Quarterly Journal of Political Science, 2009, 4: 229–249

Candidate Faces and Election Outcomes: Is the Face–Vote Correlation Caused by Candidate Selection?*

Matthew D. Atkinson*, Ryan D. Enos† and Seth J. Hill‡

Department of Political Science, University of California, Los Angeles, USA
matthewa@ucla.edu
†*renos@ucla.edu*
‡*sjhill@ucla.edu*

ABSTRACT

We estimate the effect of candidate appearance on vote choice in congressional elections using an original survey instrument. Based on estimates of the facial competence of 972 congressional candidates, we show that in more competitive races the out-party tends to run candidates with higher quality faces. We estimate the direct effect of face on vote choice when controlling for the competitiveness of the contest and for individual partisanship. Combining survey data with our facial quality scores and a measure of contest competitiveness, we find a face quality effect for Senate challengers of about 4 points for independent voters and 1–3 points for partisans. While we estimate face effects that could potentially matter in close elections, we find that the challenging candidate's face is never the difference between a challenger and incumbent victory in all 99 Senate elections in our study.

* The authors thank John Zaller, Lynn Vavreck, Jeff Lewis, R. Brian Law, Andrew Gelman, our survey participants, and especially Elisabeth Michaels for research assistance and Alex Todorov for sharing pictures, data, and feedback.

Replication Data available from:
http://dx.doi.org/10.1561/100.00008062_supp
MS submitted 18 August 2008; *final version received 18 June 2009*
ISSN 1554-0626; DOI 10.1561/100.00008062

"It's called being ugly," Senator Phil Gramm (R-TX) responded when asked on *60 Minutes* why he came across poorly on television.[1] Gramm's face is well below average in our estimate of the quality of almost 1,000 political candidate faces. Does the facial appearance of political candidates affect the voter's choice?

In this paper, we identify the effect of candidate facial competence on individual vote choice controlling for electoral context and voter partisanship. We show that higher quality challenger faces are selected into more competitive districts, suggesting that models of the effect of candidate face on vote should account for district competitiveness. When controlling for this selection, we show that the difference between a below average and above average challenger face increases the probability of voting for the challenger by between 1 and 3 points for partisan voters, and by almost 4 points for independent voters. The estimated effects of incumbent faces are not statistically different from zero.

Our work builds on a growing literature on appearance and elections in two ways. First, we demonstrate the importance of electoral context by showing that the best candidate faces are not allocated randomly across districts: in competitive districts the challenger candidate tends to have a higher quality face, either because of the decision-making process of potential candidates or because of the efforts of political elites. The selection of quality candidate faces into competitive districts has implications for the unbiased estimation of the effect of candidate faces on election outcomes, namely that district competitiveness may be an omitted variable in certain specifications.

Second, we develop a cross-district regression analysis that allows us to compare the electoral effect of particular candidate faces across elections. This cross-district analysis is possible because of a novel measurement tool for candidate facial quality. Previous measurements consider only the relative advantage of one candidate's face over the face of their opponent in a one-on-one comparison. Our method assigns individual scores to each candidate face, allowing us to compare any candidate with any other and to consider separately the effects of challenger and incumbent faces. In addition, the individual candidate face scores enable us to measure the effect of district competitiveness on the facial characteristics of individual candidates. By merging face scores to survey data, we are able to estimate the effect of candidate faces on vote choice controlling for district competitiveness and voter characteristics.

CANDIDATE FACES AND ELECTIONS

In a seminal contribution, Todorov *et al.* (2005) show that undergraduates exposed to the two faces of competing congressional candidates choose the winner of the election as more competent in 70 percent of races. This predictive power is most effective when exposure to the faces is for less than one second and when the participants have no prior knowledge about either candidate. This finding has been replicated across political

[1] Quoted in *The New York Times*, Frank Rich, "Journal; Their Own Petard," February 23, 1995. Accessed at http://query.nytimes.com/gst/fullpage.html?res=990CEFDB133CF930A15751C0A963958260.

contexts and with candidates from different countries.[2] Rapid inferences from faces also predict non-political outcomes such as success in business.[3]

Previous studies showing a correlation between faces and election outcomes have done so at the aggregate level, for example, comparing the vote percentage garnered by the winning candidate and the percentage of participants who picked that candidate more competent. Part of the aggregate correlation, however, may be due to variables omitted from the two-variable comparison. In particular, we argue that the effect of expected incumbent electoral performance on challenger selection may be an important omitted variable.

Quality potential candidates are more likely to stand for office when their chances of success are high (Jacobson and Kernell 1983, Carson 2005). Scholars of candidate quality suggest that appearance may be part of quality (Green and Krasno 1988, Jacobson 1989, Squire 1992, Squire and Smith 1996). Beyond the direct contribution of quality faces to candidate quality, the economic return to attractiveness measured by economists suggests that more attractive individuals have higher human capital endowments.[4] Thus an indirect association between facial quality and candidate quality may operate through the association between human capital and candidate quality.

These considerations suggest the possibility that the decisions made by parties and candidates during the candidate selection process contribute to the correlation between candidate faces and election outcomes. If better candidate faces are related to higher quality candidates, and if quality candidates select into districts where they have a better *a priori* chance of victory, then better candidate faces are selected into districts with a better *a priori* chance of victory. This selection process could produce a correlation between outcomes and candidate faces before any votes are cast. A model to capture the effect of face on vote, therefore, should control for each candidate's *a priori* chance of victory.

We use two methods to accurately measure the effect of face on vote controlling for potential candidate selection effects. First, we identify and control for the selection of candidate faces to districts and, second, we look at surveys of individual voters. With individual observations we can control for electoral context and potentially confounding voter characteristics in a model of congressional vote choice. Most importantly, we can hold constant the effect of voter partisanship and its relationship with the incumbent's party. This allows us to better estimate the direct effect of candidate face on individual vote choice.

MEASURING FACIAL COMPETENCE

In this section, we describe our new method for creating individual measures of candidate facial competence. Our measure improves upon previous measures of candidate facial

---

2 For example, see Benjamin and Shapiro (2006), Berggren *et al.* (2006), Ballew II and Todorov (2007), Lenz and Lawson (2007), Little *et al.* (2007), and Antonakis and Dalgas (2009).

