

Candidate Faces and Election Outcomes*

Matthew Atkinson Ryan D. Enos Seth J. Hill†

March 5, 2008

Abstract

Recent research finds that naive survey participants' rapid evaluations of the facial competence of United States Congressional candidates predict aggregate vote margins. The predictive power of facial competence has generated considerable interest because it seems to indicate a causal relationship between face and vote choice. Because there is no *a priori* reason to expect that candidate facial qualities are randomly distributed across electoral contests, we estimate the effect of facial competence using multiple regression analysis with controls. Using data from a survey we implemented with evaluations of more than 167,000 pairs of candidate faces, we find evidence that facial competence has a small but significant causal effect on vote choice; we also find evidence that candidates with high facial competence participate in the most competitive electoral contests. We argue that faces are an interactive part of a bigger political story that structures election results.

*The authors thank Mike Franks, Gary Jacobson, R. Brian Law, Jeff Lewis, Elisabeth Michaels, David Sears, Alexander Todorov, Lynn Vavreck, John Zaller, and participants at the UCLA Workshop on Political Methodology.

†All authors: Department of Political Science, University of California, Los Angeles; matthewa@ucla.edu, renos@ucla.edu, sjhill@ucla.edu.

Recent work in psychology demonstrates that naive, rapid evaluations by survey participants of the facial competence of candidates linearly predict United States Congressional candidate vote share (Todorov, Mandisodza, Goren, & Hall 2005). Todorov et al. (2005) showed college students the faces of candidates contesting United States House and Senate elections in 2000, 2002, and 2004, and asked them to unreflectively choose the more competent looking candidate in each contest. They find that the candidate in each contest more frequently selected as appearing competent won the actual election in 66.8 percent of House and 71.6 percent of Senate contests. They also find that the proportion of paired evaluations in which one candidate’s face is judged more competent than the opponent’s correlates to the difference in votes between the two candidates in both the House ($r = .44$) and in the Senate ($r = .46$).

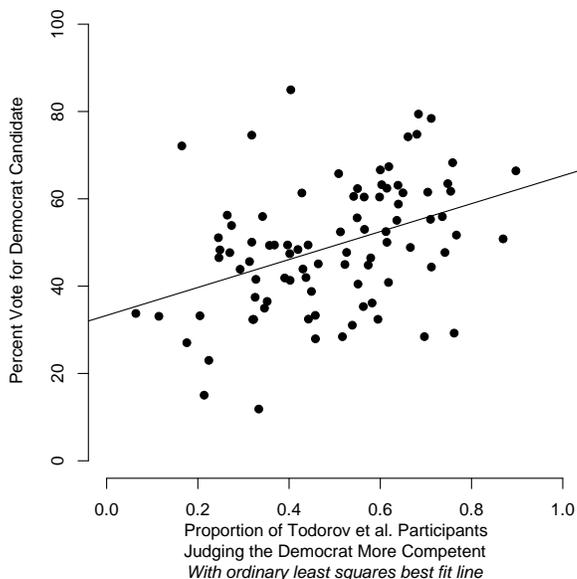
In Figure 1 we reproduce Todorov et al.’s (2005) main finding using participant evaluations to predict Democratic vote share.¹ Similar findings have subsequently shown that candidate faces predict gubernatorial elections (Benjamin & Shapiro 2006, Ballew II & Todorov 2007) and that executive faces predict corporate profits (Rule & Ambady 2008). The relevance of facial cues has previously been demonstrated in research on the effect of candidate appearance on voter emotions (McHugo, Lanzetta, Sullivan, Masters, & Englis 1985, Sullivan & Masters 1988).

Although Todorov et al. (2005) were careful to discuss their results as predictive rather than causal, there has been a tendency among popular commentators and some scholars to interpret the results as an indictment of the ability of American voters to make politically-reasoned decisions. *The New York Times*, writing about Little, Burriss, Jones, & Roberts (2007), declared that “Faces Decide Elections” (Skloot 2007), and a *National Public Radio* segment suggested we “[f]orget political polls . . . voters prefer candidates who look competent, even if they are not” (Hamilton 2005). In fact, some have implied the need for electoral reform so that voters will not be able to see candidates’ faces.²

¹We thank Alex Todorov for generously sharing with us his data.

²“Understanding the nature and origins of appearance biases has real-world value, not the least of which

Figure 1: The Relationship of Vote Share to Inferred Competence, United States Senate Elections 2000-2004.



Todorov et al. (2005) use a randomized design for measuring latent candidate facial competence. This innovative approach to measuring a latent candidate characteristic may have led some secondary authors to perceive that the correspondence between facial competence and vote share was estimated based on an experimental study. In fact, the predictive results presented in Todorov et al. (2005) are based on observational analysis. Though the explanatory factor was measured in a laboratory setting, it was not randomly assigned to electoral contests. Because we do not know that candidate facial competence is randomly allocated across districts in actual elections, measuring the causal effect of facial competence on vote choice requires the use of statistical control.

In this paper we investigate the causal relationship between evaluations of candidate facial competence and election outcomes by introducing control variables. We find evidence that candidate facial competence has a small but significant effect on individual-level vote choice. However, despite the variable’s predictive power, its causal influence is small compared to other variables. Furthermore, we find evidence that political circumstances influence the

may be identifying electoral reforms that could increase the likelihood of electing the most qualified leaders rather than those who simply look the part (Zebrowitz & Montepare 2005).”

allocation of facial competence to electoral contests.

1 Incumbency, Partisanship, and the Congressional Vote

Political science research on elections generally emphasizes a standard set of causes for congressional election outcomes and vote choice such as incumbency, partisanship, and the economy. We designate these “political” factors. Perhaps most importantly, the incumbency status of a member of the House or Senate is a consistent electoral advantage. Nearly 90 percent of House incumbents who stand are reelected, and researchers have estimated that incumbency is worth 4 to 6 percentage points of vote share in House elections (Erikson 1971, Mayhew 1974, Gelman & King 1990, Jacobson 2004). At the level of the individual voter and of the district, party identification consistently predicts the presidential and congressional vote choice (Campbell, Converse, Miller, & Stokes 1960, Green, Palmquist, & Schickler 2002, Jacobson 2004).

In spite of the focus on political factors, American political behavior research has a long tradition of working to integrate the electoral effects of both political and candidate-specific factors. Most prominently, the authors of *The American Voter* provide a framework for understanding the vote choice that integrates political and candidate-specific factors in what they term the “funnel of causality” (Campbell et al. 1960, ch. 2). They describe voting behavior as an output from a “multitude of prior factors” that creates a “converging sequence of causal chains.” An analyst can work backwards from the actual vote event to causes of greater and greater temporal distance — from election day activities to long-term predispositions such as partisanship. The authors believed that each vote could be theoretically dissected into its constituent causal parts.

The “funnel of causality” framework implies that a factor influencing voting behavior should be studied in terms of its immediate direct effects and in terms of the causally prior factors that cause variation in the factor of interest. Because they fall near the end of the

funnel of causality, candidate faces may be influenced by prior causal factors. Jacobson & Kernell (1983), for example, suggest that the types of candidates who choose to run in a given election are caused by external political and economic conditions.

To identify the effect of candidate faces in the presence of a complicated process allocating faces to districts requires a measure of facial competence that enables comparison across contests and candidates. The measure of Todorov et al. (2005) does not provide this because it only compares the opposing candidate faces in each specific contest. We implement a new survey to measure the individual competence of a set of candidate faces, pairing each face against faces from many other contests. In other words, this is the difference between a single set of a head to head contests and a “round robin” tournament where each face goes against all others. This enables us to place each face in our pool of candidates on a common scale of facial competence and use these measures to assess the interplay of face and politics.

