

Snap Judgments: Predicting Politician Competence from Photos

Katherine Casey*

Abstract

Seminal studies show that naïve lab participants accurately predict who wins real-world elections based solely on candidate photos. It is unclear what this implies for the health of democracy without knowing whether candidates who look more electable or competent in photos behave more competently in office. This study brings novel performance data to this question and shows that voters can identify which politicians divert less public money and communicate more persuasively based solely on headshots. Such inferences do not predict politician effort visiting their constituencies, but neither do other available metrics like professional qualifications. I implement these studies in a low-income country where ballots include candidate photos and weak institutional checks raise the stakes for selecting innately competent leaders. Estimates provide an example of how voters' use of heuristics need not harm democratic accountability, and can actually enhance it, in cases where these shortcuts identify traits associated with good governance.

Key words: elections, candidate appearance, heuristics, ballot photos

* Stanford University Graduate School of Business, kecasey@stanford.edu. I am grateful to Allyson Barnett, Kelly Bidwell, Rachel Glennerster and the Innovations for Poverty Action Freetown office for their collaboration. I would like to thank the Editor and three anonymous referees, as well as Elizabeth Blankespoor, James Fearon, David Laitin, Neil Malhotra, Edward Miguel and Stephan Seiler, for helpful comments. Samuel Asher, Nicholas Eubank, Agnes Lahai, Anthony Mansaray, Fatoma Momoh, Osman Nabay, Katie Parry, Carlos Sánchez-Martínez and Korinayo Thompson provided excellent research assistance. Supplementary material for this article is available in the appendix in the online edition. Replication files are available in the JOP Data Archive on Dataverse (<http://thedata.harvard.edu/dvn/dv/jop>). This research received institutional review board approval from Stanford University (protocol #24823). Support for this research was provided by the Governance Initiative at the Abdul Latif Jameel Poverty Action Lab (grant No. MIT4145).

Seminal studies show that quick, reflexive judgments made by naïve lab participants confronted with pairs of candidate photos and video snippets predict who wins American elections with accuracy greater than chance. These inferences predicted the winner in 71% of recent Senate races (Todorov et al. 2005) and explained 20% of the variance in gubernatorial vote shares (Benjamin and Shapiro 2009). Forecast power based on nothing more than a few seconds' glance at candidate faces raises concerns about citizen reliance on appearance in casting votes and what this implies for the functioning of our democracy.

There is evidence, for example, that poorly informed voters who watch a lot of television respond most strongly to appealing-looking candidates (Lenz and Lawson 2011). Children, as largely uninformed and inexperienced citizens, make strikingly similar choices over candidate photos as adults (Antonakis and Dalgas 2009). Attractive candidates fare better in a variety of countries, drawing an uncomfortable parallel between elections and beauty contests (Banducci et al. 2008; Berggren et al. 2010, 2017). And the value of high-quality faces is not lost on parties, who field challengers with more competent looking faces to contest more competitive races (Atkinson et al. 2009). Voter responsiveness to the seemingly arbitrary trait of facial appeal fits into broader concerns that citizens do not base their decisions on things that matter for the functioning of democracy—namely policy and performance—and thereby erode electoral accountability (Achen and Bartels 2016; Lenz 2012).

The high prevalence of illiterate citizens, who rely on visual cues to cast accurate votes, adds a different perspective to this debate in lower income countries and leads many governments to print candidate photos directly on the ballot.¹ Evidence from Brazil shows how such efforts to reduce voting error can affect policy: the introduction of electronic voting, which includes

¹ In their study of voting procedures around the world, Reynolds and Steenbergen (2006) count more than 30 countries in their sample of 107 that use ballots with candidate photos (page 577).

candidate photos in a verification stage, reduced the frequency of invalid votes among the poorly educated, and this change in effective enfranchisement shifted government spending towards the priorities of low-income households (Fujiwara 2015). Yet ballot photos may risk negative unintended consequences: evidence from Uganda, for example, suggests that photos prime identity-based considerations, like ethnicity, in determining vote choice (Moehler and Conroy-Krutz 2016). Thus, to the extent that candidate photos are integral to reducing error, they pose a trade-off between facial appeal as a potentially misleading heuristic (Kuklinski and Quirk 2000; Lau and Redlawsk 2001) and as a tool to safeguard enfranchisement.