3 For example, Rule and Ambady (2008).

4 See for example, Hamermesh and Biddle (1994), Biddle and Hamermesh (1998), and Mobius and Rosenblat (2006).

characteristics by estimating a numerical value for each individual face. To obtain these estimates, we wrote a computer-based survey that presents survey participants with two randomly drawn faces from the pool of all candidate faces. Each participant evaluated hundreds of face pairs. For each pairing, participants were exposed to the images for one second and asked to choose which of the two faces appeared more competent.[5] The text of the question and experimental design followed as closely as possible that used by Todorov *et al.* (2005). The word *competent* in the survey question text was not defined for the respondents, and those that asked the administrators for a definition were asked to use their understanding of the word to make the comparison. We recruited participants from political science undergraduate classes. An example of the survey can be found at http://sjhill.bol.ucla.edu/faces.

Our key improvement over the procedure used by other candidate face researchers is to present two randomly drawn faces from the pool of all contests, rather than only the two opposing candidates in each actual contest. Our measure thus allows the comparison of candidates across contests through a numerical face score of each candidate.

We used more than 167,000 binary choices by participants to build competence scores for each face in our candidate pool from 2004 House elections and 1990–2006 Senate elections.[6] The statistical method we use to calculate the scores is based on three assumptions and follows the method of Groseclose and Stewart III (1998). First, we assume that there is a latent continuum of facial competence on which each face can be placed that drives the average perceptions of all raters. Second, we assume that participant evaluations have a probabilistic, not deterministic, relationship with the latent facial competence dimension. Third, we assume that participant evaluations are transitive.

Based on these assumptions, facial competence scores can be estimated by determining the relative positions of the faces on the competence continuum that would have been most likely to produce the choices made by our participants. The estimated position provides a numerical value of the latent facial competence for each face in our pool. The estimated locations on the continuum also allow us to calculate the probability that any face will be chosen over any other face in a pairwise evaluation by our participants based upon the distance between the two faces, enabling us to compare our measurement with other measures of candidates' faces.

Our estimates match well with those of Todorov *et al.* (2005). In the Appendix we report this reproduction, provide further details of the survey, present the technical details of our estimation model, and discuss robustness checks. For interpretation

---

5 According to Todorov *et al.*'s (2005) comparison of a number of different facial traits, competence predicts election outcomes more effectively than the other traits evaluated. When we refer to the quality of a candidate's face, we refer to inferences of candidate competence, though we acknowledge that this trait is not always found most predictive (Berggren *et al.* 2006).

6 We use only white male candidates for the 2004 House candidates. We estimated the scores with a variety of robustness checks based upon recognition, respondent consistency (we repeated the same face pairs within respondents, varying left–right status of the repeat pair), and dropping early and late evaluations for fatigue and learning. None of the alternatively estimated scores substantively affected our results.

purposes, we standardize the competence scores for each chamber to mean zero and unit variance.

The scores represent a numerical estimate of the facial competence of actual politicians. Because the House and Senate are estimated separately, it is not meaningful to compare scores across chambers. For each chamber, the distribution of candidate face scores is approximately normal, though there is a leftward skew of lower-quality faces in both distributions. After standardizing the scores, the median Senate face is 0.17. The median challenger face is 0.015, and the median incumbent, Mitch McConnell (R-KY), is 0.29. The maximum Senate incumbent face is Russ Feingold (D-WI) at 1.99, less than the maximum challenger, John Thune (R-SD) with 2.22. Thune became the most competent-looking Senator in our sample when he defeated Tom Daschle in 2004. The minimum Senate challenger face is −3.98, less than the minimum incumbent face of Spencer Abraham (R-MI) at −2.49.

In the House, as in the Senate, the typical incumbent face score is higher than the typical challenger face score. The median House challenger scored −0.06, and the median incumbent 0.24. The minimum House challenger scored −1.06 and the minimum incumbent −0.74. The maximum House challenger scored 0.69, and the maximum house incumbent 0.77.

THE SELECTION OF CANDIDATE FACES TO ELECTION CONTESTS

With a numerical measurement of the facial competence of each candidate in our pool, we can investigate the selection of candidate faces into districts. In Table 1 we present evidence that district competitiveness predicts the facial quality of the challenging candidate. We regress our measure of candidate facial competence score on district competitiveness as evaluated by the *Cook Political Report* (Cook 1992–2006).[7] Cook classifies each campaign as *Tossup*, *Lean*, *Likely*, or *Safe* for each party. In an attempt to keep our measure untainted by the challenging candidate's characteristics, we use Cook publications from at least one year before each election, so that the challenger is unlikely to have yet been selected. For example, our measure of competitiveness for the 2004 elections is taken from the August 2003 *Cook Political Report* newsletter. We call Cook's measure *Incumbent Risk*.

The parameter estimates in columns 1 and 2 of Table 1 indicate a statistically and substantively significant relationship between incumbent risk and challenger facial competence. Moving from a race categorized by Cook as "safe" for the incumbent to a race categorized by Cook as a "tossup" (3 units on the scale) leads to a predicted increase of 0.9 in the facial competence of the House challenger and of 0.8 of the Senate challenger, changes of almost one standard deviation in facial competence. One standard deviation in the Senate is about the difference between the face scores of Phil Gramm (R-TX) and

---

7 For examples of scholarly work employing Cook's report, see Gimple *et al.* (2008) or Vavreck (2001). Models using the Cook variable exclude the 1990 Senate contests because we were unable to obtain a *Cook Political Report* for 1989.

Table 1. Predicting candidate facial competence with district competitiveness.

	House challengers	Senate challengers	House incumbents	Senate incumbents
Intercept	0.427	0.478	0.611	0.398
	(0.306)	(0.396)	(0.251)	(0.290)
Cook incumbent risk	0.304	0.285	0.103	−0.010
	(0.110)	(0.097)	(0.090)	(0.071)
1994 Fixed effect		−0.336		−0.036
		(0.447)		(0.328)
1996 Fixed effect		−0.408		−0.227
		(0.465)		(0.341)
1998 Fixed effect		−0.032		−0.442
		(0.444)		(0.326)
2000 Fixed effect		−0.475		−0.241
		(0.413)		(0.303)
2002 Fixed effect		0.001		0.103
		(0.413)		(0.303)
2004 Fixed effect		0.214		−0.248
		(0.422)		(0.309)
2006 Fixed effect		−0.202		−0.203
		(0.420)		(0.308)
N	148	145	167	145
R^2	0.049	0.084	0.008	0.038
Adjusted R^2	0.043	0.030	0.002	−0.019
Std. error of regression	0.989	1.160	0.852	0.850

Ordinary least squares regression coefficients with standard errors in parentheses. House models are for candidates from 2004, Senate models for candidates from 1992–2006. Dependent variable is facial competence, Cook incumbent risk is coded increasing from low to high risk.