Our contribution to the understanding of elections is threefold. First, taking account of candidate face and political variables together, we find that political variables remain strong and consistent predictors of the congressional vote, that the effect of facial evaluations is much smaller than some interpretations have suggested, and that much of the remaining effect is driven by the face of the challenging candidate. Second, using our original measure of facial competence, we present evidence that candidate facial competence, incumbency, and partisanship are correlated in a way that suggests strategic behavior by political elites. Third, we use survey data to find a direct effect of candidate face on the individual vote choice. But we show that the direct effect of face is small and uncertain, and is partially conditional on the partisan identification of the voter and the incumbency status of the candidate.

2 Candidate Faces in Congressional Elections

In this section, we estimate the effect of facial competence on vote choice. We begin our evaluation of the effect by comparing the bivariate relationship of facial competence to vote from Todorov et al. (2005) with models predicting vote share based on incumbency status and district partisanship. Following Todorov et al. (2005), we analyze data from the 2000, 2002 and 2004 Senate elections and the 2004 House elections.³ We operationalize difference in facial competence using the proportion of evaluations in which the Todorov et al. (2005) participants chose the Democratic candidate as more competent than the Republican candidate.

Because the method of Todorov et al. (2005) does not have experimental control with respect to election returns, the strong predictive power of the facial competence variable may not accurately reflect the true causal influence on vote choice. To be precise, the estimated effect of facial competence will be biased if some variable omitted from the model that causes election outcomes is correlated with the facial competence measure. The direction of the bias would depend upon the direction of correlation with the omitted variables. If the omitted variables are positively correlated with face and vote, the two-variable comparison would over-state the impact of face. On the other hand, if the omitted variables are negatively correlated with either face or vote, the comparison would under-state the impact of face on vote.

Participant ratings comparing the Democrat and Republican candidate faces are positively correlated with the political variables incumbency and partisanship. For both House and Senate candidates, the correlation between the proportion of Todorov et al. (2005) participants who choose the Democrat more competent and Democratic district partisanship is about $r = .2$. The correlation with a dichotomous indicator variable for a Democrat incumbent is about $r = .5$ and with a Republican incumbent indicator is about $r = -.4$.⁴

³Because we use presidential vote share to measure district partisanship, we omit the 2002 House elections due to redistricting.

⁴Correlations with incumbency are polyserial as the incumbency variable is dichotomous and the compe-

In order to identify the relative contributions of face, incumbency, and partisanship, we present regression models predicting Democrat vote share in Table 1. For each chamber, the first column is a model with facial competence difference only. Our regression model in column one replicates the impressive correlations between face and vote noted in Todorov et al. (2005). Moving from none of the Todorov et al. (2005) participants choosing the Democratic face more competent to all of the participants choosing the Democrat face more competent is associated with an estimated 29 point increase in Democratic vote share in the House, and an estimated 32 point increase in the Senate. These are surprisingly large effects.

However, the second and third columns for each chamber present models that also predict large differences in candidate vote share. Moving from a Democratic incumbent to a Republican incumbent is associated with a change in Democratic vote of almost 30 points in the House and of 25 points in the Senate. In the third column, the estimated effect of district partisanship is also large, with each additional point of presidential vote translating into about a point of House and Senate vote share. A first clue as to which predictors map more closely into congressional vote is found by comparing the R^2 statistic for the three models. In both House and Senate, the incumbency-only model and the partisanship-only model explain a much greater amount of variance in vote shares than the faces-only model.

In the fourth column for each chamber we present a multiple regression model to jointly estimate the effect of faces, incumbency, and partisanship. When controlling for these political factors, the point estimate for the effect of facial competence on vote share decreases more than 5-fold in the House and in the Senate. Both estimates also move out of range of standard levels for statistical significance. In both the House and Senate the estimated effect of incumbency and partisanship are both decreased, suggesting omitted-variable bias in the separately-specified models. Moving from a Democratic incumbent to a Republican incumbent is now worth about 23 points in the House and 21 points in the Senate, and presidential vote now translates about two-to-one into congressional vote. When controlling

tence variable is numeric (Fox 2007), while the correlation to district partisanship is Pearson's r .

for incumbency and district partisanship, the effect of facial competence is much-diminished and not statistically significant.

The results of Table 1 suggest that the bivariate comparison of face and vote suffers from omitted variable bias and that incumbency and partisanship remain important causes of the congressional vote. They do not necessarily indicate that faces do not matter, as in both cases the coefficient remains positive and of meaningfully substantive size even when controlling for political variables. However, the greater uncertainty in the estimate and the small effect compared to the effect of other variables undermines the claim that “faces decide elections.”

The correlation of candidate facial competence to incumbency, partisanship, and the vote outcome indicates that faces must matter somewhere along the causal chain. We next discuss research that indicates ways in which faces might matter. The research indicates that to understand the relationship of face to vote, we must move beyond the relative comparison of candidate faces and consider the relationship of individual candidate faces to political context. We find that face does matter when each candidate is considered as a separate individual.

2.1 The Association Between Political Factors and Facial Competence

The causes of the association between incumbency and facial competence are suggested by findings in the psychology research on facial cues and in the economics literature on physical appearance and vocational success. Economics research has consistently shown that facial characteristics are associated not just with higher salaries, but with greater human capital as well (Biddle & Hamermesh 1998, Mobius & Rosenblat 2006, Hamermesh & Biddle 1994, Rule & Ambady 2008). Among the more innovative recent studies demonstrating the connection between human capital and facial appearance, Zebrowitz, Hall, Murphy, & Rhodes (2002) showed research participants photographs of people unknown to the participants and asked

Table 1: Using Facial Competence, Incumbency, and Partisanship to Predict Democratic Vote Share.

	House			Senate				
	Faces Only	Incumbency Only	Partisanship Only	All	Faces Only	Incumbency Only	Partisanship Only	All
Intercept	33.44 (2.83)	41.52 (1.49)	1.27 (4.17)	20.80 (2.47)	33.30 (4.13)	48.03 (2.23)	5.93 (7.05)	19.36 (5.50)
Todorov Democrat Competence	28.86 (5.56)			4.43 (2.28)	32.00 (7.86)			6.65 (5.25)
Democrat Incumbent		22.80 (1.76)		15.42 (1.58)		14.31 (2.76)		10.76 (2.53)
Republican Incumbent		-6.71 (1.71)		-8.09 (1.32)		-10.70 (2.72)		-9.69 (2.36)
District Partisanship			1.02 (0.09)	0.50 (0.05)			0.93 (0.15)	0.57 (0.11)
N	142	142	142	142	89	89	89	89
R ²	0.16	0.80	0.47	0.89	0.16	0.59	0.31	0.70
Adjusted R ²	0.16	0.80	0.47	0.88	0.15	0.58	0.30	0.69
Std. Error of Regression	14.09	6.84	11.17	5.27	13.49	9.48	12.24	8.17

Ordinary least squares with standard errors in parentheses. Dependent variable is Democratic Congressional vote share for 2004 House races or 2000, 2002 and 2004 Senate Races. District partisanship measured by contemporaneous (2000, 2004) or lag (2002) Democratic presidential vote share of two-party vote.

them to evaluate the intelligence of the people in the photographs. Zebrowitz et al. (2002) found that attractiveness was not only correlated with perceived intelligence but it was also correlated with actual intelligence (measured by I.Q. tests). According to Hamermesh & Biddle (1994, p.1191), “the effects of an individual’s looks on his or her earnings are very robust.” In short, facial characteristics are found to be associated with human capital, intelligence, and earnings.