A key piece of missing evidence in assessing the consequences of these findings is whether and how the appearance of competence correlates with actual competence, which is where this study fits in. I combine voter assessments of politician photos with data on the performance of elected Members of Parliament (MPs) in Sierra Leone to evaluate whether voters can infer productivity-relevant traits from the way candidates look. I focus on three constructs: (lack of) corruption, constituency service, and ability to persuade. Strength in these areas improves MP performance—via minimizing the squandering of public funds, understanding what voters want and delivering relevant services, and convincing peers to support legislation—and is thus arguably valued by voters. Such innate competencies are particularly important determinants of performance where politicians are generally free from formal scrutiny: these MPs, for example, face no monitoring or reporting requirements for how they spend public funds earmarked for the development of their constituencies.

I approach this question in two steps. First, I establish proof of concept via a pre-election study that tests whether voter choices over candidate photos predict who later wins elections in Sierra Leone. This reproduces estimates that are strikingly similar to what Todorov et al. (2005)

and Benjamin and Shapiro (2009) document in the U.S., which verifies that the forecasting results generalize to this context. Second, I use data on the actual performance of both candidates and elected MPs in a subsequent study to test whether naïve voter inferences identify real skills and competencies. This is the main contribution of the paper, which I build from the following components.

To measure the three constructs of interest, I leverage rich and novel data on politician performance. Lack of corruption is captured by field audits that tracked how much of the constituency facilitation fund (CFF), a pot of public money allocated annually to each MP, could be verified on the ground as having been spent on local development projects. These audits also tracked how many times the MP had visited his or her home constituency to hold public meetings and consult with local residents, which I use as a measure of constituency service. Ability to persuade is captured by videotaped pre-election debates where Parliamentary candidates argued about which sector—e.g. education versus health—should be the top priority for additional government funding.

Study protocols elicited voter inferences about performance in these three areas based on candidate photos and other slices of audiovisual content. For corruption and constituency service, the study presented participants with a series of photo pairings among different elected MPs and asked them to guess which one in the pair performed better on that particular metric. I find that photo-based inferences correctly identify which MP spent more of the CFF on development for 60 percent of politician pairs, which is significantly greater than chance guessing. Photo inferences do not accurately identify which MP made more constituency visits, however neither do other readily available heuristics like their professional experience or years of education.

For persuasion, the study presented participants with pairs of winning and losing candidates

who argued for different sectors during the pre-election debates. Participants picked which sector to prioritize under three conditions: seeing only pairs of candidate photos; hearing audio recordings of the candidates' arguments during the debates, with no visual cue; and watching audiovisual clips of these same arguments. Estimates show that participant selections based on photos accurately predict the selections of *other* participants who only heard the audio recordings, suggesting that photo-based inferences detect superior oral communication skills. Winning candidates are, moreover, more persuasive than losers under all three conditions.

Directly connecting the results about inferred performance back to electoral forecasting is not possible, since performance in office is not observed for losing candidates. To indirectly close this loop, I show that post-election voter inferences about which politicians are less corrupt and more persuasive positively correlate with the actual vote shares those politicians received in the 2012 election. Analogous estimates for inferences about constituency service are null.

Regarding mechanisms, voter accuracy in inferring politician performance from ballot photos is broadly consistent with a common trait, like conscientiousness, driving greater politician care and effort in both presenting themselves publicly and executing the duties of public office. The persuasion results in particular are further compatible with the appearance of competence creating opportunities to practice and perfect one's leadership skills, creating a correlation between looking and sounding persuasive via socialization effects. By contrast, I find little evidence that the results are driven by voter inferences about candidate party identity or ethnicity, which is reassuring in an environment where party loyalty and ethnicity are tightly, and often contentiously, connected (Kandeh 1992).

The idea that voters may glean insights about candidates from their photos speaks to broader debates about voter responsiveness to seemingly irrelevant information (see Healy and

Malhotra 2013 for review). In a leading example, scholars dispute whether the electoral punishment associated with natural disasters reflects an emotional reaction to phenomena outside politicians' control (Achen and Bartels 2016), a measured assessment of government failure to adequately mitigate or prevent economic losses (Healy and Malhotra 2010), or a more complicated process of learning about incumbent type (Ashworth et al. 2018). The (ir-)relevance of politician appearance in photos is similarly debatable, and the central advantage of this study is data that links voter responses directly to known governance outcomes, side-stepping intermediating factors. Given the many barriers to informed citizenship, the results provide a glimmer of hope for electoral accountability and do so in a particularly challenging setting.

