Automated Facial Trait Judgment and Election Outcome Prediction: Social Dimensions of Face

Jungseock Joo[1,3], Francis F. Steen[2], and Song-Chun Zhu[1]

[1]Departments of Computer Science and Statistics, UCLA
[2]Department of Communication Studies, UCLA
[3]Facebook

jungseock@gmail.com

Abstract

*The human face is a primary medium of human communication and a prominent source of information used to infer various attributes. In this paper, we study a fully automated system that can infer the perceived traits of a person from his face – **social dimensions**, such as "intelligence," "honesty," and "competence" – and how those traits can be used to predict the outcomes of real-world social events that involve long-term commitments, such as political elections, job hires, and marriage engagements. To this end, we propose a hierarchical model for enduring traits inferred from faces, incorporating high-level perceptions and intermediate-level attributes.*

*We show that our trained model can successfully classify the outcomes of two important political events, only using the photographs of politicians' faces. Firstly, it classifies the winners of a series of recent U.S. elections with the accuracy of **67.9% (Governors) and 65.5% (Senators)**. We also reveal that the different political offices require different types of preferred traits. Secondly, our model can categorize the political party affiliations of politicians,* i.e.*, Democrats vs. Republicans, with the accuracy of **62.6% (male) and 60.1% (female)**. To the best of our knowledge, our paper is the first to use automated visual trait analysis to predict the outcomes of real-world social events. This approach is more scalable and objective than the prior behavioral studies, and opens for a range of new applications.*

1. Introduction

1.1. Trait Judgment from Faces

The human face is a highly salient medium of human communication, carrying surprisingly rich information about the subject, including gender [12], ethnicity [14], age [31], emotional state [11, 10], and identity [38, 22], each of which has been studied intensively both in human perception and computer vision. Furthermore, research in social psychology suggests that people also make inference about persistent **social traits** such as "trustworthy" or "dominant" from facial appearance [5, 25]. Evaluating others along these dimensions is a routine activity in our daily lives: when we choose a spouse, hire a new employee, or select **political leaders** to make crucial decisions on our behalf. An extensive literature indicates that facial appearance can have a significant impact on social trait judgment [32, 34, 37].

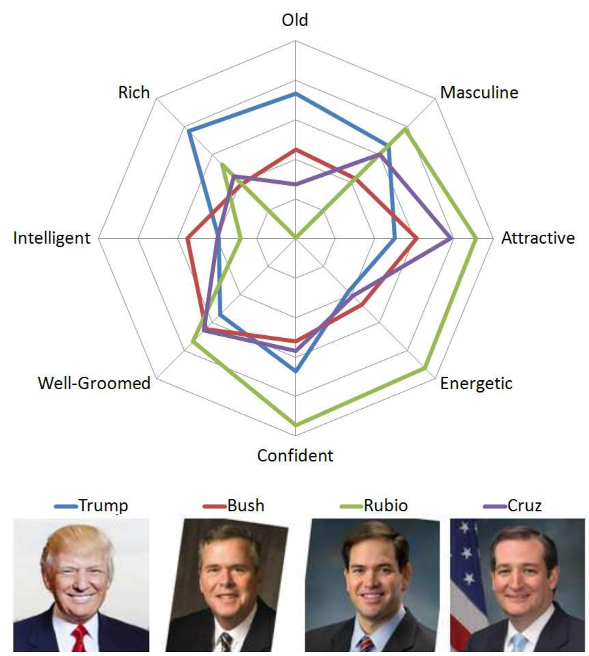

Figure 1. The inferred social dimensions of faces of a few Republican politicians who may run for 2016 presidency, predicted by our learned model.

In this paper, we study the cognitive procedure of social

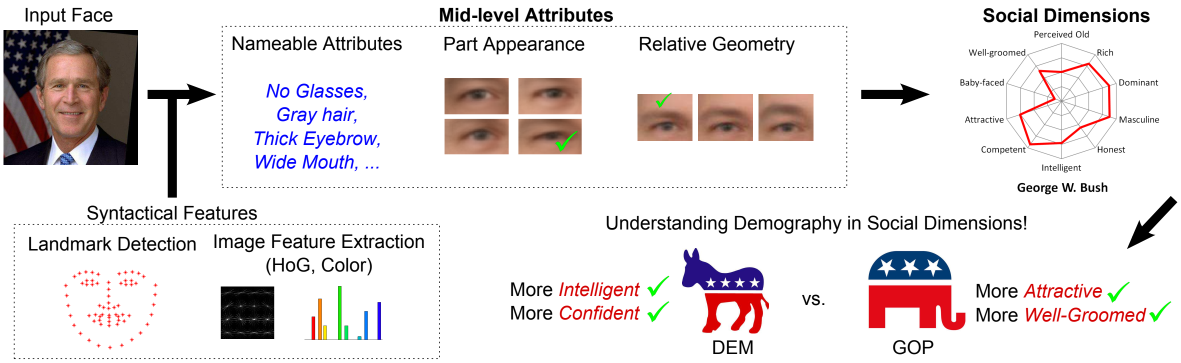

Figure 2. The overview of our model to predict the social dimensions of face. Given an image of a face, we first detect the facial landmarks and extract the low-level features from corresponding facial regions. These features are used to recognize a set of mid-level attributes: the nameable attributes (*e.g.*, gray-haired), the part appearance types, and their relative geometries. These attributes are collectively used to finally infer the social dimensions.

trait judgment from facial appearance and verify its effect on the outcome of real-world social events in a *computational* framework. Understanding the social dimensions of a face is an active research area in the social sciences, and facial appearance and perceived traits have been shown to predict the actual outcomes of real-world events such as elections [32] and criminal sentencing [6]. Prior studies in social science, however, are limited in scalability and consistency, as they are based on human experiments. Behavioral studies generalize poorly; for instance, trait annotations of public icons such as U.S. presidents or high-profile politicians are unreliable because of prior familiarity and personal preferences affect the annotators (Fig. 1). By extending and refining computer vision research in social trait judgment, we can capture an important dimension of human cognition and make a key methodological contribution to social science research. The contributions of our paper can be summarized as follows:

1. We propose a hierarchical model to predict the personal traits from a facial photograph, which incorporates high-level perceptions, intermediate-level attributes, and low-level features.