That incumbents will have greater levels of human capital than challengers on average seems uncontroversial since challengers are quite often not serious candidates. If we expect that, on average, incumbents have higher levels of human capital than challengers, then the economics and psychology findings would indicate that, on average, incumbents will have better facial characteristics than challengers. This is one explanation for the correlation of face to incumbency.

The correlation of face to district partisanship can also be explained with face as a proxy for human capital. If individuals with more human capital, on average, have better facial characteristics, then the candidate advantaged by the district’s party identification should tend to have better facial characteristics. We might see this for two, non-mutually exclusive reasons: strategic candidate behavior or strategic party behavior.

Prospective candidates with more human capital — and by implication better average facial qualities — encounter higher opportunity costs by participating in a congressional contests that they do not win. Thus potential candidates with better faces may be more likely to enter a contest when the expected probability of winning is higher, and less likely to enter a contest when the expected probability of winning is lower (e.g. Jacobson & Kernell 1983). Therefore, in districts with lopsided partisan proclivities or popular incumbents, the disadvantaged challenger candidate would be less likely to possess high facial competence.

Strategic political parties may choose to recruit with more effort in congressional contests that they expect will be the most competitive (e.g. Herrnson 1986, Jacobson 1996). This effort may lead to higher-quality faces in the competitive districts due to two not-mutually-

exclusive factors. Parties may anticipate that a candidate with better facial qualities will appeal to voters and therefore select explicitly on candidate appearance. Or, in the process of recruiting candidates with high levels of human capital, parties may unintentionally recruit candidates with better facial characteristics. In either case, better candidate faces would obtain in districts with the most party effort of recruitment.

That candidate face is non-randomly correlated with the partisanship of a district and incumbency is a testable proposition if the faces of candidates can be compared across districts. We next present a method that allows us to do this by separately measuring individual candidate facial competence. This measure allows us to test more complete models of the relationship between face, political variables, and election outcomes.

3 Measuring Individual Candidate Facial Competence

We created a survey to measure the facial competence of individual candidates. By creating a common scale of facial competence, we are able to test the relationship between political context and the facial competence of individual candidates.

Our survey measured the perceived facial competence of candidate faces for the U.S. House in 2004, and the U.S. Senate 1990-2006. Todorov et al. (2005) asked respondents to compare the two faces of opposing candidates in a given election. This method measures the competence of the two faces relative to each other, head to head, but does not indicate how the two faces would compare to other faces from outside of that election.

We wrote a web-based survey that presented each participant two randomly-drawn faces from the pool of all candidate faces. Each pair of faces was presented on a computer screen for one second, and the participant was asked to choose which of the two faces was more competent. The text of the question followed exactly that used by Todorov et al. (2005). An example of the survey can be found at <http://sjhill.bo1.ucla.edu/faces>.

We conducted two separate surveys. In the first, 296 students in a lower division political

science class at UCLA evaluated images of white male candidates from 2004 House elections.⁵ 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 the 2004 House candidates from the first survey.⁶ Images were compiled following Todorov et al. (2005) from cnn.com and other internet sources.

Each participant evaluated hundreds of face pairings, all randomly-drawn. We used more than 167,000 binary choices by participants to build competence scores for each face in the candidate pool.⁷ Three assumptions were used to build the scores.

Our assumptions can most easily be described with reference to the simple case of three faces in the pool, A, B, and C, without loss of substantive generality. First, we assume the latent competence is transitive, such that if A has higher perceived competence than B, and B has higher perceived competence than C, more often than not A will be chosen the more competent over B *and* C, and more often than not B chosen the more competent over C. Second, we assume that the participant choice is probabilistic and not determinative such that B or C might sometimes be chosen more competent than A despite A's higher latent competence.

Third, we assume the latent scores are continuous, such that A, B, and C can be placed on a scale and represented by (arbitrary) numeric values. The numeric values can be reconstructed to probabilities that any given face will be chosen over any other face. For example, if A had a competence score of 1, B a score of .5 and C a score of .45, we would expect A

⁵We 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.

⁶Before the first survey, participants were asked to identify the photo of the Member of Congress for the UCLA area, Henry Waxman, from a lineup of photos containing members of the California Assembly. This was done as a test of potential recognition of the images. The participants recognized 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. 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.

⁷We 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, to measure their consistency), dropping early and late evaluations for fatigue and learning. None substantively affected our results.

to be chosen more competent than B and C most of the time, but while B would be chosen more competent than C more often than not, the choice would not occur as frequently as A over B due to the relatively smaller difference between .45 and .5.

Given these assumptions, we are able to use the thousands of two-face competence comparisons obtained from our survey participants to infer the latent competence score for each candidate’s face. We present details of the estimation model in the Appendix.

3.1 Replication

We are able to use the scores generated from our survey method to compare our results to those of Todorov et al. (2005). The Todorov et al. (2005) facial competence measure is the proportion of times one candidate’s face is chosen more competent than their opponent’s. Given our estimated competence scores, we can construct the predicted probability that any face will be chosen more competent than any other by reversing the estimation model described in the Appendix.⁸

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 2. 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 our competence scores indicate 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 the vote (not presented).⁹

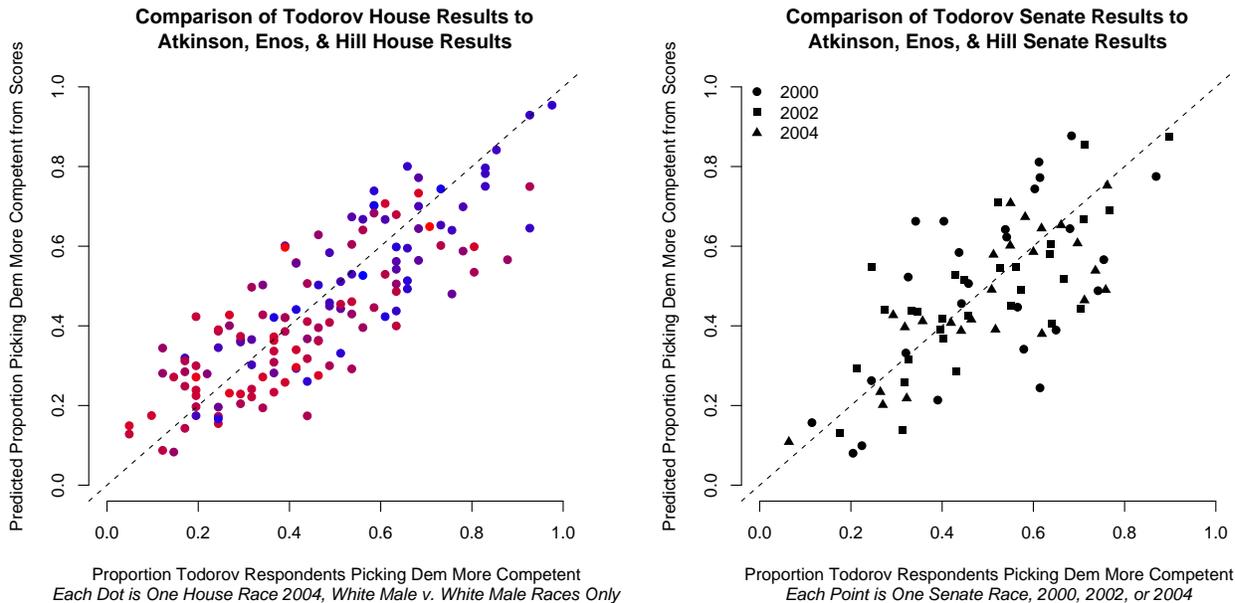
⁸Specifically, the probability that any participant chooses candidate i more competent than candidate j given estimated competence scores c_i and c_j is

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

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

⁹For the purposes of our survey, part of the face images were those used by Todorov et al. (2005). We

Figure 2: Replication of Todorov Experimental Results



4 The Competence of Politician Faces

Our survey allows us to compare the subjective facial competence of Members of Congress and their competitors. The numbers measure “inferred facial competence” based upon the evaluations of our survey participants, and we standardize the scores to have mean zero and unit variance. The scales for each house were estimated separately, so comparing the numeric values presented here across chambers is not meaningful.