In what follows, I first provide background on the literature and institutional context, and briefly present the results of the electoral forecasting study. The bulk of the article concerns the inferred performance study: I present the design and results for each of the three constructs, connect the performance inferences back to electoral outcomes, and then explore mechanisms. The paper concludes with policy considerations.

BACKGROUND

Information is a critical resource for effective democratic participation, yet opportunities to acquire it are often limited and inequitably allocated (Delli Carpini and Keeter 1996). Where information is scarce, heuristic shortcuts might help (Lupia 1994; Popkin 1991), but only if they facilitate sound judgments (Kuklinski and Quirk 1998; Lau and Redlawsk 2001). And where they do not, they contribute to the broader disconnect between drivers of vote choice and the substantive matters of policy and performance that foster accountability (Achen and Bartels 2016; Lenz 2012). Ballot photos are one heuristic, accessible to all, but it is unclear whether they convey information that facilitates good decisions.

The literature assessing the relevance and consequences of voter responsiveness to candidate appearance proceeds in three steps. First, early work shows that inferences about competence based on headshots predict who wins actual elections. Todorov et al. (2005) exposed naïve lab participants to photos of the two major party candidates competing in Congressional races from 2000 to 2004. Participants picked which candidate in the pair appeared more competent, and these inferences predict the winner in 71% (67%) of Senate (House) races. Benjamin and Shapiro (2009) showed lab participants 10 second silent video clips of major party candidates taken from televised debates in gubernatorial elections from 1988 to 2002. Participant guesses about who won the race explain 20% of the variance in two-party vote shares, performing better than traditional economic forecasters.

Second, experimental manipulation establishes that photos shift voters towards more competent-looking candidates. Ahler et al. (2016) emphasize that the above correlations do not necessarily uncover a causal relationship: candidates who put more effort into looking presentable might also put more effort into their campaigns, and if it is the campaign that matters, the appearance effect is spurious. To gain traction, the authors controlled exposure to ballots with and without photos, and show that photos shift voting intentions toward candidates who independent raters thought looked like they would be better congresspersons. This suggests that part of what drives the electoral forecasting results is a direct voter response to the appearance of competence.

This study takes the third step and explores whether politicians who look more competent are, in fact, any more or less competent in carrying out their official duties. Why might this be the case? One explanation is an extension and reinterpretation of the point made by Ahler et al. above: the effort and conscientiousness that goes into making oneself look competent for a public appearance may be positively correlated with the effort and conscientiousness that one applies to

executing the duties of public office. If so, the appearance effect is not wholly spurious because it captures voter response to a signal (albeit likely a noisy one) of a common motivation that drives politician behavior in both domains. Alternatively, those who look like competent leaders may be afforded more opportunities to practice and perfect their leadership skills, thereby establishing a positive correlation via socialization effects. I return to discussion of these and other potential mechanisms in the penultimate section.

EMPIRICAL CONTEXT

Before detailing the data and study protocols, a few words about the institutional context are in order. The studies focus on the November 2012 Parliamentary elections and subsequent performance of elected MPs in Sierra Leone. For this election, there were 112 Parliamentary constituencies, which are single-member jurisdictions elected by first-past-the-post plurality. Politics is dominated by two major political parties, the All People's Congress (APC) and the Sierra Leone People's Party (SLPP). These two parties have, respectively, long standing ties to the Temne and other ethnic groups in the north of the country, and the Mende and other ethnic groups in the south (Kandeh 1992). While these ethnic-party ties are strong, two studies provide evidence that citizens in Sierra Leone do in fact respond to candidate-specific attributes when deciding whom to support. Casey (2015) documents a higher prevalence of voting across these ethnic-party lines in local elections where citizens have better information about individual candidates; and Bidwell et al. (2020) demonstrate that voters are more likely to support candidates who perform well during public debates. Note that the Presidential election was held concurrently with the Parliamentary races; and while turnout is not mandatory, it is consistently high and

reached 87 percent for this election.²

Sierra Leone is a very poor country, ranking 184th out of 189 countries evaluated by the United Nations Human Development Index, which is a comprehensive measure of development that captures health, education and standards of living.³ Voters have correspondingly low levels of schooling, for example the majority of study participants have no formal education, and limited access to mass media: television ownership, for instance, is at 16 percent (National Public Services Survey 2011).⁴ The high prevalence of illiterate voters motivates the use of ballot photos in Sierra Leone, which provide visual cues to help citizens cast accurate ballots. This is an important context to examine voter inferences from ballot photos, because both the potential benefits—as a tool to safeguard enfranchisement—and downside risks—e.g. of priming ethnicity or misleading voters when political information to correct misperceptions is scarce—are substantial.