Dick Durbin (D-IL). The coefficient estimates predicting incumbent facial competence are not statistically different from zero (columns 3 and 4). Senate models include year of election fixed effects.

The relationship presented here between challenger face and district competitiveness has implications for the unbiased estimation of the effect of candidate faces on election outcomes. If district competitiveness is related to the eventual outcome of each election and if the Cook evaluation is also related to the challenger's facial qualities, a correlation between challenger face and election outcome will exist even if no voters are influenced by the challenger's face. Omitting district competitiveness, therefore, is likely to bias estimates in a two-variable comparison of face and vote. In the next section, we control for district competitiveness in an effort to estimate the direct effect of facial quality on voting behavior.

INDIVIDUAL VOTER RESPONSE TO CANDIDATE FACES

In this section, we identify the effect of candidate face on individual vote choice. Research using laboratory experiments demonstrates that candidate facial qualities influence attitudes about candidates,[8] and observational analysis shows a strong aggregate relationship between facial competence and electoral outcomes. Yet neither demonstrates that voters directly respond to candidate faces in actual elections. On the one hand, experiments provide empirical support for an individual-level process that may operate in experimental settings where sufficient exposure is ensured, but may lack external validity. On the other hand, aggregate post-election comparisons show that facial competence is associated with real-world election outcomes but are unable to determine whether this association is caused by individual voter response or by incidental correlations produced by the elite-level selection processes.

To estimate the effect of candidate face on vote choice at the individual level in actual elections, we merge our estimates of individual candidate facial competence to exit poll election surveys. Using surveys allows us to control for the possibility that the influence of candidate face on voters is conditioned by voter partisanship. Ballew II and Todrov (2007) suggest that competent appearance is most likely to affect the behavior of people without strong partisan attachments. In addition to strength of partisanship, the disparity of available information for evaluating incumbents and challengers may affect voter response to candidate characteristics (Green and Krasno 1988, Jacobson 1990). Competent appearance might be less influential in the evaluation of incumbent candidates with congressional track records than in the evaluation of challenging candidates without congressional track records. To account for this disparity we code our dependent variable as incumbent vote rather than partisan vote, omitting open-seat races.[9]

Surveys also allow us to better control for electoral competitiveness by including both Cook incumbent risk and individual respondent partisanship in our analysis. By putting the partisanship of respondents in the model we are able to approximate district partisan composition, which is arguably the most important measure of district competitiveness. We attempt to account for other characteristics that contribute to competitiveness, in addition to district partisan composition, by including Cook ratings as an additional covariate.

We model the House and Senate incumbent vote choice of exit poll respondents as a function of respondent characteristics, district competitiveness, and our measure of individual candidate facial competence. We measure whether the respondent shares the party of the challenger, shares the party of the incumbent, or identifies as an independent. We control for contest competitiveness and incumbent risk by coding each race according to the classifications provided by the *Cook Political Report*, and include measures of incumbent tenure and age and the square of each as potential correlates

---

8 For an interesting series of work, see Rosenberg *et al.* (1986), Rosenberg and McCafferty (1987), and Rosenberg *et al.* (1991).

9 In an analysis of a small number of open-seat races only, we find results consistent with those reported here: facial competence has a small but evident effect on vote choice.

of face evaluations as well as vote outcomes.[10] All standard errors are clustered on the state or district. We also present a more robust specification that includes controls for challenger expenditures and, in Senate models, state population and year fixed effects. In the Appendix we present a linear probability model version of all estimates (Table A1).[11]

In Table 2 we present the results from a probit regression of individual incumbent vote choice on challenger and incumbent facial competence. Of the eight coefficients estimating the effect of challenger and incumbent facial competence on vote choice, only for the effect of Senate challenger face does the 95 percent confidence interval exclude zero.

The coefficient from the ordinary least squares version of the model (Table A1) suggests that increasing a Senate challenger's face by one standard deviation (one unit) increases challenger vote probability by 1.8 points, all else equal. The House model indicates that increasing the challenger's face by one standard deviation would increase challenger vote by 2.4 points (1.4 points in the specification that includes challenger expenditures), though this point estimate is not estimated with much precision. The probit coefficients in Table 2 exhibit the same patterns of effect and uncertainty as the linear probability versions.

The coefficients estimating the effect of incumbent faces on incumbent share are all essentially zero, suggesting that incumbent facial competence has little effect on the election outcome when controlling for the partisanship of the voter and the competitiveness of the district. We do not interpret this to mean that incumbent faces do not matter, but rather that the positive benefits of incumbency plus the selection effect of faces during each incumbent's initial election are more important than the small effect of candidate faces. In fact, since many incumbents were at one point also challengers, our theory of candidate face selection suggests a negative relationship between incumbent vote and incumbent face quality. The incumbents from the most competitive districts would have higher facial quality than incumbents from the most safe incumbent districts due to the selection process of better faces to competitive districts, inducing a negative relationship between incumbent face and incumbent vote. We have found some initial evidence to this effect, but due to uncertainty in estimates, we leave the incumbent effects to the side in this study.

As a substantive interpretation of the probit challenger face effect results, we simulate predicted values given the estimated coefficients from the parsimonious models in columns 1 and 3. We hold the incumbent's facial competence at the 50th percentile, hold the Cook report of district competitiveness at likely going to the incumbent party, and

---

[10] An improved specification of this model would include a measure of exposure to political messages. Research does find that level of exposure can affect response to candidates' appearance (Lenz and Lawson 2007). With the exit poll data, however, there is no adequate measure of political exposure. Education could be used as a proxy for exposure, but education level also affects attitude stability and so has potentially conflicting effects (Zaller 1992). We therefore exclude education in models presented; including education does not substantively change results.

[11] We estimated, but do not present, a pooled version of the model including both Senate and House vote responses in one combined specification. Results are highly similar to those of the Senate-only models.

Table 2. The effect of candidate facial competence and partisanship on incumbent vote choice.