We present in Table 2 summary statistics of the distributions of estimated facial competence scores in the U.S. House and Senate. In both the House and Senate, the average and median incumbent is more competent than the average and median challenger. However, the difference is smaller in the Senate, perhaps due to the higher profile Senate campaigns attracting, on average, better challengers.

To meaningfully illustrate the numeric competence scores, we present in Table 3 an

also collected our own faces from CNN.com, as Todorov et al. (2005) had done, and supplemented those with photos that were collected from various web-sites. We standardized these photos in size and pixel count, put in a standard gray background, and turned all to black and white. See Appendix 8.2 for a discussion of our efforts to purge the pictures of “quality” that could be associated with candidate traits.

Table 2: Distributions of Competence Scores by Chamber and Incumbency

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
All House Candidates	-3.11	-0.69	0.10	-0.01	0.75	1.82
House Challengers	-3.11	-1.11	-0.40	-0.39	0.37	1.68
House Incumbents	-2.25	-0.22	0.39	0.33	0.98	1.82
All Senate Candidates	-3.98	-0.57	0.20	0.04	0.77	2.22
Senate Challengers	-3.98	-0.93	-0.04	-0.25	0.55	2.22
Senate Incumbents	-2.49	-0.35	0.37	0.30	0.88	1.99

Note: Scores are not comparable across chambers. Cell entries are the standardized coefficients of the probit estimation described in the appendix.

example grouping of faces along with their estimated facial competence. The median 2004 House incumbent face was Representative Cliff Stearns (R-FL 6th) (picture (a) in Table 3). Stearns handily defeated his opponent, David Bruderly (b) 64 to 36 percent despite having only a relatively small advantage in facial competence (the facial competence estimate of each candidate is listed below each photograph). The median challenger was Warren Redlich (c) who challenged Representative Michael R. McNulty (D-NY 21st) (d) and only won 30 percent of the two party vote.

The most competent looking candidate for the House was Representative Alan B. Mollohan (D-WV 1st) (e). Mollohan carried his district by 36 points over his opponent Alan Parks (f) and had a commanding competence advantage as well.

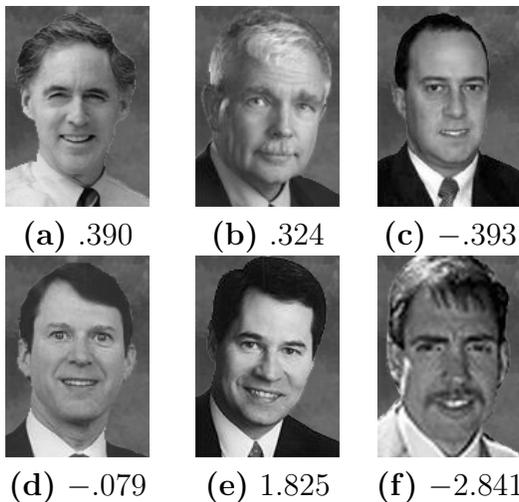
In the Senate we included candidates regardless of race or gender, except high-profile senators likely to be recognized by survey participants (e.g., Hillary Clinton). The median Senate challenger was David Walters (g) who lost to David Inhofe (R-OK) (h) despite having an advantage in facial competence. The median Senate incumbent face in our pool was John Glenn (D-OH) (i).¹⁰ Glenn easily defeated his opponent in 1992, Michael DeWine (j), over whom he had a substantial advantage in competence.¹¹ The most competent of all Senate

¹⁰In seeking a measure that avoided bias because of familiarity, we unfortunately removed easily recognizable candidates, so we cannot comment on the competence of some of the more famous members. A group of congress and elections scholars did not identify Glenn as recognizable to college students, so he was not removed.

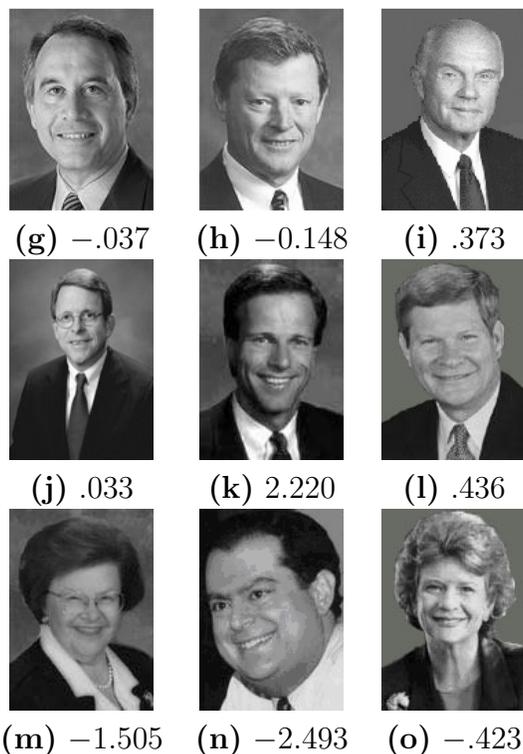
¹¹DeWine would be elected to the Senate in the following election.

Table 3: Faces and Inferred Competence

House Candidates



Senate Candidates



Competence Scores are listed below each candidate. Scores are not comparable across chambers.
House Candidates: (a) Rep. Cliff Stearns (R – FL6), (b) David Bruderly, (c) Warren Redlich, (d) Rep. Michael R. McNulty (D – NY21), (e) Rep. Alan B. Mollohan (D – WV1), (f) Alan Parks.
Senate Candidates: (g) David Walters, (h) Sen. David Inhofe (R – OK), (i) Sen. John Glenn (D – OH), (j) Michael DeWine, (k) Sen. John Thune (R – SD), (l) Sen. Tim Johnson (D – SD), (m) Sen. Barbara Mikulski (D – MD), (n) Sen. Spencer Abraham (R – MI), (o) Sen. Debbie Stabenow (D – MI).

candidates was John Thune (R-SD) (k), who defeated Tom Daschle in 2004 despite having narrowly lost to Tim Johnson (D-SD) (l) in 2002, over whom he had a considerable advantage in face.

Our pool includes Senate candidates from 1990-2006. John Thune also has the distinction of being the most competent looking current Senator. The least competent face in our pool of Senators from the 110th Congress was Barbara Mikulski (D-MD) (m). The least competent incumbent Senate face across all races and years was Spencer Abraham (R-MI) (n), who lost to the more competent but still below average Debbie Stabenow (D-MI) (o) in 2000.

Finally, Republicans typically have more competent faces than Democrats. The median Republican candidate is more competent looking than the median Democrat in the House by one-third standard deviation (-0.040 to 0.312) and the Senate by one-tenth standard deviation (0.129 to 0.234).

5 Competition and Faces in Congressional Elections

With estimates of the facial competence of individual candidates in hand, we return now to assessing the mechanism that connects face to political success. Given the findings about physical appearance and vocational success in economics and about inferences of personality traits in psychology we consider the effect of faces being distributed non-randomly across districts. One result could be that voters are not simply responding to the relative distance between the two candidates' faces, but rather are responding to each face separately. We find evidence that they do. We then present graphical evidence of the selection of higher quality faces into more competitive and partisan-advantaged districts.