PROOF OF CONCEPT: ELECTORAL FORECASTING

The first pre-election study establishes that naïve inferences predict who wins elections in Sierra Leone. The study adapts protocols developed by Todorov et al. (2005) and Benjamin and Shapiro (2009), with two enhancements: it uses a random sample of registered voters, who more closely represent the population of interest than self-selected students (Henrich et al. 2010); and was conducted *before* the election. I use photos and video clips of candidates from the two major political parties competing in twelve Parliamentary races. This media comes from the filmed pre-election debates, converted into black and white (see Appendix Figure A1 for an example).

Enumerators administered study tasks via tablet device during home visits to 407 registered

² National Electoral Commission of Sierra Leone, accessible here: <http://necsl.org/PDF/Documents/STATEMENT%20%20FROM%20NEC%20CHAIRPERSON.pdf>

³ Accessible here: <http://hdr.undp.org/en/2018-update>.

⁴ Accessible here: <https://doi.org/10.7910/DVN/E3AOFV>.

voters in the 5 weeks preceding Election Day. They informed participants that the images were of real, contemporary candidates and that the order of the two major party candidates varied randomly across pairs. Participants evaluated each candidate individually on a few dimensions (e.g. leadership, corruption) and were then shown the two candidates together and asked: “if you had to vote today, which candidate would you choose?” (Data and replication code for analysis in this article are available in Casey 2021.)

These “vote today” choices predict election winners with probability greater than chance. Figure 1 shows that participants selected the eventual winner 55% of the time, which is significantly better than random guessing at the 99% confidence level for both the photo and video conditions ($N = 3,581$ participant ratings for each condition). Aggregated up to the race level ($N = 12$), photo selections accurately forecast 75% of elections, which is statistically distinct from chance at 90% confidence. Video inferences similarly forecast 67% of races, however this is not statistically distinguishable from chance at the race level. Regression analysis (in Appendix Table A1) shows that “vote today” choices explain one third of the total variation in actual vote shares, and assessments of who looks less corrupt correlate positively with differences in vote shares. The results are quite similar to those from the U.S., suggesting that links between appearance and electoral success cut across cultures (Lawson et al. 2010). It further attests to the strength of the results in the original studies, in an environment where the (non-)reproducibility of findings in social science has become a widely debated concern (Open Science Collaboration 2015).

INFERRING POLITICIAN PERFORMANCE

Having thus established the existence of a face-vote correlation in Sierra Leone, the second study asks whether snap judgments detect qualities associated with performance: i.e. given that “vote today” choices predict winners, and corruption assessments contribute to this, is there evidence

that politicians who look less corrupt in their headshots handle public funds in a more trustworthy way? The associated tests use data on politician performance over their first 18 months in office, which was collected in related work (Bidwell et al. 2020). Analysis centers around three competencies: public spending performance, constituency service via in-person visits, and persuasive oral communication.

The first two competencies studied are by definition only observed for elected officials, necessitating a tactical shift away from winner-loser comparisons. My strategy is to first maximally leverage the performance data by forming many different pairings among winners. Linking these results back to electoral forecasting requires an inferential leap of faith: it assumes that if performance is detectable in photos of winners, it is similarly observable in photos of losers, which is difficult to test. I return to this challenge below. Second, I examine persuasion as a third competency that is observable for both winners and losers, and conveniently for research purposes, public oratory occurs naturally both during the campaign and in Parliament.

Empirical tests are based on voter evaluations elicited during a second wave of data collection, which recruited participants from household listings of registered voters and administered tasks during home visits. In total, 399 voters were randomly sampled from twenty rural polling centers, located in three districts around the country. Overall, 51% of participants were female, with an average age of 38, and 53% had no formal schooling. By construction, the study was conducted after the election (in January 2016), necessitating precautions that participant choices do not reflect external knowledge. Here the weak information environment works to our advantage: the vast majority of respondents have no access to television or newspapers, so the likelihood of seeing pictures of politicians outside their own constituency is low. Moreover, no participant is registered to vote in a constituency represented by an MP in the sample, and I drop

observations where the participant recognized either politician (which affects less than 1% of participant-photo pair observations). To keep the workload manageable, and create independence between statistical tests, the many distinct tasks were apportioned across participants (see schema in Appendix Table A3).

