Winning Faces: Model of basic primate visual processing predicts elections

Darren Schreiber[1, 2*+], Nicolas Pinto[3, 4 +], Patrick Rogers[5], James J. DiCarlo[3, 4]

[1] *Department of Political Science, Central European University, Budapest, Hungary.*

[2] *Department of Politics, University of Exeter, Exeter, EX4 4RJ, United Kingdom.*

[3] *McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts.*

[4] *Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts.*

[5] *Department of Political Science, University of California San Diego, La Jolla, CA 92103 USA*

[*]Correspondence can be sent to darren.schreiber@gmail.com
[+]Contributed equally

Abstract

Todorov (2005) demonstrated that people could make competence judgments about faces with only 100 millisecond exposures that were predictive of electoral victories. Others have since shown that Swiss children can predict the outcome of French elections when they pick which face they would want to be the 'captain of the boat' in an adventure game. And, Indians and Americans can pick the winners of Mexican and Brazilian political contests just by looking at their faces. The automaticity, early development, and universality of this ability suggests it is likely instantiated in a basic cognitive process. We trained a computer model of the V1 region of the primate visual cortex that has performed very well in object recognition tasks to identify winning candidates with an accuracy of 70-80%.

Voters must rely on mental shortcuts (such as party affiliation, ideology, endorsements, success in polls, and appearance) rather than on a complete knowledge of political candidates' policy preferences, qualifications, and alliances when choosing a representative (*1*). Despite very low levels of political knowledge and the use of these heuristics, voters are able to correctly choose a candidate that matches their policy preferences 75% of the time (*2*). As primates, our ability to quickly navigate the complex and dynamic politics of our social context is critical to survival and we need to make nearly instant judgments about the likely winners and losers in tribal conflicts (*3*). Our modern politics appear to be constructed on these very ancient foundations (*4*).

Attributions about personal characteristics of political candidates made solely on the basis of facial appearance have been repeatedly demonstrated to predict election results (*5*). Since these judgments can be made based on only 100 millisecond glimpses, can be made by young children (*6*), and appear to be culture independent (*7*), we hypothesized that these assessments might be the result of rapid, automatic visual processes. Although other brain structures are known to be involved in complex social cognition tasks, the primate's ventral visual processing stream is critical for swiftly discriminating among visually-presented objects (*8*), including faces.

To test whether basic ventral visual processes could be sufficient for predicting electoral outcomes from only facial appearance, we began by using a computer algorithm that models only the first stage of primate cortical visual processing – the primary visual cortex (area V1). This "V1-like" model functions as a neuroscience "null" model because it is only a first order description of the way the early visual system codes visual images and would not *a priori* be expected it to perform well with real-world object and face recognition tasks. Nonetheless, in some contexts the output of this relatively simple visual code can beat other state of the art computer vision approaches (*9*).

We trained the standard linear classifiers on the V1-like visual code, with some of the same set of candidate photos that have been used in many studies with human subjects (*5*) and then tested the election prediction performance using photos of candidates that it had not previously seen. Using only face-cropped photos, we found that this V1-like model predicted previous U.S. Senate and House of Representatives races with accuracy comparable to human subjects in previous appearance-based studies (U.S. House 74.6% correct, sem: 2.3%; U.S. Senate 75.5% correct, sem: 2.4%; chance is 50%). Because performance was much lower using only a raw pixel code (55.1% and 51.4% respectively), this suggests that the predictive power results from the way the primate brain represents those images (approximated here by the V1-like model).

To test generalizability across cultures, we first used the same V1-like model (without any additional training data) to successfully predict elections in Mexico (55.6% correct) and Brazil (70.4% correct). Second, we made accurate (65.1% correct, $p<0.001$) prospective predictions of the 2010 U.S. House, Senate, and Gubernatorial elections using the V1-like model trained on previous elections. Consistent with studies using human raters (*10*), the accuracy depended on whether the election was in an uncompetitive (75.8% correct, $p<0.001$) or highly competitive district (35.1% correct, $p=0.16$). In districts that are evaluated to be more politically competitive, both parties appear to put their "best face forward" neutralizing any facial appearance advantage.