We first regress incumbent vote share on the incumbent and challenger's facial competence scores as separate variables. Re-coding the dependent variable as incumbent vote share is driven by the importance of incumbency and challenger quality in congressional elections (Jacobson 2004). It is plausible that the effect of a challenger's face may be different from

the effect of an incumbent’s face.

We present these results in the left column for each chamber in Table 4.¹² The effect of challenger and incumbent faces do differ. Moving from the least competent challenger face to the most competent challenger face is estimated to decrease the incumbent’s vote share by almost 6 percentage points in the House and by almost 20 percentage points in the Senate. Oddly, in both chambers a more competent incumbent face is estimated to *decrease* the incumbent’s vote share. In the House this estimate is statistically different from zero, with an estimated decrease across incumbent faces almost as large as that for the challenger face.

We suggest that the counter-intuitive notion that a better incumbent face causes lower incumbent vote share is unlikely to be true. It is more likely that we are again observing the effects of omitted variable bias. Strategic candidate entry and strategic party recruitment suggest that the faces-only model in Table 4 is misspecified.

What should we make of the negative coefficient on incumbent facial competence? On the one hand, were strategic candidate behavior driving selection, a candidate’s facial competence would be proportionate to her chances of winning an election because individuals with more competent faces, and therefore higher wages in the private sector, would incur higher opportunity costs for participating in elections. In this case, we would observe linearly-increasing human capital and facial competence as the incumbent’s advantage in the district increases, thus inducing a positive coefficient for the incumbent face variable when it is regressed on incumbent vote share.

On the other hand, were strategic party recruitment driving selection, parties would expend effort to recruit the most competent-looking faces to the most closely contested elections. This would present an empirically different pattern than the candidate opportunity cost model. If parties engage in this strategic allocation of recruitment effort, then incumbents from marginal districts (i.e. expected incumbent share of about 50 percent) would tend

¹²Including year fixed effects does not affect the results.

to have more competent faces than incumbents from safe districts (i.e. expected incumbent share of 80 percent) because the party would have recruited a high quality face to the 50-50 district. This would induce a negative correlation between incumbent facial competence and incumbent vote share because the best faces are in the districts with lower incumbent vote share.

We find evidence of party recruitment in the second and fourth columns of Table 4. When we add a measure of district competitiveness to the model, the coefficient on incumbent face moves towards zero, which we would expect to occur if incumbent face were negatively correlated with expected probability of incumbent victory. We measure the competitiveness of a contest by coding each race according to the classifications provided by the *Cook Political Report* (Cook 1992-2006).¹³ 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 competence, 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 are taken from the August 2003 *Cook Political Report* newsletter. We recode the measure of competitiveness in the direction of the expected probability of incumbent victory, and include dummy indicator variables for each category. For the House results, the excluded category is “Tossup”, and for the Senate, the excluded category is “Challenger Lean”.¹⁴

In both chambers, adding the district competitiveness variable increases the amount of explained vote share variance. For the House, the perverse coefficient on incumbent face is cut in half and loses statistical significance as does the estimated effect of challenger face. In

¹³For examples of scholarly work employing Cook’s report, see Gimple, Karnes, McTague, & Pearson-Merkowitz (2008) or Vavreck (2001).

¹⁴This occurred in the 1992 North Carolina Senate Election, where the district was coded by Cook as Lean Republican. Incumbent Democratic Senator Terry Sanford lost to the challenging Republican by 10,000 votes. It is possible that even the Cook measure is endogenous to the characteristics of the challenger. We ran the same models with the presidential vote share of the district or state from the previous election in lieu of Cook as a measure of competitiveness. This results in little change in the substance of what we describe here. We are more satisfied, however, that Cook’s measure better captures the competitiveness of district, which is based on more holistic information than just the result of the previous presidential election.

Table 4: Using Facial Competence and District Competitiveness to Predict Incumbent Vote Share.

	House 2004		Senate 1992-2006	
	Faces Only	With Expectation	Faces Only	With Expectation
Intercept	64.51 (0.54)	57.42 (2.76)	60.14 (0.80)	49.66 (6.66)
Challenger Facial Competence	-1.30 (0.47)	-0.57 (0.41)	-3.37 (0.63)	-2.05 (0.50)
Incumbent Facial Competence	-1.17 (0.56)	-0.72 (0.48)	-0.45 (0.88)	-0.30 (0.66)
Cook: Incumbent Lean		-0.39 (2.98)		5.08 (6.77)
Cook: Incumbent Likely		1.53 (3.20)		8.32 (6.76)
Cook: Incumbent Safe		8.72 (2.79)		16.89 (6.73)
Cook: Tossup				-0.37 (6.88)
N	145	145	147	147
R^2	0.08	0.38	0.17	0.56
Adjusted R^2	0.06	0.35	0.16	0.54
Std. Error of Regression	5.62	4.67	8.97	6.65

Ordinary least squares with standard errors in parentheses. Dependent variable is incumbent vote share.

the Senate, the effect of challenger face is decreased by a third but maintains a confidence interval well outside of zero. Moving from the least competent Senate challenger to the most competent Senate challenger is estimated to decrease incumbent vote share by 12 points. The effect of the incumbent's face remains uncertainly-estimated yet maintains its perverse sign.

Adding this measure of competitiveness did not fully resolve the situation, as both incumbent face coefficients remain negative, if uncertain. We suspect that our current estimate of the effect of face may still be biased due to our inability to fully capture the strategic political calculations at play.

If the competitiveness of contests is correlated with the effect of challenger faces, it is not surprising that the effect is stronger in the Senate. Senate races are higher salience elections

and voters seem more likely to receive information about the appearance of the challenging candidate. We will explore the effect that faces have on individual voters in Senate elections in more detail in the next section.

