. First, we found a strong negative correlation between dominance and trustworthiness ratings for female faces but not for male faces, and we showed that this gender-specific discrepancy cannot be captured by gender-neutral regression models based on dominance and trustworthiness ratings. On the other hand, gender-specific regression models on social traits were able to capture the dominance-trustworthiness correlations for their corresponding face gender. This result provides an important insight into why female faces are less likely to be perceived as highly dominant and trustworthy at the same time, in comparison to male faces: people are using different *criteria* to evaluate the same trait for male and female faces.

Second, we showed that both dominance and trustworthiness are positive contributors to preference in election and job application scenarios. This result implies that the "backlash effects" females experience may not be simply due to the direct effect of dominance, but instead is a combination of both dominance and trustworthiness. Females are deemed less trustworthy when perceived to be dominant because of the interaction effect between these two traits. In the case of election and job hiring, where both dominance and trustworthiness are important traits looked for in the candidates, dominant, power-seeking, and self-promoting women tend to also be perceived as less trustworthy women, while trustworthy-looking women are not perceived to be dominant enough for the high level leadership positions. Male candidates do not experience such dilemma since their faces are judged based on a different set of facial features that do not compromise between the perceived dominance and trustworthiness. This male advantage, based on our simulations, should be most apparent in highly competitive scenarios, where the candidates have to be already "good enough."

Our findings about the role of gender information on subsequent face perception is also consistent with the neuroscience literature about the temporal dynamics of face processing, where gender information has been found to be available earlier than other types of information (Dobs et al., 2019). One limitation of our study is that, with the available data, we cannot estimate all the parameters in our model (Equation 1) for both genders. For example, in addition to how trustworthiness and dominance are assessed differently between the genders, how the two respectively contribute to social decisions may also be different between genders, in such a way so as to exacerbate or perhaps mitigate gender biases in ultimate social decisions. Thus, future studies, including collecting data on male faces, is needed to have a more complete understanding of the connections between perceptual differentiation in face genders and gender bias. This direction of research will be an interesting addition to the understanding of the functional and social importance of gender information prioritization in perception, and might also help to develop novel solutions to combat gender biases in real-life scenarios.

Methods

Dataset

We use two publicly available face datasets: Chicago Face Database (CFD) (Ma et al., 2015) and Bainbridge (Bainbridge et al., 2013). CFD comprise 109 East Asian (57 female), 197 black (104 female), 108 Hispanic (56 female) and 181 white (90 female) faces. Bainbridge consists of 2222 US adult faces that are collected from Google Images.

Experimental Design

We recruited 613 undergraduate students (331 female) participated for a social decision-making study. We use 72 white female faces with direct gaze and natural expression from Bainbridge et al., 2013. As part of the study, 36 pairs of female faces were rated on six social traits on a scale from 1-9 within 2s, including dominance and trustworthiness. After that, for five social scenarios, including election and job interview, the participants respond with a preference between the two faces displayed on the screen, on the scale of 1-9 (1 indicates a preference for the face on the left and 9 for the face on the right), to . The questions are: a) Suppose you represent a company at a job fair, and two individuals approached you to discuss job openings. Which of them would you be more willing to talk to? b) Suppose two people are candidates for a state-wide election. Which of them are you more likely to vote for?

Model Descriptions

Active Appearance Model (AAM) AAM is a well-establish computer vision model and appears to have neural relevance (Chang & Tsao, 2017). It consists of *shape* and *texture* features. The shape features are the coordinates of a set of pre-defined landmarks. The texture features are the pixel values of each face image after warping it to the average landmark. We train AAM with faces from CFD and Bainbridge. We perform additional PCA on shape and texture features and retain the first 60 PCs as the vector representation in a 60-dimensional "face space".

Linear Trait Axis (LTA) The Linear Trait Axis (LTA) $\tilde{\beta}$ (Guan et al., 2018; Huang et al., 2018) for each social trait is computed as the normalized regression coefficients of ratings regressed against AAM featuers: $y = \beta\vec{x} + \varepsilon$, where y is the standardized ratings for the trait, $\vec{x}$ is the AAM features, and β is the vector of regression coefficients. The LDA is defined as $\tilde{\beta} = \frac{\beta}{||\beta||}$. The LDA specifies a direction in the face space that (linearly) maximally alter the response of the task.