2. We introduce a novel dataset of faces labeled with 14 social dimensions. Our dataset contains 650 images of real-world politicians (491 male and 159 female subjects).

3. We show that our trained models can be used to analyze and predict the outcomes of important social events (*e.g.*, elections).

1.2. Related Works

Social Attributes of Human Faces. The social dimensions of face can be viewed as a special type of *facial attribute* that has been studied intensively in computer vision literature. These attributes typically represent categorical and fine-grained information about human subjects, such as gender, ethnicity, age group, or emotional states. Recently, there have been also interesting reports to leverage these traditional visual attributes to achieve higher-level goals – to infer *perceptual attributes* from visuals.

Attractiveness, or facial beauty, is a good example of high-level perceptual concept which can be formed from the characteristics of individual facial components [13, 23]. Research has shown patterns of consensus among different people; judgments of beauty are predictable and manipulable [21]. Memorability, *i.e.*, how long one can remember a specific face, is another high-level concept that has been studied recently [4].

In this paper, we are interested in learning a variety of facial perceptions but in particular focus on the **socially-important dimensions** and how they predict the **outcomes of real-world social events** such as political elections. Social trait judgment is an active research topic across several disciplines including cognitive science, political science [32], and psychology [25], but with lack of scalability and reproducibility as they are based on behavioral experiments.

Our approach is fully automated and we take advantage of a hierarchical model that leverages a variety of mid-level facial attributes to infer the high level social dimensions. Recently, computer vision and machine learning research started to reach out to social science disciplines and tackle large-scale visual problems by massive processing capacity. For instance, Joo *et al.* [17] proposed to recognize the hidden communicative intents of social images in the news media and link the sentiments to the public opinion. Zhu *et al.* [39] also studied the relations between visual features of images and their effects on the viewer engagement, *i.e.*, the number of comments. These studies develop a new research field of multimodal computational social science, in which computational methods take on challenges in massive datasets and feature dimensions that are intractable with traditional methods.

Finally, our paper is related to the work of Vernon *et al.* [33] who trained neural networks to predict the perceived facial traits but didn't extend their approach to real-world application. Rojas *et al.* [27] also studied the effectiveness of various models for classification of facial traits from synthetic images, and [2] used a similar method to classify perceived personality. These are all limited to psychological analysis at the personal level without seeking a connection to social behaviors. To the best of our knowledge, our paper is the first to establish a complete pipeline of analysis which goes through feature analysis, trait prediction, and election outcome prediction all together. Therefore, it is our unique contribution to unite psychological analysis (facial trait perception) and its effects to social construction of leadership (election) within a unified computational framework.

2. Dataset

2.1. Collection

We introduce a novel dataset of facial photographs of U.S. politicians (491 male and 159 female subjects) each of which is labeled with 14 dimensions of perceived traits, which we discuss in details shortly. Specifically, we collected the facial images of politicians who have run for a political office in U.S. (Senator, Congressman, Governor) in 2000-2012 from the Wikipedia, the other election-related sites, or their own homepages[1]. As the majority of U.S. politicians are of White ethnicity, we only consider those with White ethnic background. Each image has been processed and normalized by cutting out the background and clothing regions, brightness normalization, and aligning the center and the size. A few examples are presented in Fig. 4. Note that we distinguish gender, and the trait visualizations for female politicians are presented in the supplementary material.

All our annotators were the Amazon Mechanical Turk workers located in the U.S. and likely shared the same cultural background as the U.S. voters. Once a pair of images are given, the annotators were asked if they recognize either politician, in which case we discarded their responses (but still paid for the responses). Since the trait evaluation is a subjective task, we asked 10 different workers to respond to the same question and took the average rating. In addition, we asked the annotators to "compare" a pair of images in given dimensions rather than evaluating each image individually. This scheme of annotation is popular when one needs to retrieve the relative ranking order among examples (*e.g.*, relative attributes [26, 19]). One particular advantage of comparison scheme in our case is that the annotators do not need to establish the absolute baseline or scales in these

[1]We also used the images from 2014 elections (without annotation) in analysis, but these were only used for election outcome prediction, not for training.