5.1 The Distribution of Faces Across Districts

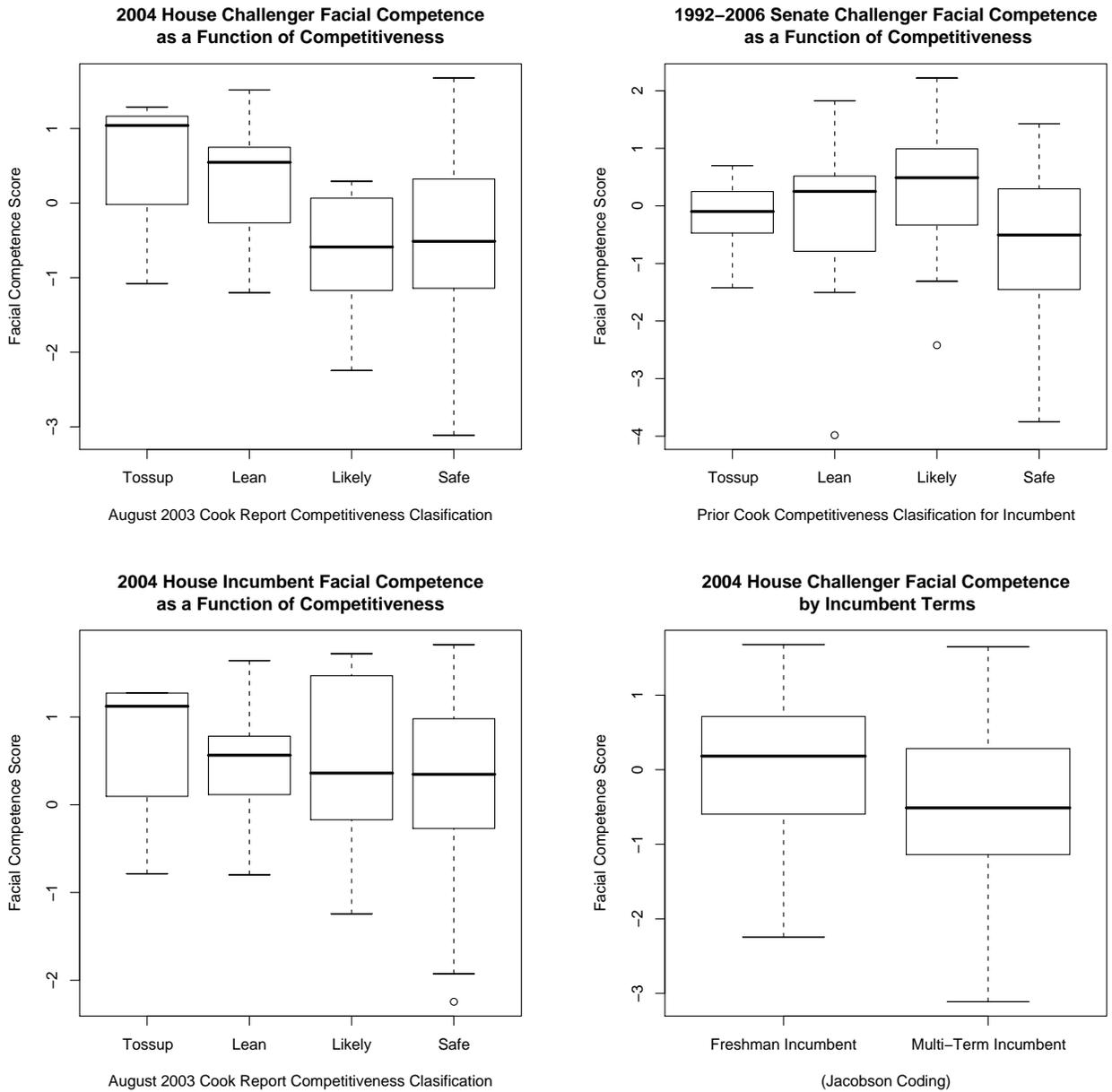
We turn now to a graphical exploration of the distribution of faces by competitiveness. Our regression results indicate that faces are correlated with district and election characteristics, perhaps due to strategic behavior. In Figure 3, we present the distribution of candidate facial competence by competitiveness measures. The first three panels use *Cook Political Report* measures of district competitiveness. The top two panels are boxplots of House and Senate challenger competence as a function of year-prior Cook district classification. Both indicate that more competent challenger faces enter the more competitive elections, though in the Senate the pattern is somewhat less clear. In both chambers the worst challenger faces contest elections in districts considered safe for the incumbent.

The third panel shows the distribution of House incumbent faces by district competitiveness. Even though the Cook measure should be expected to take account of incumbent face, again we see a pattern of more competent faces in more competitive districts. Finally, the fourth panel presents the distribution of House challenger faces by the tenure of the incumbent, a measure likely to be robust to potential endogeneity in the Cook measure. The graphic demonstrates more competent faces challenge freshmen incumbents, generally the most vulnerable of incumbents. The polyserial correlation between challenger face and a freshman incumbent indicator is $r = .28$.

6 Individual Voters and Faces

In the preceding analysis we found evidence that the effect of face on vote share is not solely a voter-level process, but rather that the allocation of faces to districts is related to

Figure 3: Facial Competence by District Competitiveness



incumbency and district partisanship. In this section, we extend our investigation to the analysis of individual-level data to see how incumbency and district partisanship structure individual response to candidate faces.

There are two avenues through which the candidate faces might influence the vote choice. On the one hand, the quality of each candidate's face may add or subtract some constant probability across all voters. On the other hand, the face may interact with partisanship such that the response to the face is conditional on the voter's and the candidate's partisanship. In statistical parlance, face may produce a simple intercept shift in the vote equation, or be interactive such that the slope mapping face to vote varies by partisanship.

How face functions as a cue to different types of partisans is a subject of theoretical importance to understanding the relationship between faces and election outcomes. If voters were naively making comparisons of faces independent of partisanship or incumbency, we would expect the influence of facial competence on vote choice to be limited to non-partisan voters. Those voters that lack a partisan cue to structure their vote choice would be more influenced by candidate characteristics such as face than voters whose vote choice is structured by their partisan identification. If, on the other hand, face comparisons in the real world are not made naively, but instead interact with other voter dispositions such as partisanship, we would expect face to effect the choices of partisans as well.

We pursue this exploration by combining the facial competence measures from our survey with exit poll data.¹⁵ We chose to use this data, rather than other common survey data sets, such as the American National Election Study because of the advantages of a relatively large sample size and the proximity to the actual vote choice. Because we are only concerned with vote choice, not voter turnout, there is no danger in selection bias by ignoring the non-voters that comprise part of the ANES sample. Additionally, because vote choice is the key variable, a poll that is conducted immediately following the casting of a ballot is less likely

¹⁵We make use of network exit poll data for the 2004 House elections and all Senate elections covered in our survey of facial competence with the exception of the 2004 Senate elections. The *National Election Pool* exit poll questionnaires did not include a question on Senate vote choice in 2004.

affected by measurement error than a survey conducted up to a month after the election.

In Table 5 we present the results from a probit regression analysis of the effect on vote choice produced by the interaction between facial competence and respondent partisanship.¹⁶ Because the results are probit coefficients and include complicated interactions, we discuss the results of Table 5 with reference to predicted values using the coefficients from the full House and Senate models. Our predicted values are calculated by holding one candidate’s facial competence at the 50th percentile and holding the Cook report of district competitiveness at “likely” going to the incumbent party. We estimate the change in vote choice of moving the other candidate’s facial competence from the 25th percentile to the 75th percentile. In Table 6 we present predicted changes with 95 percent confidence intervals.

The first column of Table 6 presents the predicted effect of changing incumbent facial competence from the 25th to the 75th percentile for each of three types of voters and overall given the underlying distribution of voters from the three groups in the electorate. For independents, the point estimate indicates that increasing incumbent facial competence *decreases* the probability of a House incumbent vote by 4 percentage points, though the 95 percent confidence interval spans zero. Similar uncertainty exists in the estimated effect for challenger co-partisans and incumbent co-partisans.

The estimated effects of changing candidate face in the House are unstable, with only the predicted effect of challenger face on challenger co-partisan vote being statistically different from zero. In the Senate, however, more than just challenger co-partisans respond to face. A better incumbent face is predicted to increase the probability of the challenger’s party voters defecting to the incumbent by 2.0 percentage points, while once again a better incumbent face leads to a decreased probability of incumbent party voter loyalty of 1.4 points. The predicted effect of changing a Senate challenger’s face is significant for all three subpopulations, with increased probability of challenger vote of 3.5 points for independents, 2.5 points for challenger-party voters, and 1.5 points for incumbent-party voters. Note also that the size

¹⁶Including year fixed effects does not affect the results.

Table 5: Using Facial Competence and Partisanship to Predict Individual-Level Vote Choice.