The algorithm appears to rely on facial features similar to those that are primarily attended to by humans (e.g. eyes, nose, mouth) when engaged in face processing (**Figure 1**). While it remains unclear precisely what characteristics of the image are enabling this algorithm to make accurate forecasts of electoral results and it is unclear if humans rely on these same characteristics, this computational approach should enable a more systematic analysis of the relationship between appearance and electoral outcomes. Although many candidate characteristics that drive voting behavior can be readily identified (party affiliation, endorsements) or quantified (ideology, polling results) using traditional tools of political science, candidate appearance has been far more difficult to study rigorously. These results suggest that features of a candidate's appearance that are identifiable by rapidly computed, low-level visual processes are predictive of electoral outcomes. Since the complexity of human politics demands that we rely upon shortcuts to make decisions (*1*), further study of the neural mechanisms which implement these shortcuts is critical for our understanding of the democratic process.

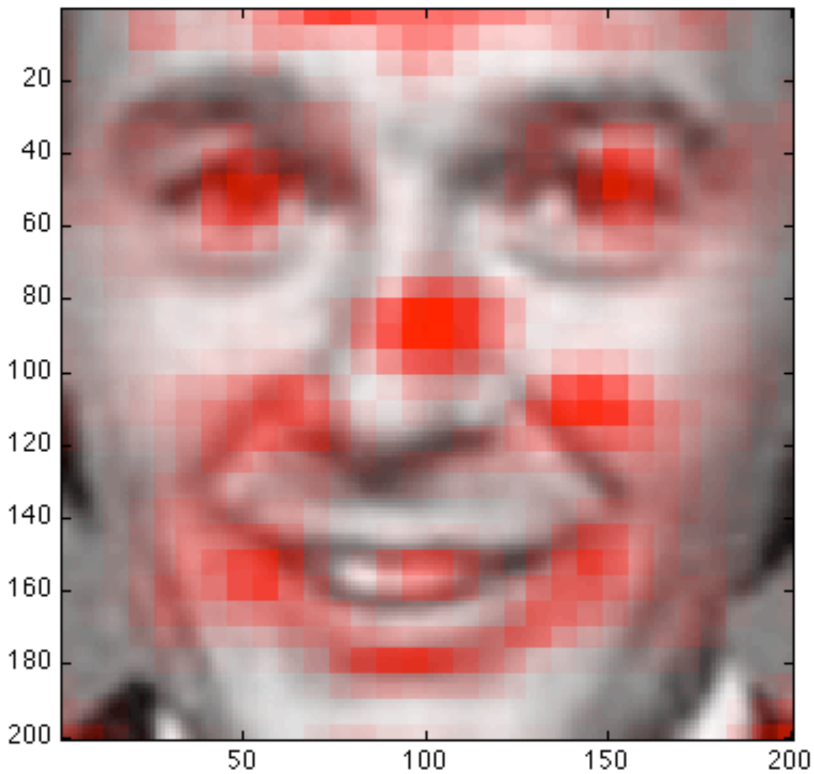

Fig. 1. Red areas show regions of the face where differences in the two candidates' appearance had the strongest impact on predictions of electoral outcome. These regions and the detailed weighting on the underlying V1-like visual feature (see methods) were discovered from photos of pairs of political candidates in which the electoral outcome of each race was provided. For reference, an example image of a candidate is shown in the background.

Work Cited

1. A. Lupia, M. D. McCubbins, *The democratic dilemma : can citizens learn what they need to know?* , Political economy of institutions and decisions (Cambridge University Press, Cambridge, U.K. ; New York, 1998), pp. xiv, 282 p.
2. R. Lau, D. Andersen, D. Redlawsk, An exploration of correct voting in recent US presidential elections. *American Journal of Political Science* **52**, 395 (2008).
3. F. B. M. de Waal, *Chimpanzee politics : power and sex among apes*. (Johns Hopkins University Press, Baltimore, ed. Rev., 1998), pp. xv, 235 , [16] of plates.
4. J. H. Fowler, D. Schreiber, Biology, politics, and the emerging science of human nature. *Science* **322**, 912 (Nov 7, 2008).
5. A. Todorov, A. N. Mandisodza, A. Goren, C. C. Hall, Inferences of competence from faces predict election outcomes. *Science* **308**, 1623 (Jun 10, 2005).