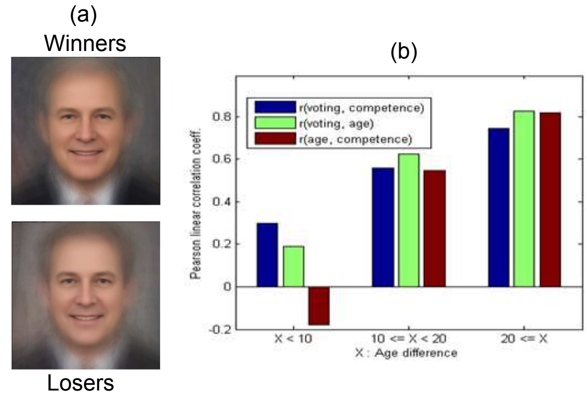

Figure 3. (a) The mean face of winners and that of losers in Senator elections, where the age difference between two groups is apparent. (b) The age, the competence rating, and the voting share are all intercorrelated, particularly when the age difference of two rival candidates is large. The statistics was computed from the data of [32].

social dimensions, which would be inconsistent (*i.e.*, what does a score of 0.3 mean?). The comparison-based ratings naturally identify the strength of each example in the context of relational distance from the other examples, generating a more reliable ranking of subtle signal differences.

2.2. Control Variables

Prior studies in social trait judgment [32] claimed to find strong, predictive signals in facial competence assessments to electoral success. A signal advantage of our three-tiered model is that the individual components of the trait judgment become visible; in attempting to replicate this study, we discovered an undetected artifact effect due to picture quality, clothing, and facial expression. Not surprisingly, high-quality and better-posed images tend to be more frequent for successful than for failed political candidates. To ensure we examined effects exclusively from facial features, we first removed the **clothing and background**, which alone have been shown powerful enough to predict the election outcomes [29]. We also standardized on a smiling face, ruling out the confounding effects of transient **facial expressions**. One may also use emotionally neutral faces [1, 28], but they are typically not available other than from a laboratory environment. In public self-presentation, the most common expression is a smile, which we adopt as normative in this dataset.

In addition, we control for the dimension of **biological age** by restricting pair-wise comparisons to similar ages. The problem is illustrated in Fig. 3, where the older people (senior and often incumbent candidates) have strong advantages in elections and are also rated more competent[2]. That is, any measure correlated with the candidates' ages can predict the election outcomes with no surprise. An older

[2]This was not the case for gubernatorial races with the term limitation.

person is likely to be judged to be more "competent" simply because of his or her age, having by default more experience than a younger person. To determine the contribution of facial features to the election outcomes, one must block this factor. Note that we still keep the dimension "perceived age," a judged dimension distinct from biological age.

Again, it is important to note that all of these controlled variables were carefully selected to separate the pure effect of facial appearance from the effects of covariates – so that we can eventually confirm the true effects of face on the elections. This is the reason why we need to have a complete and unified pipeline of analysis, with a specifically designed dataset, which differs from more general purpose, uncontrolled datasets such as US-10K [4] with the trait annotations for generic facial images.

3. Ontology of Face Perception

3.1. Social Dimensions

In evaluating others based on facial features, people rely on multiple trait dimensions [3]. We develop an ontology of five perceptual categories: essential, biological, moral, social, and professional history. To achieve a broad representation, we make a selection of traits from each category.

1. **Essential**: The face functions as an index for essential or invariant traits such as sex and race. Studies indicate that Americans tend to perceive intelligence as an inborn trait, while other cultures view it as socially contingent [30]. Our dataset includes male and female faces, while race is restricted to white. We consider the traits "intelligent" and "masculine" ("feminine" for female's case), also known for "sex-typicality." [7]

2. **Biological History**: Faces also reflect a person's biological history, giving rise to traits such as perceived age, health, and level of energy, important dimensions for predicting future performance. We consider the traits "perceived age," "baby-faced," [37] and "energetic" in this category.

3. **Moral History**: Habitual and recurring intentions and emotions might leave their mark in the face, giving rise to traits such as perceived kindness, generosity, and honesty. We select "generous" and "honest" as important moral dimensions for interpersonal assessments.

4. **Social History**: A person's social history is commonly inferred from the face, resulting in perceived traits such as levels of grooming, wealth, confidence, and dominance. We select the dimensions "well-groomed," "rich," "confident," and "dominant" to represent this important category.

5. **Professional History**: Professional history is inferred in perceived traits such as experience, competence, and trustworthiness; we select the dimensions "competence" [32] and "trustworthy."

Composite traits are inferred from across these categories; for instance, the trait "baby-faced" commonly used in the literature [37], is negatively correlated with masculinity of inherent traits, biological history, and social history. The correlations of trait dimensions are shown in Fig. 5.

As discussed in previous section, we use Amazon Mechanical Turk to obtain ground-truth annotations of 14 dimensions by pair-wise comparison scheme where an annotator can choose either image to have a stronger signal in each given dimension (*e.g.*, which person is more competent?) For an image pair (I_i, I_j), the score of each dimension is then simply obtained by the average ratings such that:

$$s(I_i, I_j) = \frac{(\text{\# of preference on } \mathrm{I}_i) - (\text{\# of preference on } \mathrm{I}_j)}{\text{\# of total responses}},$$

and we denote $(\mathrm{I}_i \succ \mathrm{I}_j)$ if $s(\mathrm{I}_i, \mathrm{I}_j) > 0$. From these pair-wise ratings, we retrieve the global ranking orders of all examples by HodgeRank [15] to resolve loopy orderings.

3.2. Mid-Level Attributes

In order to effectively infer the ultimate social dimensions, we use two types of mid-level facial attributes: nameable attributes and part appearance type attributes.