Competence 1: Public Spending Performance

The first task in the post-election study links photo-based inferences about corruption to audit data on how elected MPs managed public funds. It focuses specifically on MPs' handling of the CFF: did they use it for development projects or spend it on themselves? How the CFF was spent in practice was measured via extensive field audits that tracked each expenditure line item that an MP claimed to have used for development to its source in the MP's home constituency. The audits covered 27 Parliamentary constituencies located in 11 districts around the country. Audit enumeration teams traveled to the specific location of each purported project within the constituency, where they: i) conducted multiple interviews and two focus group discussions with local stakeholders to establish the timing, cost and funding sources for each project; ii) physically examined all purported projects; and iii) reviewed all available records associated with each project (e.g. receipts for purchased inputs, community log books). Any remaining unverified funds were either spent on legitimate transport costs for MP travel to the constituency or diverted to other unsanctioned uses. In practice, diversion of the CFF is substantial: audits verified zero development expenditures for 21 percent of the sampled MPs.

The study used the official MP headshots posted on Parliament's website, converted into black and white and rescaled for comparability (see Appendix Figure A1 panel B for an example). To form pairs, I ordered the 27 MPs in the audit data by their CFF performance and made 55 different matches between MPs in the top and bottom half of the distribution. Since these MPs

participated in a field experiment (analyzed in Bidwell et al. 2020), I form all pairings within that treatment assignment and regressions include an assignment indicator.⁵ The left-to-right ordering of the high versus low performer varied randomly both across MP pairs and across different respondents evaluating a given MP pair. Enumerators introduced the task with the following prompt: “Every year, Honorables receive 44 Million Leones [US\$11,000] in the Constituency Facilitation Fund (CFF). Some MPs use this money for development projects in their constituencies, some use it to travel back and forth to Freetown, and some put the money in their own pocket. I am going to show you photos of real Honorables (MPs) who are not from your constituency, but are from other parts of Sierra Leone. At the end of the year, we checked how much of the money they actually spent on development. In each pair, one MP spent more of the CFF on development than the other one. We mixed up the order of who spent more and who spent less in each pair. Do your best to guess which MP spent more of the CFF on development and put less in his own pocket.”

These photo-based inferences strongly predict MP performance managing public funds. Across 2,158 evaluations of photo pairs, participants chose the MP who in fact performed better 54.2% of the time (standard error 1.1), which is significantly better than random guessing at 99% confidence (p -value <0.001). The wisdom of crowds is somewhat stronger: Figure 2 plots the distribution of correct guesses across the 55 distinct photo pairings and shows that study selections on average correctly identify the true better performer in 60% of pairs. Interestingly, there is wide variation across pairs: in some cases, nearly all participants picked the better performer, while others seemed harder to read.

⁵ The Bidwell et al. (2020) study evaluates the impact of participating in debates as candidates on MP behavior in office. By forming photo pairings within that study’s treatment assignment (e.g. comparing photos of treated MPs only to photos of other treated MPs) analysis here controls for any impacts the debates had on MP behavior.

Appendix Figure A2 shows this relationship in the raw data, plotting the monetary difference in what was verified in the CFF audits for each MP in the pair (first minus second) against the share of participants selecting the first MP as the bigger spender. As expected, the upward sloping line suggests that photo assessments predict spending differences.

Photo inferences are more reliable than other observable characteristics in predicting development spending. To gauge this, I leverage the fact that the same MP was compared to several peers (4 on average) in different photo pairings, producing 110 observations of trustworthiness shares by pair. I regress these on fixed effects for the 27 MPs in the sample. The estimated coefficients, which I use as an explanatory variable in Table 1, capture the intangible features that make a particular MP appear more or less trustworthy than peers. These inferences strongly outperform professional qualifications in predicting who spent the CFF on development.