	House 2004	House 2004	Senate 1992–2006	Senate 1992–2006
Intercept	0.336	0.293	−1.270	−0.812
	(0.671)	(0.694)	(1.255)	(1.595)
Cook incumbent risk	−0.139	−0.116	−0.110	−0.116
	(0.061)	(0.065)	(0.025)	(0.026)
Respondent shares	−1.311	−1.313	−1.037	−1.040
challenger party	(0.081)	(0.081)	(0.043)	(0.042)
Respondent shares	1.392	1.393	1.080	1.080
incumbent party	(0.090)	(0.090)	(0.036)	(0.035)
Challenger facial	−0.119	−0.080	−0.066	−0.062
competence	(0.095)	(0.098)	(0.026)	(0.027)
Incumbent facial	−0.024	−0.018	−0.003	−0.009
competence	(0.128)	(0.128)	(0.032)	(0.037)
Incumbent tenure	−0.009	−0.005	−0.015	−0.013
	(0.016)	(0.018)	(0.010)	(0.012)
Tenure squared	−0.000	−0.000	0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.000)
Incumbent age	−0.020	−0.021	0.058	0.042
	(0.030)	(0.031)	(0.043)	(0.053)
Age squared	0.000	0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Challenger expenditures (logged)		−0.022		0.016
		(0.019)		(0.019)
State population (millions)				−0.007
				(0.008)
1994 Fixed effect				0.014
				(0.094)
1996 Fixed effect				0.044
				(0.080)
1998 Fixed effect				0.059
				(0.078)
2000 Fixed effect				0.030
				(0.085)
2002 Fixed effect				0.121
				(0.114)
2006 Fixed effect				0.002
				(0.135)
N	4250	4250	26454	26454
AIC	3372.005	3372.261	25329.216	25312.920

Probit regression coefficients with standard errors in parentheses. Dependent variable is respondent vote for incumbent candidate. Cook Incumbent Risk is coded from zero for contests classified as "safe" to 3 for contests classified as "tossup." Robust standard errors clustered on state/district.

set incumbent age and tenure at their means. For each chamber and for each of three respondent partisan types (challenger co-partisans, incumbent co-partisans, and independents) we estimate the change in the predicted probability that a respondent votes for the challenger candidate when the challenger candidate's facial competence score is moved from the 25th percentile to the 75th percentile of all candidate faces.[12] We present in Figure 1 a graphical representation of the estimated effect for voters in House and Senate elections.

In the Senate, the increase in predicted probability of challenger vote from increasing the challenger's face from the 25th (−0.61) to the 75th percentile (0.73) is 3.5 points for independents with a 95 percent confidence interval of [0.3, 6.8]. For challenger co-partisans, the effect is 2.5 points with a 95 percent confidence interval of [0.4, 4.6] and for incumbent co-partisans 1.6 points [0.1, 3.1].

In the House, the uncertainty around the challenger coefficient is reflected in first difference confidence intervals that span zero. The increase in predicted probability of challenger vote from increasing the challenger's face from the 25th percentile (−0.18) to the 75th percentile (0.39) of all House candidates is 2.7 points for independents with a 95 percent confidence interval of [−1.8, 7.4]. For challenger co-partisans, the effect is 1.2 points with a 95 percent confidence interval of [−0.8, 3.5] and for incumbent co-partisans 1.0 points [−0.7, 2.9].

CHALLENGER EFFECTS BY CONTEST

Publication of Todorov *et al.*'s (2005) study was met by substantial speculation in the mass media about the extent to which candidate faces affect real-world election outcomes.[13] Our cross-district analysis enables us to assess the extent to which candidate faces influence actual vote outcomes in the 99 Senate races in our study.[14]

We can use our model to estimate how much election vote share would differ if the challenger's facial appearance score were different. For example, John Thune, who had the highest recorded facial appearance score of our candidates, defeated Tom Daschle by 1.1 percentage points in 2004. If Thune had more ordinary facial appearance — say, the facial appearance of the median congressional candidate, would he have still defeated Daschle? In this section, we answer this question for all challengers in our study by comparing their actual vote share with the share we predict they would have garnered if they possessed ordinary facial appearance.

[12] For all simulated effects presented we take repeated random samples of model coefficients from the multivariate normal coefficient distribution with the clustered variance–covariance matrix. For each sampled vector of coefficients we calculate the predicted vote given our hypothetical respondent and control values. The means and percentiles across coefficient samples become the point estimate and confidence intervals of our quantities of interest.

[13] For example, "Faces Decide Elections" (Skloot 2007) or "Scientists Search For That Winning Look" (Hamilton 2005).

[14] We do not consider the effects in our House contests due to the larger confidence intervals on face coefficients estimated in House models.

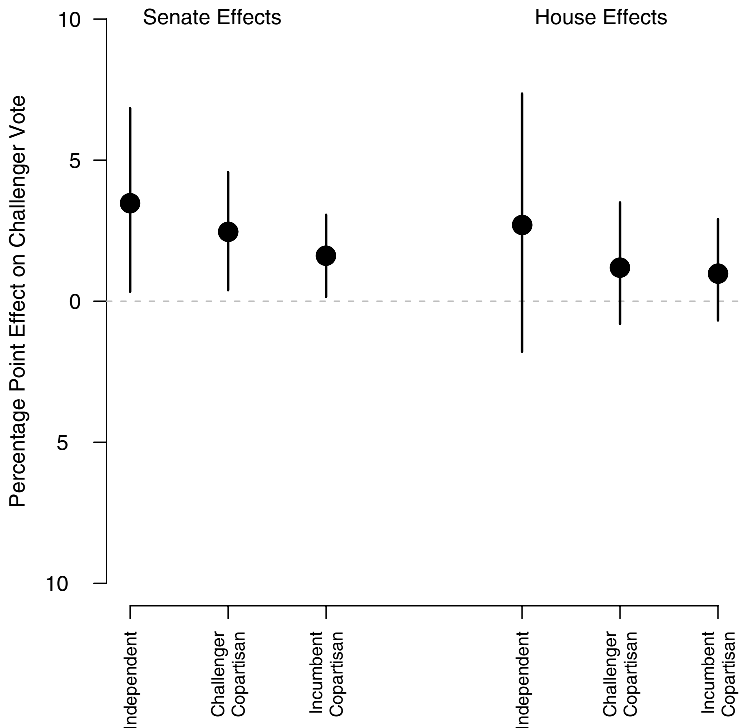

Figure 1. Estimated effect on challenger vote probability of increasing challenger facial competence, by respondent partisan affiliation with 95 percent confidence intervals. *Each point represents the estimated difference in challenger vote probability moving the challenger's face from the 25th percentile to the 75th percentile, holding incumbent age and tenure at their means, the Cook report at likely going to the incumbent, and the incumbent's face at the chamber median. The points represent the average estimate across 500 samples from the clustered coefficient distribution, and the lines the 2.5th percentile to the 97.5th percentile of the sampled effects.*

There is no intuitive baseline for measuring the absolute effect of facial competence. What follows is a simple test of whether candidate facial appearance has a widespread influence on election outcomes. If facial appearance affects a significant number of election outcomes, then a significant number of election outcomes should change when we use our model to predict counterfactual election outcomes with challengers having the facial appearance of the median candidate, everything else held constant.