3.2.1 Nameable Attributes

We first consider a set of nameable attributes such as "bald" or "mustache". These attributes carry semantically meaningful information and also enable more interpretable analysis on the result. Such attributes have been also shown to be useful in facial beauty understanding [23]. Therefore, we consider a number of binary attributes: Glasses, Bald (M), Mustache (M), Blonde (F), Dark-hair (F), Curly-hair (F), Long-hair (F), Gray-hair, Thick-eyebrow, as well as scalar attributes: Eye-height, Eye-width, Drooping-eye, Angular eyebrow, Tall-nose, Sharp-nosetip, Wide-face, Tall-face, Wide-mouth, Thick-lip, etc – geometric attributes. Some attributes are gender-specific because they are only applicable to either gender.

For the binary attributes, we train a linear SVM from the binary annotation (positive vs. negative) from the low-level image features that we will discuss shortly. The scalar attributes are simply obtained from the positions of detected facial landmarks and represented by real value pixel distances. Therefore, each image is represented by a real-valued vector containing the responses from SVM classifiers (binary) and the measured distance (scalar). We denote this vector by $\mathbf{f}_a(\mathrm{I}_i) \in \mathbb{R}^p$.

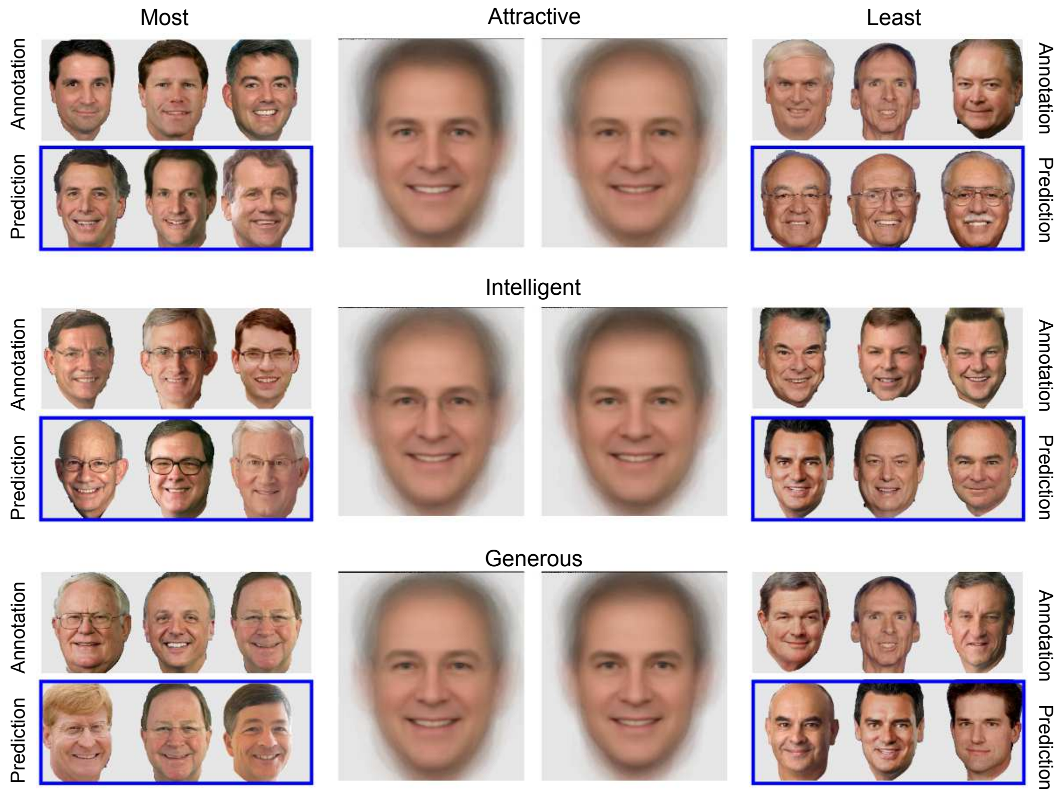

Figure 4. Illustration of three selected social dimensions. For each dimension: (*middle*) the averaged images of examples with high responses and low responses, (*left*) the examples with the highest ratings from (*top-left*) annotation and (*bottom-left*) prediction, marked with blue rectangles, and (*right*) the examples with the lowest ratings. More visuals and female traits are presented in the supplementary material.

3.2.2 Part Shape Vocabulary

While the nameable attributes provide the relevant semantics that can be naturally quantized as a binary or 1-d scalar value ("wide-mouth"), the finer-grained part shape may further enhance to explain the subtle difference in perception although not interpreted verbally. One popular approach to model the facial parts is to utilize a part vocabulary [36, 24], or a dictionary, either pre-defined or obtained by clustering. In this paper we adopt a recent part learning approach in [18] learning the part dictionary directly from the images.

Specifically, we first specify 7 facial regions: head-top, (left, right) eye-brow, (left, right) eye, nose, and mouth. The locations and sizes of corresponding regions in each image are estimated from the facial landmarks. For each specified facial region, we crop the corresponding image patches and extract the low-level image features, HoG [9] and RGB color histogram. Then we obtain the initial clusters of the patches by K-means clustering ($K = 15$) and train a linear SVM detector for each cluster. These detectors are then iteratively refined by updating the cluster memberships of the patches and re-training the detectors. Therefore, each facial part is represented by a real-valued vector of length K containing the responses from the learned K detectors. We denote this vector by $\mathbf{f}_p(\mathrm{I}_i) \in \mathbb{R}^{7\cdot K}$ and construct the full feature vector $\mathbf{f}(\mathrm{I}_i)$ by concatenating with the attribute feature vector.