	House 2004		Senate 1992-2006	
	Faces Only	With Expectation	Faces Only	With Expectation
Intercept	0.20 (0.05)	0.34 (0.24)	0.19 (0.02)	-0.04 (0.03)
Respondent Shares Challenger Party	-1.29 (0.08)	-1.29 (0.08)	-1.01 (0.02)	-1.03 (0.02)
Respondent Shares Incumbent Party	1.32 (0.08)	1.31 (0.08)	1.03 (0.03)	1.02 (0.03)
Challenger Facial Competence	-0.01 (0.10)	0.02 (0.11)	-0.09 (0.02)	-0.07 (0.02)
Incumbent Facial Competence	-0.25 (0.13)	-0.20 (0.13)	0.00 (0.02)	-0.01 (0.02)
Respondent Shares Challenger Party *Challenger Facial Competence	-0.32 (0.16)	-0.31 (0.16)	-0.01 (0.02)	-0.00 (0.02)
Respondent Shares Incumbent Party *Challenger Facial Competence	-0.14 (0.17)	-0.14 (0.17)	0.02 (0.02)	0.01 (0.02)
Respondent Shares Challenger Party *Incumbent Facial Competence	0.02 (0.20)	0.02 (0.20)	0.05 (0.03)	0.07 (0.03)
Respondent Shares Incumbent Party *Incumbent Facial Competence	0.36 (0.21)	0.38 (0.20)	-0.05 (0.03)	-0.04 (0.03)
Cook: Incumbent Lean		-0.42 (0.24)		0.18 (0.04)
Cook: Incumbent Likely		-0.26 (0.24)		0.15 (0.04)
Cook: Incumbent Safe		-0.12 (0.23)		0.42 (0.03)
N	4350	4350	21974	21619
AIC	3460.51	3453.03	21655.22	21113.43

Probit regression coefficients with standard errors in parentheses. Dependent variable is respondent voted for incumbent candidate.

of these effects are in line with the estimated aggregate effect we presented in the multiple regressions of Table 1.

The predicted effects of candidate faces on vote choice are not limited to political independents. Rather, even individuals who identify with a political party can be influenced by faces under the right circumstances. Generally it is the challenger’s face that matters, except to the extent that the incumbent’s face is predicted to *decrease* the incumbent vote share.

Among partisans, voters who share the party label of the challenger candidate are the only group demonstrating a statistically significant response to candidate face in voting across both chambers. That the most distinctive effects occur among this group suggests that voter response to candidate faces is structured by political attitudes. Therefore, we suggest it is likely that the process producing the causal effect of candidate faces on vote choice is more complex than naive voters making decisions based on a rapid comparison of faces. Indeed, voters appear to be influenced by facial qualities through mechanisms consistent with traditional theories of political campaigns and partisanship.

Strikingly Table 6 predicts the same counter-intuitive negative relationship between incumbent facial competence and electoral success in the individual level data. Though challenger candidates appear to do better with increasing levels of facial competence, the same does not hold true for incumbents — incumbents are predicted to lose votes as their facial competence increases. This again suggests to us the presence of strategic behavior by elite politicians.

7 Discussion

In this paper, we have measured the effect of candidate faces on vote share. We show that the considerable bivariate relationship of face to vote is reduced in the presence of observational controls. Holding incumbency and partisanship constant, the effect of facial competence is small. However, we do find an effect of face to vote when the two candidate faces are

Table 6: First Difference Predictions Based on Probit Analysis in Table 5.

Category of Voter	House 2004		Senate 1992-2006	
	Incumbent competence from 25th to 75th	Challenger competence from 25th to 75th	Incumbent competence from 25th to 75th	Challenger competence from 25th to 75th
Independent	-4.46	0.51	-0.41	-3.51
(45% House respondents, 27% Senate respondents)	[-10.64,1.15]	[-4.28,5.34]	[-2.59,1.8]	[-5.17,-1.75]
Shares Challenger's Party	-1.92	-2.89	2.03	-2.46
(24% House respondents, 35% Senate respondents)	[-5.62,1.3]	[-5.59,-0.33]	[0.68,3.32]	[-3.52,-1.39]
Shares Incumbent's Party	1.48	-1.05	-1.44	-1.51
(31% House respondents, 38% Senate respondents)	[-1.26,4.4]	[-3.34,1.1]	[-2.59,-0.25]	[-2.5,-0.55]
Net Effect of Change	-2.01	-0.79	0.05	-2.38

Percentage point change in the probability that respondent supports incumbent candidate with simulated 95 percent confidence intervals in brackets (Imai, King, & Lau 2007a, Imai, King, & Lau 2007b). Predictions calculated holding competitiveness at "likely incumbent victory," holding the rival candidate's face at the 50th percentile of all candidate faces for the applicable chamber, and moving from 25th percentile face to 75th percentile face.

accounted for separately, and also at the level of the individual voter. We find evidence that the effect of candidate face on individual voters is more often found among party identifiers, not independent voters.

One important conclusion to note is that the presence of randomization in a study does not imply that observational controls are not needed. It is only when treatment is randomized that estimated treatment effects in two-variable comparisons are free of concern about lurking variables.¹⁷ Randomization in measurement is not the same as randomization of treatment.

We have also constructed an original measure of candidate quality. To the extent that economists have demonstrated that appearance is correlated with other success (Rule & Ambady 2008, Mobius & Rosenblat 2006, Hamermesh & Biddle 1994, Biddle & Hamermesh 1998), this measure is not inconsistent with the standard measures of candidate quality (e.g. Jacobson 2004). If this is an accurate measure of candidate quality, we have found that higher quality candidates are more likely to enter the subset of congressional elections in which the contest is competitive. How these candidates come to enter these contests has important implications for understanding the dynamics of party and candidate influence in shaping the choices available to voters. On the one hand, higher quality candidates with more to lose could be selecting into contests which they are more likely to win. On the other hand, parties could be recruiting the best candidates into the most competitive contests (Dominguez N.d., Gibson, Cotter, Bibby, & Huckshorn 1985, Gibson, Cotter, Bibby, & Huckshorn 1983, Herrnson 1986, Mann & Ornstein 1981).

Our evidence suggests that candidates, parties, or likely both, are making decisions which limit the range of faces and candidate quality from which voters are able to choose. Because many congressional contests in the United States are not competitive and because candidates with high competence are more likely to enter the contests in which they have a reasonable chance of success, we find high competence candidates defeating their usually low competence challengers in the majority of contests — a dynamic that produces a high correlation

¹⁷Of course, the implementation of the randomization should be investigated.

between facial competence and election outcomes. That candidates are not distributed randomly across contests and that it is likely that elite level actors, in the form of parties and candidates, are making decisions which affect the allocation of candidates across races suggests avenues for future research. For example, our unique measure of candidate quality may be useful in research adjudicating between candidate and party centered models of electoral politics.

The attribution of election outcomes to potentially-objectionable voter decision processes also has implications for normative political theory. We have demonstrated however that, even if voters do sometimes make decisions based on candidate appearance, the effect does not appear large enough to decide any but the closest election outcomes, and what effect there is appears to be conditioned by voter partisanship and candidate incumbency status.

Some of the media attention surrounding the research by Todorov et al. (2005) was probably generated by the sense that the finding demonstrates that the voting public is uninformed. We have demonstrated that this plays a much smaller role in election outcomes than one might infer from a casual reading of Todorov et al. (2005) and its representation in the media. However, a world where many districts pit a good-looking incumbent who holds partisanship consistent with the vast majority of the district against a challenger with little chance of victory might raise normative concerns as well.

References

- Ballew II, Charles C., & Alexander Todorov. 2007. "Predicting Political Elections From Rapid and Unreflective Face Judgments." *Proceedings of the National Academy of Sciences*.
- Benjamin, Daniel J., & Jesse M. Shapiro. 2006. "Thin-Slice Forecasts of Gubernatorial Elections." *National Bureau of Economic Research Working Paper Series* (Working Paper No. 12660).
- Biddle, Jeff E., & Daniel S. Hamermesh. 1998. "Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre." *Journal of Labor Economics* 16 (1): 172–201.
- Campbell, Angus, Philip E. Converse, Warren E. Miller, & Donald E. Stokes. 1960. *The American Voter*. New York: Wiley.