For each election, we simulate the effect of moving the challenger's face from its actual score to the median face score of all Senate candidates. Hence in cases where the challenger's face is above the median, this change in the face score benefits the incumbent since the counterfactual has a worse challenger face, and in cases where the challenger's face is below the median, the change benefits the challenger since the counterfactual has a better challenger face. We simulate the effect for independent, challenger co-partisan, and incumbent co-partisan voters in each race, and then create a weighted contest face effect using the statewide partisanship estimates of Wright *et al.* (N.d.).

For example, in New Mexico's 2000 Senate contest, the challenger had a face score of 0.74. Relative to a counterfactual contest where the challenger has a median face score (0.17), our model predicts that the challenger's actual face increases propensity to support the challenger by 1.46 points among independents, 0.84 points among incumbent co-partisans, and 0.83 points among challenger co-partisans, all else equal. Wright *et al.* (N.d.) estimate New Mexico's 2000 partisan proportions as 27 percent independent, 39 percent incumbent co-partisan, and 34 percent challenger co-partisan. The estimated effect of the actual challenger's face relative to the median face, therefore, is $0.27*1.46 + 0.39*0.84 + 0.34*0.83 = 1.0$ points of challenger vote share. In this example the challenger's real face is above the median face so we estimate that the actual challenger's face helped the challenging party. But in many races the challenger's face is below median, so our estimates show a negative effect of these challengers' faces.

In Figure 2 we present the estimated challenger face effect from each Senate election in our sample. The x-axis is the estimated effect of challenger face on challenger vote share relative to the median face, and the y-axis identifies each of 99 races sorted by the effect from the largest pro-challenger effect at the top to the largest pro-incumbent effect on the bottom.

In three, and almost four, contests the estimated effect of the challenger's face was larger than the margin of victory in the actual vote outcome. The three contests are Nevada 1998 with an incumbent victory by 0.1 points, Washington 2000 with an incumbent loss of 0.1 points, and South Dakota 2002 with an incumbent victory of 0.2 points. This suggests that in roughly three to five percent of Senate contests the quality of the challenger's face has a reasonable ability to affect the outcome of the election. In all three of these races, however, the direction of the face effect was opposite to that of the margin, meaning a median candidate face would not have changed the outcome. For example, in Maria Cantwell's (D) 2000 race against incumbent Slade Gordon (R-WA), we estimate that Cantwell's below median facial quality came close to costing the out-party Democrats the election. Cantwell ultimately won by 2,200 votes or 0.1 percentage points, but our simulations suggest that an alternative Democratic candidate of median facial quality would have won by just over 2 points, all else equal.

Our simulations indicate that a fourth contest, Rick Santorum's 1994 victory in Pennsylvania, would have been very close to a different outcome had the Republican challenger had a median candidate face rather than Santorum's above average face. We estimate that Santorum's face was worth 2.1 points, and he won election by a margin of 2.5 points, a difference certainly smaller than our estimation model's precision. Had the Pennsylvania Republicans run a candidate with a median face rather than Santorum, we estimate they may have lost the election.

The biggest positive challenger face effect in our study is produced by John Thune in his 2002 race against incumbent Tim Johnston (D-SD), an estimated pro-challenger

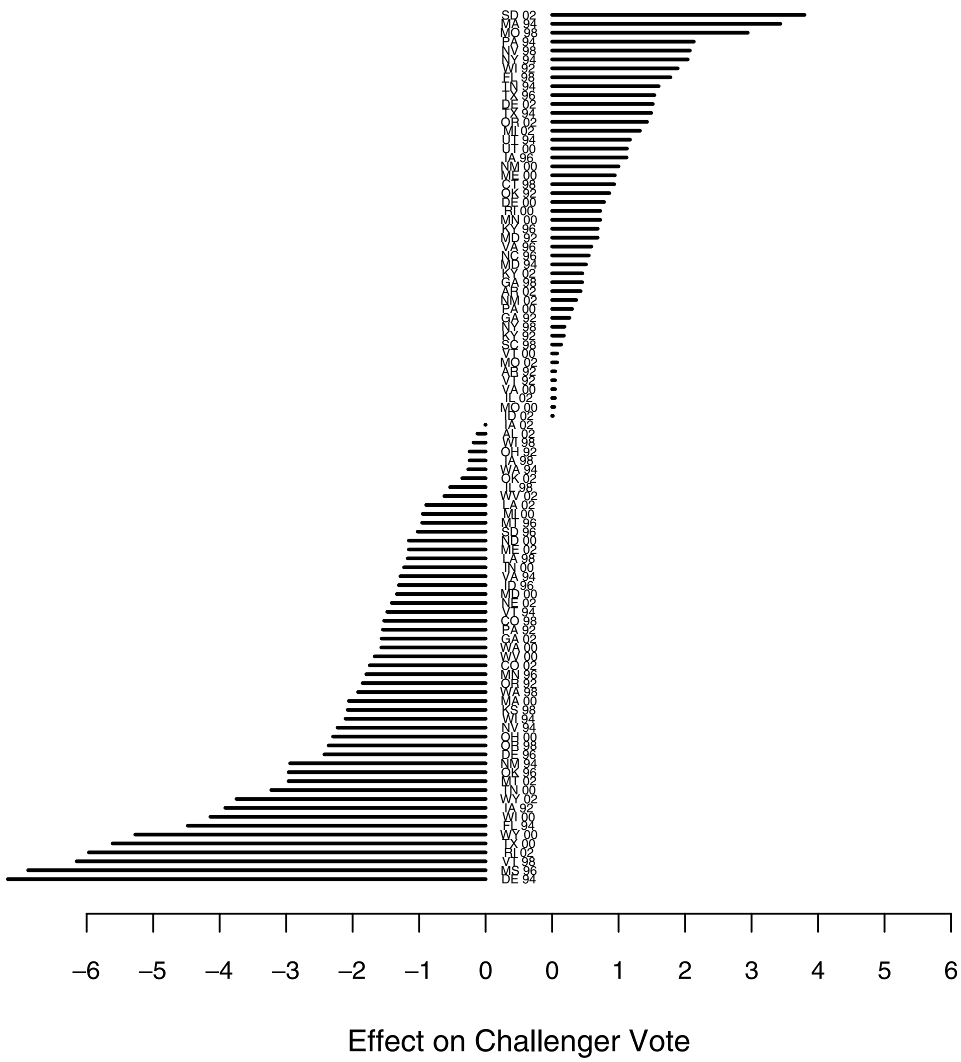

Figure 2. Estimated effect on challenger vote share moving challenger face from median Senate candidate face to actual challenger face, by Senate contest.
Each line represents the estimated difference in vote share between an election with the actual challenger's face and a hypothetical election with a challenger face at the median of all Senate candidates. Each election outcome is estimated using the results from the probit model in Table 2, column 3, weighted by the statewide partisan proportions estimated by Wright et al. (N.d.).

effect of almost 4 percentage points. While Thune did not unseat Johnston, two years later he did defeat the Democratic Senate Majority Leader, Tom Daschle, in a contest decided by only 4,500 votes or 1.1 percentage points. We suspect that Thune's 2004 victory is an instance of facial quality being consequential in our time period. Unfortunately the 2004 South Dakota contest is not in our study because we omitted high-profile politicians such as Daschle from our sample of candidate faces to avoid potential recognition by our participants.