Low-Level Representation. We briefly explain the low-level image features that are used to infer the mid-level attributes. First, we detect the 76 facial landmarks and use the positions. We implemented the Supervised Descent Method of [35] to detect internal facial fiducial points. Then from each facial region, we also extract the HoG feature and RGB color histogram. Thus we have three types of low-level cue: structural cue from landmarks, shape and color appearance cue from HoG and color histogram. We skip further elaboration since these are very common image feature types.

4. Learning To Rank

In order to train our model to predict the social dimensions, which essentially represent the relative ranking orders among the examples, we use Ranking SVM (RankSVM) [16]. Unlike a binary, or multi-class SVM which maximizes margin between the groups of examples (*i.e.*, positive and negative), RankSVM is more suitable for our task because it aims to preserve the pre-specified pairwise ranking orders in training examples. This advantage has allowed RankSVM to be very popular in the literature of information retrieval (web-search) and also in the recent works in computer vision such as relative attributes [26]. In the experimental section, it will be also shown to provide a straightforward framework to classify the election outcomes (*i.e.*, "who looks more like a winner?").

Specifically, we introduce a formalism to learn one social dimension as follows. We do not specify the index to trait dimension for notational simplicity. Given N training images and their global ranking orders, $D = \{(i, j)|\mathrm{I}_i \succ \mathrm{I}_j\}_{i,j=1}^{N}$, our goal is to learn a linear ranking function, $r(\mathrm{I}) = \langle \mathbf{w}, \mathbf{f}(\mathrm{I})\rangle$, with the following objective:

$$\begin{aligned} \text{minimize} :& \quad \frac{1}{2}||\mathbf{w}||_2^2 + C\sum \xi_{i,j} \\ \text{subject to} :& \quad \mathbf{w}^\top \mathbf{f}(\mathrm{I_i}) \geq \mathbf{w}^\top \mathbf{f}(\mathrm{I_j}) + 1 - \xi_{i,j}, \\ & \quad \xi_{i,j} \geq 0, \forall (i,j) \in D, \end{aligned} \tag{1}$$

where $\mathbf{w}$ is the model parameter to learn, $\xi_{i,j}$ is a non-negative slack variable for every pair in D, and C controls the trade-off between training error and margin maximization. $\mathbf{f}(\mathrm{I})$ is the mid-level representation of each image, discussed in the previous section. We use the implementation of [8] to solve this optimization problem.

5. Experiments

5.1. Dataset Statistics

Inter-annotator Reliability. The annotators' judgments of social traits are subjective; we averaged the scores to obtain the typical judgment (collective judgment). In addition, we deployed two measures of inter-annotator reliability. First, we measure the linear correlation coefficients of the ratings of two groups of contributing annotators, which has been also used in [4] (memorability). Specifically, we first divided all annotators into two groups, obtained two different scores from two groups for each question, and performed correlation test, from which we observed reliable correlations (0.53 for the male image set and 0.57 for the female set). This measure is useful in the annotating setting of AMT, where each annotator responds to only a subset of the data (*i.e.*, missing data). We also measured Cronbach's alpha which indicates most dimensions are reliable (higher than 0.6)[3].

Dimension Correlations. Not surprisingly, the fourteen social trait dimensions are related to each other: the ratings for "Honest," "Generous," and "Trustworthy" are for instance pairwise strongly correlated (Fig. 5). Still, the dimensions remain clearly distinct. "Old," for instance, is not simply the opposite of "Baby-faced" – their differences are revealed in their relations to the other dimensions, such as "Masculine," which is positively related to the latter only.

[3]The full statistics is provided in the supplementary material.

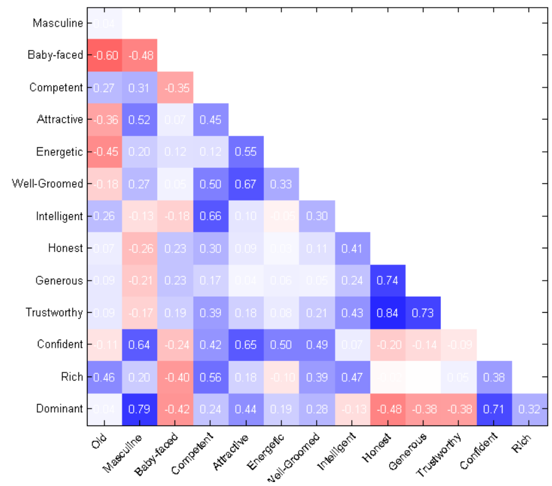

Figure 5. Inter-correlation among different trait dimensions. Blue indicates levels of positive correlation, red negative.

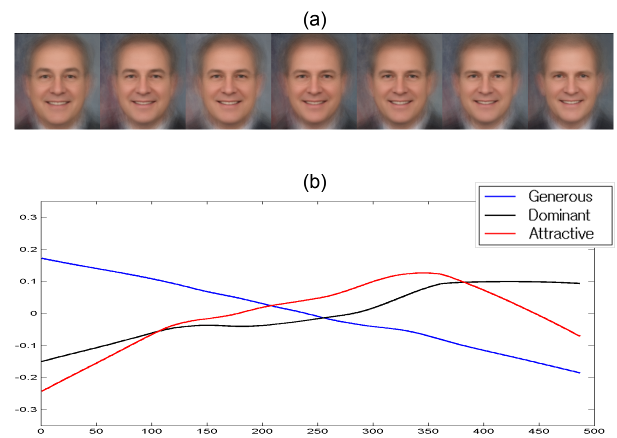

Figure 7. The effect of an individual facial feature – the vertical distance between eyebrow and eyes. (a) The mean image templates obtained from morphed faces in each group of faces sorted by distance. (b) The (smoothed) plot showing the relations between the facial feature and the annotated perceived traits.