6. J. Antonakis, O. Dalgas, Predicting Elections: Child's Play! *Science* **323**, 1183 (Feb 27, 2009).
7. C. Lawson, G. S. Lenz, A. Baker, M. Myers, Looking Like a Winner: Candidate Appearance and Electoral Success in New Democracies. *World Politics*, (2010).
8. J. J. DiCarlo, D. D. Cox, Untangling invariant object recognition. *Trends in Cognitive Science* **11**, 333 (Aug 1, 2007).
9. N. Pinto, D. D. Cox, J. J. DiCarlo, Why is real-world visual object recognition hard? *PLoS Comput Biol* **4**, e27 (Jan 1, 2008).
10. M. Atkinson, R. Enos, S. Hill, Candidate Faces and Election Outcomes: Is the Face–Vote Correlation Caused by Candidate Selection? *Quarterly Journal of Political Science* **4**, 229 (2009).

Supplementary Online Materials

1) Data Collection

a) Image collection and preprocessing

To build the database of candidates for the 2010 elections we first identified all 72 major party candidates for governor in the United States, all 73 major party candidates for the Senate, and all 850 major party candidates for the House of Representatives. In selected races, we also identified 3rd party candidates that either had the potential to win election outright, or act as "spoiler" candidates--third party candidates who would attract voters from predominantly a single party, allowing the other party to gain an electoral advantage.

From this master list of candidates, we searched for digital headshots – with the subject facing directly into the camera—for each candidate. Candidate photos were predominantly recovered from the candidate's own websites, although a substantial portion instead opted to host their image directories on social networking sites, in particular, Facebook. Additionally, if candidates did not directly offer appropriate images for the analysis, then public image directories were searched, primarily via Google Image Search as well as Flickr.com. Once images had been collected, the images were pre-processed. Four alterations were made to each image. First, the images were cropped just below the chin, above the head, and on either side. They were then re-sized to be 160 pixels in width by 220 pixels in height. Next, they were desaturated. The face images were finally "aligned" using the detection algorithm from face.com's public API. Specifically, each face was rotated (in plane) and cropped so that the eyes and mouth are roughly aligned in each image.

b) Election data collection

Additionally, each race was ranked according to district competitiveness. We looked at 3 different measures for this. The first score we used was the Cook's Race Rankings, which is a subjective ranking based off of expert estimates. Candidates are ranked on a seven-point scale, ranging from: (1) Solid Democrat, (2) Likely Democrat, (3) Lean Democrat, (4) Toss-Up, (5) Lean Republican, (6) Likely Republican, and (7) Solid Republican. Furthermore, for House races we also used the Cook's Partisan Voting Index, which is a measure of a district's partisanship based on the margin of votes cast for each party's presidential candidate in the prior presidential election, compared with the overall margin of votes cast for each presidential candidate in the country as a whole. Finally, we also used the FiveThirtyEight competitiveness score, which is based in part off of polling data as well as running simulated elections 10,000 times to generated predicted probabilities of winning for each candidate.

2) Prediction Algorithm

For each race, a pair of images is given as inputs to the prediction algorithm outlined below. The output is a simple binary decision, i.e. is the winning face on the left or on the right?

a) "V1-like" Visual Representation

We represent each image using the "V1-like" visual representation taken without additional optimization from Pinto et al's (*1*) "V1S+".

This visual representation is based on a first-order description of primary visual cortex V1 and consists of a collection of locally-normalized, thresholded Gabor wavelet functions spanning a range of orientations and spatial frequencies.

We have proposed these V1-like features as a neuroscientist "null" model (a baseline against which performance can be compared) for object and face recognition since they do not contain a particularly sophisticated representation of shape or appearance, nor do they possess any explicit mechanism designed to tolerate image variation (e.g. changes in view, lighting, position, etc. (*2*).

Here, this model serves as a lower bound on the level of performance that can be achieved by only relying on relatively low-level regularities that exist in the test set. Previous work demonstrates that even simple ad-hoc additions to the model to incorporate expected features of the images can significantly improve preformance (*1*).