- Cook, Charles E. 1992-2006. "The Cook Political Report.". Washington, D.C.: National Political Review, Inc.
- Cressie, Noel A. C. 1993. *Statistics for Spatal Data*. New York: Wiley-Interscience.
- Dominguez, Casey. N.d. "How Party Elites and Ambitious Candidates Respond to Anticipated General Election Competitiveness.". Working Paper, University of California, San Diego.
- Erikson, Robert S. 1971. "The Advantage of Incumbency in Congressional Elections." *Polity* 3 (3): 395–405.
- Fox, John. 2007. *polycor: Polychoric and Polyserial Correlations*. R package version 0.7-5.
- Gelman, Andrew, & Gary King. 1990. "Estimating Incumbency Advantage without Bias." *American Journal of Political Science* 34 (4): 1142–1164.
- Gibson, James L., Cornelius P. Cotter, John F. Bibby, & Robert J. Huckshorn. 1983. "Assessing Party Organizational Strength." *American Journal of Political Science* 27 (2): 193–222.
- Gibson, James L., Cornelius P. Cotter, John F. Bibby, & Robert J. Huckshorn. 1985. "Whither the Local Parties?: A Cross-Sectional and Longitudinal Analysis of the Strength of Party Organizations." *American Journal of Political Science* 29 (1): 139–160.
- Gimple, James G., Kimberly A. Karnes, John McTague, & Shanna Pearson-Merkowitz. 2008. "Distance-decay in the political geography of friends-and-neighbors voting." *Political Geography* 27: 231-252.
- Green, Donald, Bradley Palmquist, & Eric Schickler. 2002. *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. New Haven: Yale University Press.
- Groseclose, Tim, & Charles Stewart III. 1998. "The Value of Committee Seats in the House, 1947-91." *American Journal of Political Science* 42 (2): 453–474.
- Hamermesh, Daniel S., & Jeff E. Biddle. 1994. "Beauty and the Labor Market." *American Economic Review* 84 (5): 1174–1194.
- Hamilton, John. 2005. "Scientists Search for that Winning Look." *National Public Radio All Things Considered* (June).
- Herrnson, Paul S. 1986. "Do Parties Make a Difference? The Role of Party Organizations in Congressional Elections." *Journal of Politics* 48 (3): 589–615.
- Imai, Kosuke, Gary King, & Olivia Lau. 2007a. *probit: Probit Regression for Dichotomous Dependent Variables*. In Kosuke Imai, Gary King, and Olivia Lau, "Zelig: Everyones Statistical Software," <http://gking.harvard.edu/zelig>.

- Imai, Kosuke, Gary King, & Olivia Lau. 2007b. *sim: Simulating Quantities of Interest*. In Kosuke Imai, Gary King, and Olivia Lau, “Zelig: Everyones Statistical Software,” <http://gking.harvard.edu/zelig>.
- Jacobson, Gary C. 1996. “The 1994 House Elections in Perspective.” *Political Science Quarterly* 111 (2): 203–223.
- Jacobson, Gary C. 2004. *The Politics of Congressional Elections*. 6 ed. New York: Longman.
- Jacobson, Gary C., & Samuel Kernell. 1983. *Strategy and Choice in Congressional Elections*. New Haven: Yale University Press.
- Little, Anthony C., Robert P. Burriss, Benedict C. Jones, & S. Craig Roberts. 2007. “Facial Appearance Affects Voting Decisions.” *Evolution and Human Behavior* 28 (1): 18–27.
- Mann, Thomas E., & Norman J. Ornstein. 1981. “The Republican Surge in Congress.” In *The American Elections of 1980*, ed. Austin Ranney. American Enterprise Institute.
- Mayhew, David R. 1974. *Congress: The Electoral Connection*. New Haven: Yale.
- McHugo, Gergory J., John T. Lanzetta, Denis G. Sullivan, Roger D. Masters, & Basil G. Englis. 1985. “Emotional Reactions to a Political Leader’s Expressive Displays.” *Journal of Personality and Social Psychology* 49 (6): 1513–1529.
- Mobius, Markus M., & Tanya S. Rosenblat. 2006. “Why Beauty Matters.” *American Economic Review* 96: 222–235.
- R Development Core Team. 2007. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Ripley, Brian D. 2004. *Spatial Statistics*. New York: Wiley-Interscience.
- Rule, Nicholas O., & Nalini Ambady. 2008. “The Face of Success: Inferences From Chief Executive Officers’ Appearance Predict Company Profits.” *Psychological Science* 19 (2): 109–111.
- Skloot, Rebecca. 2007. “Faces Decide Elections.” *New York Times Magazine*.
- Sullivan, Denis G., & Roger D. Masters. 1988. “‘Happy Warriors’: Leaders’ Facial Displays, Viewers’ Emotions, and Political Support.” *American Journal of Political Science* 32 (2): 345–368.
- Todorov, Alexander, Anesu N. Mandisodza, Amir Goren, & Crystal C. Hall. 2005. “Inferences of Competence from Faces Predict Election Outcomes.” *Science* 308 (5728): 1623–1626.
- Vavreck, Lynn. 2001. “The Reasoning Voter Meets the Strategic Candidate: Signals and Specificity in Campaign Advertising, 1998.” *American Politics Research* 29 (5): 507–529.

Zebrowitz, Leslie A., & Joann M. Montepare. 2005. "PSYCHOLOGY: Appearance DOES Matter." *Science* 308 (5728): 1565-1566.

Zebrowitz, Leslie A., Judith A. Hall, Nora A. Murphy, & Gillian Rhodes. 2002. "Looking Smart and Looking Good: Facial Cues to Intelligence and Their Origins." *Personality and Social Psychology Bulletin* 28 (2): 238-249.

8 Appendix

8.1 Details of Facial Competence Estimation

Facial competence scores for each candidate are calculated using the binary choices made by survey participants between two candidate faces. The scores are estimated similar to a unique method used to model congressional committee choice (Groseclose & Stewart III 1998). We assume each of the i candidate faces has a location, c_i , on the latent competence scale. 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

$$\Phi\left(\frac{c_i - c_j}{\sigma\sqrt{2}}\right) \quad (1)$$

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 k th face was selected the more competent, -1 if the k th face was selected not the more competent, and zero if the k th face was 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

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

This implies a likelihood function for parameters c given data V

$$L(c|V) = \prod_{k=1}^K \Phi\left(\frac{\sum_{i=1}^I c_i v_{ki}}{\sigma\sqrt{2}}\right). \quad (3)$$

Again following Groseclose & 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 . After implementing the estimation, the probit coefficients are utilized as facial competence scores.

8.2 Details of Image Quality Estimation

A casual look at the candidate images will reveal that some are of higher quality than others. That some photos are taken on 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 in campaigns if one were to argue that facial competence has a direct effect on voters.

To ascertain 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 color 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, the dependent variable of interest.

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 spill over into such an effort. Instead, we implemented a measurement based on spatial statistics using each pixel in the image.

We first constructed a *variogram* of each image. The variogram is a common concept in spatial statistical studies that measures variance γ between the value of a process k at an arbitrary point s_i in a space S and every other point $s_i + 1, \dots, s_n$ in S that is at each distance d_0, d_1, \dots, 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. This number can range from 0 to 255.¹⁸ Each photograph was standardized to a matrix of 105x147 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

$$I = \frac{\gamma d_1}{\gamma d_{49}}. \quad (4)$$

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

¹⁸In a color image, each pixel would have three values, representing the red, blue, and green color values of that point

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.