5.2. What makes for a Competent Face?

A key contribution of our model based analysis is the explicit decomposition of the holistic character of human perception into its constituent components. This is important not only for improving predictive power, but how for modeling perceptual and cognitive processes, core concerns of psychology and cognitive science.

Feature Contribution: Fig. 6 shows the correlation between a set of mid-level attributes and the social dimensions. The magnitude of correlations vary and some of the individual features are quite strong, which suggest these attributes can be effectively utilized as a feature set to predict the social dimensions. One can also notice the strong effects of the hairstyle and the glasses, which is interesting because these are not necessarily constant and inherent facial features. We also present the effect of a particularly

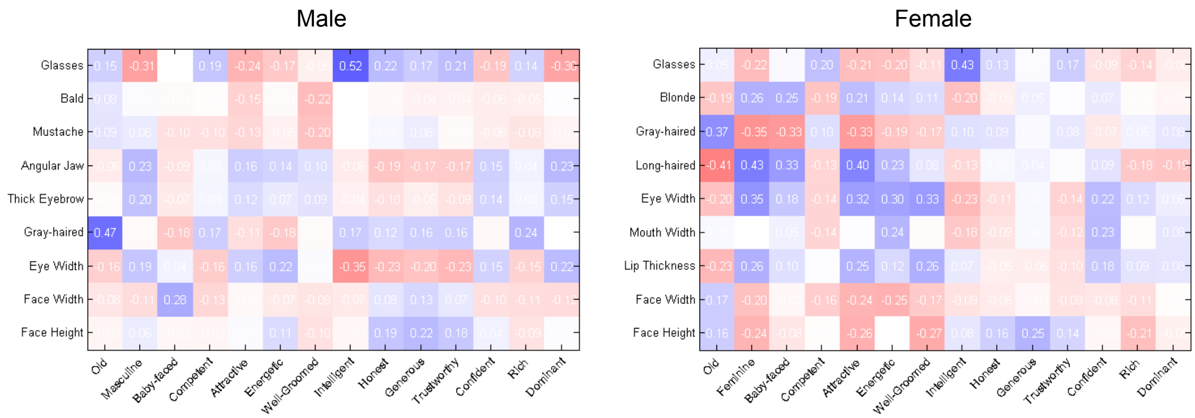

Figure 6. Contributions of the mid-level attributes to the social dimensions, indicated by correlation coefficients. We only present a set of the most interesting attributes. Note that the same attribute can have different meanings in different genders.

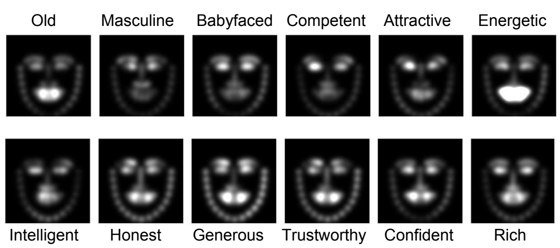

Figure 8. Saliency maps representing the contributions of different facial regions to trait judgment.

strong individual facial feature – the vertical distance between eyebrow and eyes in Fig. 7. Here, Two dimensions (generous and dominant) exhibit linear increase or decrease with the distance but the dimension of attractive shows a non-linear relation, which could be related to 'average' nature of beautiful faces [20].

Facial Region Contribution: We also investigate which facial regions contribute more to each trait dimension. Each facial region may be more or less useful for evaluating social trait judgment; in addition, some regions (*e.g.*, the eye) may play a particularly important role in one particular dimension. To identify this, we fit a regression model for each trait dimension on each facial landmark position (x-y image coordinate). Fig. 8 shows the salient map indicating the magnitude of learned coefficients (*i.e.*, contributions).

5.3. Trait Perception Prediction

We verify the accuracy of the social dimensions prediction made by our trained model. Since we are interested in validating our design choice of hierarchical model, we choose the baseline model that directly infers the social dimensions from the low-level image features. We measure the accuracy of predictions by classification test – comparing predicted social dimensions of every pair of face examples in the dataset. We perform 10-fold cross-validation and report the average performance of each method in Table 1.

The result demonstrates the advantage of our hierarchical approach to exploit a variety of attributes as mid-level representation. The intermediate layer of our model summarizes much of necessary information to be used for inferring the high level traits.

Table 2. Linear correlation coefficients between the predicted social traits and the actual voting shares of politicians. We only show statistically significant results.

	Governor ($n = 112$)		Senator ($n = 110$)	
Traits	r	p-value	r	p-value
Attractive	.433	(< .0001)		
Masculine	.368	(.0001)		
Confident	.347	(.0002)		
Dominant	.332	(.0003)		
Energetic	.302	(.001)	-.314	(.0006)
Well-groomed	.301	(.001)		
Rich			.257	(.005)
Perceived Old	-.191	(.04)	.242	(.008)
Intelligent	-.288	(.002)	.201	(.03)
Competent			.198	(.03)

5.4. What makes for a Winning Face?

The outcome of U.S. election affects the lives of millions of people. and the careers of thousands. There is consequently a high level of interest in predicting the results. It has been reported [32] that manual judgment of facial traits by participants ignorant of the candidates' political experience or platforms predicted the outcome of Senatorial races with a 72% accuracy. Our attempt to reproduce their results, however, revealed that this impressive score may be an artifact of uncontrolled variables such as image quality and biological age. As discussed earlier in Sec.2.2, our dataset only contains the photographs of good quality and we remove clothing and background regions to ensure that each image only contains a facial region.