While many of the challenger faces in our study have estimated effects that could swing close elections, our results suggest that none of the 99 contests in our sample were decided by a challenger's face.

DISCUSSION

We believe that candidate facial characteristics are one part of a broader explanation about the role of candidate trait advantages in elections. Although we have argued that individual voters appear to respond to a candidate's face and that candidates select to districts based upon appearance, there is a potential confounding factor. The competence of faces may be correlated with other desirable candidate qualities such as wealth and human capital, unmeasured in our model. Research in economics and psychology suggests that facial characteristics are associated with a variety of life outcomes. It is possible that candidates with other desirable traits select into competitive districts, and that quality faces are incidental due to its correlation with these traits. We believe this outcome is not inconsistent with a more general story: as higher quality challenging candidates select to the most competitive districts, a correlation between the qualities selected and the election outcome will be induced even before voters become involved.

Our research has important further implications. Who are the voters that might be persuaded by candidate faces, and what percentage of the full electorate do they comprise? As American demographics change during the 21st century and the candidates for national office become more diverse, the influence of candidate appearance on vote choice will merit more study. Recent elections have brought the race and gender of national candidates to the forefront, and so future research should consider if race- and gender-based voting might partially be a function of trait inference based upon physical appearance. We have created an easy-to-implement web-based technique that can be used to capture the characteristics of a large number of candidate faces for any trait of interest. Although our pool of candidates drawn from the U.S. House and Senate is mainly white males, the procedure could easily be extended to candidates or faces from other demographic groups to consider whether there is variation in the rapid inference of traits when other groups are used.

It would also be valuable to investigate the extent to which candidate facial characteristics are related to other forms of candidate quality, both in terms of political skill and with respect to general human capital. A further question of interest is what exactly humans consider when they infer a politician face to be competent. What actual attributes do voters hope a competent face will bring to representation? Does the inference of competence vary by the type of office being sought, for example might executive competence and legislative competence be different?

Despite a focus in the mass media on the importance of candidate characteristics in election outcomes, we have demonstrated that part of the relationship between candidate face and election outcomes is the selection of candidate faces to competitive districts. This selection may create the illusion that candidate traits such as facial appearance drive a large part of congressional election results. Although the effect of face is large enough to potentially swing close elections such as the 1994 Pennsylvania Senate race, expectations about the election context well in advance of election day are important contributors to the candidate choices available to the electorate.

APPENDIX

We conducted two separate surveys to gather evaluations of candidate facial characteristics. In the first, 296 students in a lower division political science class at UCLA evaluated images of white male candidates from the 2004 House elections.[15] In the second survey, 349 students from an upper division UCLA political science class evaluated images of 1990–2006 Senate candidates, of all races and genders, and, separately, the 2004 House candidates from the first survey. We used images of candidate faces provided to us by Todorov *et al.* (2005); for other candidates, we followed the methodology described in Todorov *et al.* (2005) of obtaining pictures from CNN.com and supplementing them as necessary with pictures from other Internet sites. We standardized these photos in size and pixel count, turned all to black and white, and added a standard gray background. See below for a discussion of our efforts to purge the pictures of quality that could be associated with candidate traits.

Before the first survey, participants were asked to identify the photo of the Member of Congress for the UCLA area from a lineup of photos containing members of the California Assembly as a test of candidate recognition. The participants recognized Henry Waxman at levels barely better than chance. Following the survey of Senator faces, participants were asked to identify the faces of Senators from the current Senate that they recognized. Prior to estimation, we removed the evaluations in which a participant claimed to recognize the face. In both surveys, images of individuals that we felt had a high probability of recognition, such as members of the leadership, presidential candidates, and those with high-profile scandals were not included.

[15] We limited our initial survey to this subset of candidates because we were unsure of the number of evaluations needed to get a precise measure of competence. When we determined the effectiveness of the survey and estimation procedure, we were able to add more faces into the second survey.

Details of the Estimation Process

The scores are estimated by a method used to model congressional committee choice (Groseclose and Stewart III 1998). We assume each candidate face i has a location c_i on a latent competence continuum. Each participant evaluates two faces i and j with some amount of measurement error ϵ_i and ϵ_j. Participants choose face i over face j if and only if $c_j + \epsilon_j < c_i + \epsilon_i$ which is the same as $\epsilon_j - \epsilon_i < c_i - c_j$. Without loss of generality, we assume that the ϵ_i are identically and independently distributed according to a mean-zero normal distribution with standard deviation σ. Given these assumptions, the probability that the respondent reports candidate i more competent than candidate j is

$$\Phi\left(\frac{c_i - c_j}{\sigma\sqrt{2}}\right)$$

where $\Phi(\cdot)$ is the cumulative normal probability function.

To estimate the c_i we let each observation be the evaluation by one participant of two faces. We define an indicator matrix V, with K rows of observations and I columns of faces, where each element v_{ki} takes the value 1 if the *ki*th face is selected the more competent, -1 if the *ki*th face is not selected the more competent, and zero if the *ki*th face is not evaluated. For each observation k, therefore, the probability that the respondent evaluated the face pair randomly presented in the way that they did is

$$\Phi\left(\frac{\sum_{i=1}^{I} c_i v_{ki}}{\sigma\sqrt{2}}\right).$$

This implies a likelihood function for parameters c given data V:

$$L(c|V) = \prod_{k=1}^{K} \Phi\left(\frac{\sum_{i=1}^{I} c_i v_{ki}}{\sigma\sqrt{2}}\right).$$

Again following Groseclose and Stewart III (1998) we implement this estimation with an intercept-free probit model, with the number of explanatory variables equal to the number of candidates. As with standard probit estimation, we set σ to 1, which means that our estimates of the c_i are in units of σ. For identification, one face is set to the value of 0 on the competence dimension. After implementing the estimation, the probit coefficients are the measures of facial competence.