Predictive Model for Social Traits We assume the following linear models for feature-based social trait judgments:

$$O_i = a(X_i \cdot T_g) + b(X_i \cdot D_g) + \varepsilon_i \quad (1)$$

where O_i is the overall likability of the i-th face, X_i is the vector representation of the i-th face in the feature space, and T_g and D_g are linear models for trustworthiness and dominance of gender g respectively. The scalars a and b are coefficients (model parameters). Essentially, the preference of a face in politics and workplace can be characterized by a linear combination of perceived trustworthiness and the perceived dominance, plus Gaussian noise ε_i.

Acknowledgement

We thank Jingya Huang, Severine Soltani, and Aya Konishi for collecting and sharing the data; Christine Lind for assistance with the AAM model; Dalin Guo and Corey Yishan Zhou for helpful feedback on the paper and proofreading.

References

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of experimental psychology. General*, *142 4*, 1323-34.

Chang, L., & Tsao, D. Y. (2017). The code for facial identity in the primate brain. *Cell*, *169*(6), 1013–1028.

Cootes, T., Edwards, G., & Taylor, C. (2001). Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.*, *23*, 681-685.

Cuddy, A. J. C., Fiske, S., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the bias map. *Advances in Experimental Social Psychology*, *40*, 61-149.

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over time. *Nature Communications*, *10*(1), 1–10. Retrieved from http://dx.doi.org/10.1038/s41467-019-09239-1 doi: 10.1038/s41467-019-09239-1

Foran, C. (2016, Sept 17). The curse of hillary clinton's ambition. *The Atlantic*. Retrieved from https://www.theatlantic.com/

Guan, J., Ryali, C., & Yu, A. J. (2018). Computational modeling of social face perception in humans: Leveraging the active appearance model. *bioRxiv*(2), 360776. Retrieved from https://www.biorxiv.org/content/10.1101/360776v1 doi: 10.1101/360776

Huang, S. J., Ryali, C. K., Liu, J., Guo, D., Guan, J., Li, Y., & Yu, A. J. (2018). A Model-Based Investigation of the Biological Origin of Human Social Perception of Faces. Retrieved from https://www.faceplusplus.com

Ma, D., Correll, J., & Wittenbrink, B. (2015, 01). The chicago face database: A free stimulus set of faces and norming data. *Behavior research methods*, *47*. doi: 10.3758/s13428-014-0532-5

Meco, L. D. (2019, Dec 17). *Missing from the Conversations about Tech and Elections? Women.* Retrieved from https://www.genderontheballot.org/

Oh, D. W., Buck, E. A., & Todorov, A. (2019). Revealing Hidden Gender Biases in Competence Impressions of Faces. *Psychological Science*, *30*(1), 65–79. doi: 10.1177/0956797618813092

Okimoto, T. G., & Brescoll, V. L. (2010). The price of power: Power seeking and backlash against female politicians. *Personality and Social Psychology Bulletin*, *36*(7), 923–936. doi: 10.1177/0146167210371949

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences*, *18*(11), 566–570. Retrieved from http://dx.doi.org/10.1016/j.tics.2014.09.007 doi: 10.1016/j.tics.2014.09.007

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior*, *34*(2), 83–110. doi: 10.1007/s10919-009-0082-1

Paul, D., & Smith, J. L. (2008). Subtle sexism? examining vote preferences when women run against men for the presidency. *Journal of Women, Politics and Policy*, *29*(4), 451–476. doi: 10.1080/15544770802092576

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, *48*(1), 165–179. Retrieved from http://dx.doi.org/10.1016/j.jesp.2011.10.008 doi: 10.1016/j.jesp.2011.10.008

Ryali, C., Wang, X., & Yu, A. J. (2020). Leveraging computer vision face representation to understand human face representation. *Proceedings of the 42th Cognitive Science Society Conference*.

Sutherland, C. A., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2015). Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *British Journal of Psychology*, *106*(2), 186–208. doi: 10.1111/bjop.12085

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*, 519–545. doi: 10.1146/annurev-psych-113011-143831

Williams, M. J., & Tiedens, L. Z. (2016). The subtle suspension of backlash: A meta-analysis of penalties for women's implicit and explicit dominance behavior. *Psychological Bulletin*, *142*(2), 165.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological science*, *17*(7), 592–598.