Table. 2 shows the correlation between the predicted facial traits and the voting share in actual elections, indicating that different tasks will be characterized by distinctive decision profiles. In the Governor races, the traits "confidence," "attractive," "energetic," and "masculine" have the

Table 1. Accuracy of trait prediction of trained models, measured by pair-wise classification.

Method	Per. Old	Mascl.	Baby-face	Compe-tant	Attract.	Ener-getic	Well-groom.	Intelli-gent	Honest	Gene-rous	Trust-wrt.	Confi-dent.	Rich	Domi-nant	Mean
Full	.743	.723	.737	.654	.715	.800	.711	.723	.675	.692	.682	.714	.652	.693	**.708**
LDK+HOG	.670	.623	.661	.590	.646	.707	.608	.640	.602	.599	.611	.624	.575	.612	**.626**
Landmark	.609	.598	.660	.558	.575	.718	.535	.578	.603	.629	.597	.623	.571	.596	**.603**

* Full: Our full hierarchical model with mid-level attributes,
LDK+HOG: A simplified model only using facial landmark positions and HoG features.

Table 3. Accuracy of election outcome classification.

	Senators (110)	Governors (112)
Accuracy (p-val)	.655 (0.01)	.679 (<0.01)

strongest favorable influence on outcomes, while the trait "old" has a negative effect. In sharp contrast, for Senatorial elections, the trait "old" is favorable along with "rich" and "competent," while "energetic" is negatively correlated with the electoral success.

Pair-wise Classification: We also performed a categorization test of election result with our learned model. Specifically, We first obtained a 14-dimensional trait vector for each image from the learned model, discarding a few uncorrelated dimensions, and use these vectors as input features. We only use the pairs of politicians who actually ran against each other in a race and train a RankSVM. We only consider the races of male politicians because the races of two female politicians are very rare. We followed "Leave-one-out" protocol such that it repeats n times to pick up one image pair as a test case and train a classifier with the rest of examples in training set. Table. 3 presents the accuracy showing the election result can be indeed predicted by social dimensions inferred from face. Note that the accuracy is not directly comparable to the prior reports based on human experiments [32] because we explicitly excluded the photographs of low quality and clothing in our dataset and the age factor was discarded.

5.5. Political Party Affiliation Categorization

We also examine whether the social dimensions of face can predict the political party affiliations of politicians, *i.e.*, whether they are **Republicans** or **Democrats**. The political parties in many nations are composed of their member politicians who likely share a more similar ideology, *i.e.*, conservative or liberal. Therefore the question to ask is whether each politician's ideology can be identified from his face or not. In addition, the politicians in our dataset, to run for the major elections, should be selected in primary elections to represent their parties, which means the party affiliation reflects the outcome of another election. This has been studied by prior behavioral studies [7] which reported an accuracy better than chance ($53 \sim 55\%$).

To verify this hypothesis, we take the same procedure as in Sec. 5.4 to train a binary classifier in "Leave-one-out" and Table 4. shows the accuracy ranging from 54% $\sim$

Table 4. Accuracy of party categorization.

	Male		Female	
	Whole	Winners	Whole	Winners
Accuracy	.597*	.626*	.543	.601*

* Statistically significant (p-value < 0.01).

Table 5. Correlation coefficients between political party affiliation and predicted traits.

	Whole Set (491)		Winner Set (343)	
Traits	r	p-value	r	p-value
Intelligent	.155	(.0006)	.199	(.0002)
Perceived Old	.113	(.01)	.160	(.003)
Attractive	-.110	(.01)	-.105	(.05)
Babyfaced	-.106	(.01)	-.143	(.008)

* Positive coefficient: the Democrat side.

62%, which in fact outperforms the recorded human performance in prior work. Also, the classification performance was slightly better in the winner subset. It may imply that each party has its own typical and desirable traits (Table. 5) and thus, the candidates with such faces could have advantages.

6. Conclusion

In this paper, we introduce a hierarchical model for inferring social judgments about persistent social traits based on facial appearance, integrating high-level perceptions, intermediate-level attributes, and low-level image features. By controlling more carefully for variables than prior studies, we demonstrate the trained model is able to predict party affiliation and the outcome of elections, indicating a general potential for utilizing automated social trait judgment for predicting behavior in a broad range of human social relations, such as mate selection, job placement, and political and commercial negotiations. The predictive power over human behavior is not contingent on the accuracy of social trait judgments, but merely on the persistence and regularity of such judgments. Further studies may reveal recurring topologies of flaws in social trait judgment that correct on more reliable sources of information; first impressions can nevertheless be strategically exploited.

Acknowledgement. We would like to thank Erik Bucy for fruitful discussions and suggestions. This work was supported by NSF CNS 1028381. The first author was also partially supported by Kwanjeong Educational Foundation.