In spite of their simplicity, these features have been shown to be among the best-performing non-blended features set on many standard object and face recognition benchmarks (*1*) << ref >>,<< ref >> and they are a key component of the best blended solutions for some of these same benchmarks (*3*).

We used the publicly available source code to generate these features and followed the same basic read-out/classification procedure as detailed in << ref >>, with two minor modifications. Specifically, no PCA dimensionality reduction was performed prior to classification (the full vector was used), and different SVM regularization hyperparameter were chosen (see below).

For a detailed description of the V1-like visual representation, we refer the interested reader to the methods of << ref >> << ref >><< ref >>, and the publicly available open-source code << ref >><< ref >>,

c) Classification

For each race, the algorithm produced two "V1-like" feature vectors, one for each candidate's image. This pair is then "fused" by taking the element-wise difference between the two vectors. Finally, for each pair, the resulting feature vector was labeled as "left" or "right," and the task of labeling new (test) examples was treated as a two-category classification learning problem (theoretical chance being 50%). Training and test images were carefully separated to ensure proper validation and the performance was reported as the *average classification accuracy*.

To prevent artifacts and to learn a more robust predictor, we augment the training set with identical races but with input images flipped vertically (we assume that this simple image manipulation would not have changed the outcome of the race). The resulting training set is four times larger (i.e. pair 1: "A" vs "B", pair 2: "A" vs "B flipped", "A flipped" vs "B flipped", "A flipped" vs "B").

<u>1) Previous Elections: U.S. House and U.S. Senate</u> (*4, 5*)
For these two image sets, we used a SVM classifier with precomputed linear kernel (LIBSVM solver (*6*)) with a 10-trial random subsampling cross-validation scheme (90% training and 10% testing). Here, the regularization hyperparameter *C* was fixed to 10^4 (resulting in a quasi-parameter-free hard-margin SVM). In the main text, we reported the performance as the mean *average classification accuracy* and +/- s.e.m. over the 10 splits with this fixed set-up.

<u>2) Previous Elections: Mexico and Brazil</u> (*7*)
For these two image sets, we also used a linear SVM classifier trained on all the pairs from the previous elections of U.S. House and U.S. Senate (see above). The regularization hyperparameter *C* was chosen by cross-validation on the training set only (containing only U.S. elections). We tested 5 values for *C* on a log10 scale (1e0, 1e1, 1e2, 1e3, 1e4) and found *C = 100* to be the optimal one. In the main text, we reported the performance as the *average classification accuracy* on the testing set (containing only Mexican or Brazilian races).

<u>3) 2010 U.S. Elections</u>
We used the same procedure as in Mexico and Brazil elections to predict the 2010 U.S. elections. The prospective predictions were posted to a public website.

d) Source code

In order to facilitate the reproduction of our results, we will distribute the image sets (original and aligned) and we will make open source code (with no commercial dependency) available on the social coding website github.com.

Work Cited

1. N. Pinto, D. D. Cox, J. J. DiCarlo, Why is real-world visual object recognition hard? *PLoS Comput Biol* **4**, e27 (Jan 1, 2008).

2. J. J. DiCarlo, D. D. Cox, Untangling invariant object recognition. *Trends in Cognitive Science* **11**, 333 (Aug 1, 2007).
3. P. Gehler, S. Nowozin, paper presented at the IEEE International Conference on Computer Vision (ICCV 2009), Kyoto, Japan, 2009.
4. M. Atkinson, R. Enos, S. Hill, Candidate Faces and Election Outcomes: Is the Face–Vote Correlation Caused by Candidate Selection? *Quarterly Journal of Political Science* **4**, 229 (2009).
5. A. Todorov, A. N. Mandisodza, A. Goren, C. C. Hall, Inferences of competence from faces predict election outcomes. *Science* **308**, 1623 (Jun 10, 2005).
6. C.-C. Chang, C.-J. Lin. (2010).
7. C. Lawson, G. S. Lenz, A. Baker, M. Myers, Looking Like a Winner: Candidate Appearance and Electoral Success in New Democracies. *World Politics*, (2010).