Specifically, the positive and statistically significant coefficient in the bivariate specification of Column 1 reproduces the predictive relationship of Figure 2 in regression form. Compare this to column 3, which includes only the observable characteristics, where an *F*-test cannot reject that all coefficients are jointly equal to zero. Column 5 evaluates all characteristics in a single regression. The coefficient on inferred spending is positive, highly significant, and large: moving from the 25th to 75th percentile of the inferred spending distribution corresponds to the audits verifying \$6,518 more development spending, or 59% of the annual allotment. Only one professional qualification—education—enters significantly, and is in fact negative. Note that adding inferred spending to the set of explanatory variables more than doubles the R^2 (when moving from column 3 to 5). Given the small number of observations, I run two-way comparisons between inferred spending and each observable characteristic: the difference in coefficients is positive and significant at 94% confidence or above for all six tests (see Appendix Table A4).

Competence 2: Constituency Service via Visits

In contrast to public spending, voters came to the wrong conclusions about politician delivery of constituency service based on headshots. As a measure of service, the enumerators conducting the audits surveyed several respondents in each MPs' constituency about how many times the MP had visited his or her constituency since the election (which varies from 0 to 6 visits in the data). I formed 56 headshot pairs between high and low performers. Enumerators introduced the study task with an explanatory prompt: "It is the Honorable's job to represent the people from their constituency, and they can make sure they are doing a good job by visiting their constituency regularly. I am now going to show photos of real Honorables who are not from your constituency, but are from other parts of Sierra Leone. We checked how often they visited their own constituencies. In each pair, one MP visited his constituency more than the other, and the order is mixed up. Do your best to guess which MP works harder by visiting his constituency more often."

Participants selected the MP who actually made more constituency visits 45.7% of the time, which is significantly worse than random guessing at 99% confidence. This negative effect is apparent across the analogous series of regression estimates as completed for public spending, albeit with somewhat less precision (in Tables 2 and A5). Considering the omnibus specification of Table 2 column 5, the coefficient on photo inferences remains negative, yet falls below standard confidence levels. Notably, the one coefficient that is highly significant in this specification is age, where its negative sign is intuitive if older politicians are less inclined to undertake taxing journeys. Yet note also that the respective *F*-tests for inferred visits and all observables taken together cannot reject that both are equal to zero. This suggests that items typically found on an MP's vitae—with the potential exception of age—do not present a strong alternative forecaster for voters to rely upon. Thus, while photo-based inferences do not seem useful for voters in inferring an MP's likely

level of constituency engagement via visits, it appears that other readily available measures are not good predictors, either.

Competence 3: Persuasion and Oral Communication Skills

Tests of voter ability to infer the third competence, persuasion, are structured somewhat differently than those for the other two competencies above. Study tasks use the pre-election debates, where candidates argued for a specific sector to receive greater government funding (the modal sector advocated was education, followed by health, roads and agriculture), to form 72 pairings of winning and losing candidates (some within, some across races) who argued for different sectors. Enumerators explained to study participants that “Each year, Parliament is responsible for approving the budget for the Government of Sierra Leone. MPs have the opportunity to influence what kinds of projects are prioritized in the country. There is a limited amount of funds, so some problems must receive more attention than others.” For each pair, participants picked which sector to prioritize for greater funding, under three conditions: sectors accompanied by candidate photos; audio recordings of the candidates’ arguments promoting the sector, with no visual cue; and audiovisual clips of the same arguments.

More specifically, in the photo condition the participant sees the two photos side-by-side and the enumerator asks a question like, “The first MP candidate argues for making education a priority and the second MP candidate argues for making agriculture a priority. Which issue would you choose?” (The prompts are tailored to match the actual sectors the candidates argued for in all conditions.) For the audio condition, participants listen to the two candidates making their arguments during the debates, and answer an enumerator prompt like, “Would you choose to make education a priority or to make agriculture a priority?” The video condition is the same as audio, save participants both see and hear the debate arguments. Questions positioned at the beginning

of the study protocol elicited participants' own preferences over all sectors (see Appendix Table A3). Analysis uses these conditions to test whether: (i) voters identify candidates with stronger oral communication skills from photos alone; and (ii) candidates who subsequently win are more persuasive than those who lose.