References

[1] R. B. Adams Jr, A. J. Nelson, J. A. Soto, U. Hess, and R. E. Kleck. Emotion in the neutral face: A mechanism for impression formation? *Cognition & emotion*, 26(3):431–441, 2012. 3

[2] N. Al Moubayed, Y. Vazquez-Alvarez, A. McKay, and A. Vinciarelli. Face-based automatic personality perception. In *ACM MM*. ACM, 2014. 3

[3] A. P. Atkinson and R. Adolphs. The neuropsychology of face perception: beyond simple dissociations and functional selectivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1571):1726–1738, 2011. 4

[4] W. A. Bainbridge, P. Isola, and A. Oliva. The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General*, 142(4):1323, 2013. 2, 4, 6

[5] D. S. Berry. Taking people at face value: Evidence for the kernel of truth hypothesis. *Social Cognition*, 8(4):343–361, 1990. 1

[6] I. V. Blair, C. M. Judd, and K. M. Chapleau. The influence of afrocentric facial features in criminal sentencing. *Psychological Science*, 15(10):674–679, 2004. 2

[7] C. M. Carpinella and K. L. Johnson. Appearance-based politics: Sex-typed facial cues communicate political party affiliation. *Journal of Experimental Social Psychology*, 49(1):156–160, 2013. 4, 8

[8] O. Chapelle and S. S. Keerthi. Efficient algorithms for ranking with svms. *Information Retrieval*, 13(3), 2010. 6

[9] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005. 5

[10] S. Du, Y. Tao, and A. M. Martinez. Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*, 111(15):E1454–E1462, 2014. 1

[11] B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. *Pattern Recognition*, 36(1):259–275, 2003. 1

[12] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. Sexnet: A neural network identifies sex from human faces. In *NIPS*, 1990. 1

[13] H. Gunes. A survey of perception and computation of human beauty. In *ACM workshop on Human gesture and behavior understanding*, 2011. 2

[14] S. Gutta, H. Wechsler, and P. J. Phillips. Gender and ethnic classification of face images. In *FG*, 1998. 1

[15] X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye. Statistical ranking and combinatorial hodge theory. *Mathematical Programming*, 127(1):203–244, 2011. 4

[16] T. Joachims. Optimizing search engines using clickthrough data. In *ACM SIGKDD*, 2002. 5

[17] J. Joo, W. Li, F. F. Steen, and S.-C. Zhu. Visual persuasion: Inferring communicative intents of images. In *CVPR*. IEEE, 2014. 2

[18] J. Joo, S. Wang, and S.-C. Zhu. Human attribute recognition by rich appearance dictionary. In *ICCV*. IEEE, 2013. 5

[19] A. Kovashka, D. Parikh, and K. Grauman. Whittlesearch: Image search with relative attribute feedback. In *CVPR*. IEEE, 2012. 3

[20] J. H. Langlois and L. A. Roggman. Attractive faces are only average. *Psychological science*, 1(2):115–121, 1990. 7

[21] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski. Data-driven enhancement of facial attractiveness. In *SIGGRAPH*, volume 27, page 38. ACM, 2008. 2

[22] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang. Probabilistic elastic matching for pose variant face verification. In *CVPR*. IEEE, 2013. 1

[23] L. Liu, H. Xu, J. Xing, S. Liu, X. Zhou, and S. Yan. Wow! you are so beautiful today! In *ACM MM*, 2013. 2, 4

[24] M. Meng, M. Zhao, and S.-C. Zhu. Artistic paper-cut of human portraits. In *MM*. ACM, 2010. 5

[25] N. N. Oosterhof and A. Todorov. The functional basis of face evaluation. *PNAS*, 105(32):11087–11092, 2008. 1, 2

[26] D. Parikh and K. Grauman. Relative attributes. In *ICCV*. IEEE, 2011. 3, 6

[27] M. Rojas, D. Masip, A. Todorov, and J. Vitria. Automatic prediction of facial trait judgments: Appearance vs. structural models. *PloS one*, 6(8), 2011. 3

[28] C. P. Said, N. Sebe, and A. Todorov. Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, 9(2):260, 2009. 3

[29] M. L. Spezio, L. Loesch, F. Gosselin, K. Mattes, and R. M. Alvarez. Thin-slice decisions do not need faces to be predictive of election outcomes. *Political Psychology*, 33(3):331–341, 2012. 3

[30] H. Stevenson and J. W. Stigler. *Learning gap: Why our schools are failing and what we can learn from Japanese and Chinese educ*. Simon and Schuster, 1994. 4

[31] J. Suo, S.-C. Zhu, S. Shan, and X. Chen. A compositional and dynamic model for face aging. *TPAMI*, 32(3):385–401, 2010. 1

[32] A. Todorov, A. N. Mandisodza, A. Goren, and C. C. Hall. Inferences of competence from faces predict election outcomes. *Science*, 308(5728):1623–1626, 2005. 1, 2, 3, 4, 7, 8

[33] R. J. Vernon, C. A. Sutherland, A. W. Young, and T. Hartley. Modeling first impressions from highly variable facial images. *Proceedings of the National Academy of Sciences*, 111(32):E3353–E3361, 2014. 3

[34] J. Willis and A. Todorov. First impressions making up your mind after a 100-ms exposure to a face. *Psychological science*, 17(7):592–598, 2006. 1

[35] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. In *CVPR*, 2013. 5

[36] Z. Xu, H. Chen, S.-C. Zhu, and J. Luo. A hierarchical compositional model for face representation and sketching. *TPAMI*, 30(6):955–969, 2008. 5

[37] L. A. Zebrowitz, J.-M. Fellous, A. Mignault, and C. Andreoletti. Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Personality and social psychology review*, 7(3):194–215, 2003. 1, 4

[38] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *Acm Computing Surveys (CSUR)*, 35(4):399–458, 2003. 1

[39] J. Zhu, J. Luo, Q. You, and J. R. Smith. Towards understanding the effectiveness of election related images in social media. In *ICDMW*, 2013. 2