With the estimated competence scores, we can calculate the probability that any one face will be chosen the more competent over any other face by our participants by reversing the estimation model. Specifically, the probability that a participant will choose candidate i more competent than candidate j given estimated competence scores c_i and c_j is:

$$\Phi\left(\frac{c_i - c_j}{\sqrt{2}}\right).$$

The quantity σ is ignored as it is fixed at 1 in the estimation.

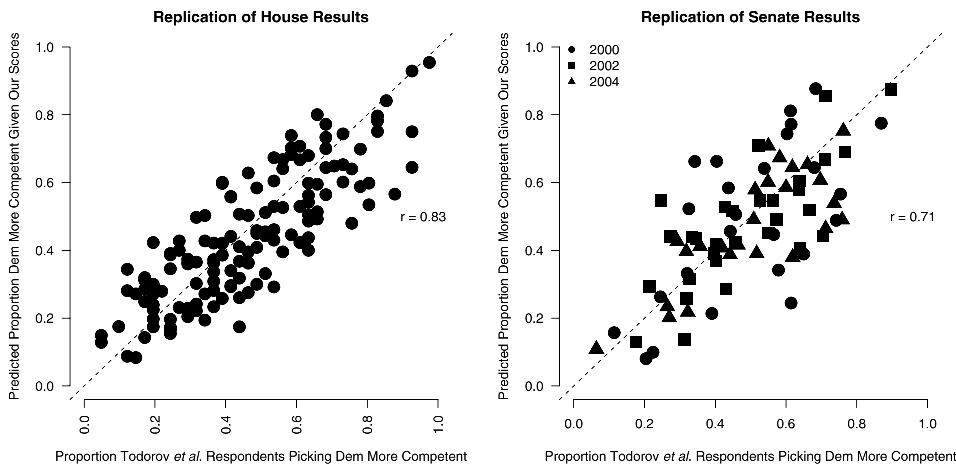

Figure 3. Replication of Todorov *et al.* experimental results.

Consistency with Previous Measures

We were able to closely replicate Todorov *et al.*'s (2005) measure of facial competence using evaluations from our survey. We present our replication of House and Senate evaluations in Figure 3. In each frame, the x-axis plots the proportion of the Todorov *et al.* 2005 participants who chose the Democratic candidate more competent, and the y-axis the predicted proportion of pairwise evaluations in which the Democrat would be picked the more competent by our participants. The dashed line is a 45-degree line indicating perfect correspondence. Our method effectively replicates the choices of the Todorov *et al.* (2005) participants. Our scores also reproduce the relationship between facial competence and aggregate vote share (not presented).

Robustness of Estimation in Measurement

We estimated the scores with a variety of robustness checks based upon recognition, respondent consistency (we repeated the same face pairs within respondents, varying left–right status of the repeat pair), and dropping early and late evaluations for fatigue and learning. None of the alternatively estimated scores substantively affected our results.

We were also concerned that the evaluations of candidate facial competence are affected by the quality of the photographs used to present the faces. A casual look at the candidate images will reveal that some are of higher quality than others. That some photos are taken with more high-quality cameras or simply produced by better photographers could be a reflection of the quality of the candidate and the campaign. It is possible then that participant responses are not based solely on qualities of the candidate's face but also to the qualities of the candidate's image. This would present a difficulty for accurate estimates of the role of facial competence alone.

To determine how much of election results are directly attributable to face, we attempt to control for image quality. We constructed a unique measure of photo quality derived from the variance of pixel shade at different points in an image. We find that image quality is related to individual's evaluations of the competence of the faces in the images. However, we find that even when controlling for image quality, evaluations are still related to election outcomes and our substantive results are unchanged.

We detail here our efforts to objectively measure image quality. We considered hand-coding by eye each image for quality, but were concerned the facial characteristics would influence such an effort. Instead, we implemented an objective measurement based on spatial statistics using each pixel in the image.

We first constructed a *variogram* of each image. The variogram measures the variance γ between the value of a process k at an arbitrary point s_i in a space S and every other point $s_{i+1}, \ldots, s_n$ in S that is at each distance $d_0, d_1, \ldots, d_m$ from s_i, where m is usually $1/3$ the maximum possible distance between points (for details on variogram estimation, see Cressie 1993, Ripley 2004). In this case, S is the pixel matrix of the image.

A black and white digital image has for each pixel a number representing the place on the gray scale of that point on the image, ranging from 0 to 255. Each photograph was standardized to a matrix of 105×147 pixels. In constructing the variogram, we first median-polished the matrix so that each pixel was the residual after the median tendency of each row and column of the matrix had been removed.

Using this variogram, we constructed an image quality measure, I, for each photograph such that:

$$I = \frac{\gamma d_1}{\gamma d_{49}}.$$

This image quality measure is designed to capture the average variance of any two adjacent pixels given the average variance of two pixels at a distance in which the pixels would be expected to have no autocorrelation. This is based on the assumption that adjacent pixels are autocorrelated because they are capturing the same feature of an image. For example, in a photograph of a man, two adjacent pixels might both be capturing the man's tie and should be the same color. As pixels are further apart, the autocorrelation goes to zero and pixels have no dependence on each other. We assume that images of higher quality have more autocorrelation between pixels because of higher original density of the photograph. However, this is relative to the overall variance of the photograph, for example the man might wear a solid tie (low variance) or a patterned tie (high variance), so the denominator of the image quality measure represents the variance at a point in which no autocorrelation is expected to exist.

Results presented in this paper are based upon competence scores purged of this image quality. We regress competence scores on the image quality measure I and use the residuals for our analysis. We standardized the residuals to mean zero and unit variance.

Alternative Estimates of Vote Model

In Table A1 we present a linear probability model of the exit poll analysis reported in Table 2.

Table A1. Using candidate facial competence and partisanship to predict individual-level vote choice.