For the first question, note that if candidate image does not matter, then participants in the photo condition should always select the sector that ranks higher in their own preferences. But this happens only 74% of the time. For the remainder, voters either tradeoff information gleaned from candidate photos or make arbitrary mistakes. Consistent with their responding to something about persuasive skill, Table 3 shows that the mean share of participants who picked the first sector in a given candidate pair based on photos is highly predictive of whether *other* participants select this same project under the audio condition. The magnitude of this effect is substantial: a 10 percentage point increase in the photo share corresponds with a 6 percentage point increase in the individual likelihood of picking that project based on oral arguments. This suggests that photos convey information about persuasive skill that goes beyond any appearance advantage enjoyed by winners. Column 3 shows that the relationship is robust to controlling for sectoral rankings and whether the first project was endorsed by the eventual winner. Appendix Figure A3 presents an analogous two-way scatterplot of the raw data and shows a strong positive relationship between looking persuasive in photos and having better oral communication skills.

On the second question, I find that candidates who subsequently won are more persuasive than those who lost under all three study conditions. Figure 3 shows that for the 72 candidate pairs, on average 60.5% of participants selected the sector backed by the eventual winner's photograph, consistent with winners having an appearance-based advantage in persuasion. This is strongly distinct from random choice (which would be 50.0%), as the associated two-sided *t*-test rejects

statistical equivalence at 99% confidence (p -value = 0.001). Without any visual cue, 57.8% of participants selected the eventual winner's sector based on audio arguments, consistent with them possessing superior communication skills (p -value = 0.001). The estimate for video, which combines visual and oral cues, is slightly larger in magnitude than both at 61.0%, again strongly distinct from chance (p -value < 0.001), but only statistically larger than the audio condition (p -value on the relevant difference in estimates is 0.065). Results are robust to controlling for partisanship and sectoral alignment (see regression analogues in Appendix Table A6).

Multiple Inference Adjustments

Analyses here and in the appendix test several inter-related predictions—voter inferences about four political outcomes under up to three content dissemination conditions (photo, audio, and video)—calling for some adjustment to account for the fact that these are not wholly independent tests. To be conservative, I pool my 4 main predictions, about photo inferences, together with all additional study conditions (another 5 audio and video conditions). For the pre-election study, this is a straightforward correction to a multiple inference problem, as analysis tests the accuracy of forecasts under two conditions on the same participant dataset (all participants completed both tasks). The post-election studies are more nuanced, as they administer 20 different versions of the protocol to both apportion the work across participants and to create independence between tests (see Appendix Table A2). Appendix Table A7 thus pursues a prudent approach and adjusts across all 9 outcome predictions collected in both studies taken together. These multiple inference adjustments do not substantively change the interpretation of results. Most importantly, for the central question of what voters might be able to infer from ballot photos of candidates, the four relevant photo predictions are robust to false discovery rate (FDR) adjustments at both the individual voter- and race-level (in columns 3 and 7), and 7 of these 8 total estimates are robust to

conservative family-wise error rate (FWER) adjustments (in columns 4 and 8).

CONNECTING INFERRED COMPETENCIES TO ELECTORAL OUTCOMES

This article began with the question of how much of the predictive power of photo-based inferences might be reasonable because they identify real-world competencies. Analysis thus far has not closed the loop on this question, as it has not connected the forecasting results of the pre-election study to the inferred performance estimates of the post-election studies. The ideal way to do this would be to evaluate whether pre-election forecasts about which candidate is more likely to win predict productivity differences between the same two candidates in office. This is fundamentally not possible given the unobserved counterfactual for losers: we never see whether they would have performed better or worse than the winner if they had instead won the race. To nevertheless make modest progress on this front, this section attempts to link the forecasting and performance results by seeing whether inferences about candidate competencies correlate with electoral outcomes. There is a timing disconnect here, as the performance outcomes and associated study inferences occurred after the election, which renders any correlation as speculative about, but not evidence of, the reasonability of forecasting.

Table 4 presents estimates from a regression of realized vote shares from the 2012 Parliamentary election on politician competencies inferred from photo pairings. Columns 1 through 3 extend the empirical approach of Tables 1 and 2 by leveraging the fact that winning MPs are compared to several different peers in the public spending and visit tasks. Making comparisons among winning MPs only, the positive and significant coefficient on inferred trustworthiness in column 1 suggests that MPs who on average look less corrupt than their peers in photos received larger vote shares in the 2012 election. The null result for inferred visits in column 2 suggests that inferences about constituency service do not predict differences in vote shares. Column 3

compares both inferences to the characteristics that might be easily observed on a politician's vitae, and shows that spending inferences remain positive and significant, visit inferences are weakly negative, and no observable characteristic has predictive power at conventional levels. Results for the winners' first versus second place finishing margins are similar, save inferred visits is not significant in any specification (Appendix Table A8).