	House 2004	House 2004	Senate 1992–2006	Senate 1992–2006
Intercept	0.605	0.593	0.212	0.330
	(0.155)	(0.159)	(0.338)	(0.427)
Cook incumbent risk	−0.032	−0.026	−0.030	−0.031
	(0.014)	(0.015)	(0.007)	(0.007)
Respondent shares challenger party	−0.431	−0.432	−0.371	−0.371
	(0.025)	(0.025)	(0.015)	(0.015)
Respondent shares incumbent party	0.379	0.378	0.324	0.324
	(0.025)	(0.025)	(0.013)	(0.012)
Challenger facial competence	−0.024	−0.014	−0.017	−0.016
	(0.021)	(0.022)	(0.007)	(0.007)
Incumbent facial competence	−0.005	−0.004	−0.001	−0.002
	(0.028)	(0.028)	(0.009)	(0.010)
Incumbent tenure	−0.001	−0.001	−0.004	−0.004
	(0.004)	(0.004)	(0.003)	(0.003)
Tenure squared	−0.000	−0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Incumbent age	−0.005	−0.005	0.014	0.010
	(0.007)	(0.007)	(0.012)	(0.014)
Age squared	0.000	0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Challenger expenditures (logged)		−0.005		0.004
		(0.004)		(0.005)
State population (millions)				−0.002
				(0.002)
1994 Fixed effect				0.004
				(0.026)
1996 Fixed effect				0.014
				(0.022)
1998 Fixed effect				0.016
				(0.022)

(*Continued*)

Table A1. (*Continued*)

	House 2004	House 2004	Senate 1992–2006	Senate 1992–2006
2000 Fixed effect				0.006
				(0.023)
2002 Fixed effect				0.030
				(0.031)
2006 Fixed effect				−0.002
				(0.037)
N	4250	4250	26454	26454
Adjusted *R*-squared	0.496	0.496	0.372	0.372
Std. error of regression	0.351	0.351	0.393	0.393

Ordinary least squares regression coefficients with standard errors in parentheses. Dependent variable is respondent vote for incumbent candidate. Cook Incumbent Risk is coded from zero for contests classified as "safe" to 3 for contests classified as "tossup." Robust standard errors clustered on state/district.

REFERENCES

Antonakis J., and O. Dalgas. 2009. "Predicting Elections: Child's Play!" *Science* 323(5918): 1183.

Ballew II C. C., and A. Todorov. 2007. "Predicting Political Elections From Rapid and Unreflective Face Judgments." *Proceedings of the National Academy of Sciences*.

Benjamin D. J., and J. M. Shapiro. 2006. "Thin-Slice Forecasts of Gubernatorial Elections." *National Bureau of Economic Research Working Paper Series* (Working Paper No. 12660).

Berggren, N., H. Jordahl, and P. Poutvaara. 2006. "The Looks of a Winner: Beauty, Gender and Electoral Success." Discussion Paper No. 119/September 2006. Helsinki Center of Economic Research, Discussion Papers, ISSN 1795–0562.

Biddle J. E., and D. S. Hamermesh. 1998. "Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre." *Journal of Labor Economics* 16(1): 172–201.

Carson, J. L. 2005. "Strategy, Selection, and Candidate Competition in U.S. House and Senate Elections." *Journal of Politics* 67(1): 1–28.

Cook, C. E. 1992–2006. *The Cook Political Report*. Washington, D.C.: National Political Review, Inc.

Cressie, N. A. C. 1993. *Statistics for Spatal Data*. New York: Wiley-Interscience.

Gimple, J. G., K. A. Karnes, J. McTague, and S. Pearson-Merkowitz. 2008. "Distance-Decay in the Political Geography of Friends-and-Neighbors Voting." *Political Geography* 27: 231–252.

Green D. P., and J. S. Krasno. 1988. "Salvation for the Spendthrift Incumbent: Reestimating the Effects of Campaign Spending in House Elections." *American Journal of Political Science* 32(4): 884–907.

Groseclose, T., and C. Stewart III. 1998. "The Value of Committee Seats in the House, 1947–1991." *American Journal of Political Science* 42(2): 453–474.

Hamermesh, D. S. and J. E. Biddle. 1994. "Beauty and the Labor Market," *American Economic Review* 84(5): 1174–1194.

Hamilton, J. 2005. "Scientists Search for that Winning Look." *National Public Radio All Things Considered* (June).

Jacobson, G. C. 1989. "Strategic Politicians and the Dynamics of U.S. House Elections, 1946–1986." *American Political Science Review* 83(3): 773–793.

Jacobson, G. C. 1990. "The Effects of Campaign Spending in House Elections: New Evidence for Old Arguments." *American Journal of Political Science* 34(2): 334–362.

Jacobson, G. C. and S. Kernell. 1983. *Strategy and Choice in Congressional Elections*. New Haven: Yale University Press.

Lenz, G., and C. Lawson. 2007. "Looking the Part: Television Leads Less Informed Citizens to Vote Based on Candidates' Appearance." Paper presented at the Annual Meetings of the American Political Science Association.

Little, A. C., R. P. Burriss, B. C. Jones, and S. C. Roberts. 2007. "Facial Appearance Affects Voting Decisions." *Evolution and Human Behavior* 28(1): 18–27.

Mobius, M. M., and T. S. Rosenblat. 2006. "Why Beauty Matters." *American Economic Review* 96: 222–235.

R Development Core Team. 2009. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Ripley, B. D. 2004. *Spatal Statistics*. New York: Wiley-Interscience.

Rosenberg, S. W., L. Bohan, P. McCafferty, and K. Harris. 1986. "The Image and the Vote: The Effect of Candidate Presentation on Voter Preference." *American Journal of Political Science* 30(1): 108–127.

Rosenberg, S. W., and P. McCafferty. 1987. "The Image and the Vote: Manipulating Voters' Preferences." *Public Opinion Quarterly* 51(1): 31–47.

Rosenberg, S. W., S. Kahn, and T. Tran. 1991. "Creating a Political Image: Shaping Appearance and Manipulating the Vote." *Political Behavior* 13(4): 345–367.

Rule, N. O., and N. Ambady. 2008. "The Face of Success: Inferences From Chief Executive Officers' Appearance Predict Company Profits." *Psychological Science* 19(2): 109–111.

Skloot, R. 2007. "Faces Decide Elections." *New York Times Magazine* (December).

Squire, P. 1992. "Challenger Quality and Voting Behavior in U. S. Senate Elections." *Legislative Studies Quarterly* 17(2): 247–263.

Squire, P., and E. R. Smith. (1996). "A Further Examination of Challenger Quality in Senate Elections." *Legislative Studies Quarterly* 21(2): 235–248.

Todorov, A., A. N. Mandisodza, A. Goren, and C. C. Hall. 2005. "Inferences of Competence from Faces Predict Election Outcomes." *Science* 308(5728): 1623–1626.

Vavreck, L. 2001. "The Reasoning Voter Meets the Strategic Candidate: Signals and Specificity in Campaign Advertising, 1998." *American Politics Research* 29(5): 507–529.

Wright, G. C., J. P. McIver, and R. S. Erikson. N.d. "Aggregated CBS News/New York Times National Polls." Electronic File.

Zaller, J. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.