Columns 4 and 5 extend this approach to the variety of winning-losing candidate pairings used in the persuasion study tasks. Here each winner (loser) was on average compared to four different losers (winners). The positive and significant estimates for inferred persuasiveness suggest that candidates whose priority sector was chosen by study participants more frequently (averaged across multiple photo pairings) received larger vote shares in the 2012 election. Column 5 shows that a couple of resume characteristics—namely professional experience managing ten or more employees and years of education—also positively correlate with vote shares, suggesting that voters reward these observable traits in the polling booth. The fact that these observables have more predictive power in this specification compared to column 3 could be due to the fact that there is more variation in the winner-loser sample (for example the standard deviation for education is slightly larger at 2.2 versus 1.9 years) or that the sample of observations is somewhat larger (34 versus 27).

Drawing connections across studies, these results are cautiously optimistic: they suggest that voter inferences about public spending and persuasiveness predict both politician performance and electoral success. The null estimates for inferred constituency service somewhat assuage concerns about the risk that these inferences systematically mislead voters: inferred visits does not robustly predict electoral outcomes, and to the extent that it does, the negative sign suggests that it pushes vote shares toward higher effort MPs (since inferred visits negatively correlates with

actual constituency visits in Table 2).

MECHANISMS

Political images, like ballot photos, are highly selective acts of self-promotion: what politician's wear, their hairstyle, facial expression and posture are deliberate choices intended to boost their electoral appeal. Such images also reveal features that are harder to manipulate—like those shaped by a candidate's age, gender or ethnicity. This combination creates scope for multiple channels through which political imagery could convey information about how a politician might perform in office. This section attempts to adjudicate which of these potential mechanisms seem more or less consistent with the observed constellation of results.

Conscientiousness

If candidates who invest more care and effort in presenting themselves publicly also invest more care and effort in executing the duties of public office, then the accuracy of perceived competence may reflect a common motivation—call it conscientiousness—driving candidate behavior in both realms. This mechanism would rationalize voter accuracy in inferring the spending performance of elected MPs if those who are more meticulous in preparing themselves for their official headshot are also more meticulous in handling public funds. It would similarly rationalize the persuasion results if candidates who more carefully prepare their appearance for a televised debate also more carefully prepare what they plan to say about policy during the debate. And, since effectively managing public funds and advocating for policy are key to getting things done in Parliament, it would be reasonable for voters to respond to visual cues of conscientiousness in the polling booth.

For this mechanism to hold, the appearance of competence must to some extent be manipulable. Studying the attractiveness of celebrities, for example, Jenkins et al. (2011) establish

substantial variation in participant evaluations of the same person across different photos, and conclude that “facial attractiveness is not determined solely by the face, it is also determined by the photo” (p. 7).⁶ The mechanism further requires that voters respond favorably to strategically crafted or selected images. On this point, Todorov and Porter (2014) show that when presented with multiple photos of the same individual, lab participants broadly agree about which photo best suits a particular scenario—e.g. the dominant photo choice for a hypothetical mayoral campaign is not the same as that for a social media post. Moreover, when these scenarios are presented to *other* study participants, they in turn draw more favorable inferences from those accompanied by a purposefully selected photo than those paired at random with a photo of the same individual.

Applying the first point to this empirical context, I can measure within-politician variance in perceived competence for a subset of MPs who were evaluated twice over the course of this study using different images: first based on a photo still taken from the public debates (in study 1) and second based on their official Parliamentary headshot (in study 2). Consistent with the variable nature of appearance, Appendix Table A9 shows that there is as much variation in mean inferred leadership and corruption scores within-MP across the two photos as there is across MPs for a given set of photos.⁷

On the second point, I find that voter accuracy in inferring which MP spent more public funds on development is strongest when the true higher spender is the one who looks relatively more favorable—as indicated by looking wealthier—in the photo pair. Note that interpreting this pattern through the lens of conscientiousness relies on the plausible but untested assumption that

⁶ For an application to political candidates, see for example, Rosenberg et al. (1991)’s exploration of the role of professional make-up artists as distinct from candidate attractiveness in boosting electoral appeal.

⁷ More precisely, for each MP trait we have a mean score (i.e. the average photo evaluation across all participants) for each study. Within-MP variation is the variance in mean scores for a given MP across the two studies, averaged over the 12 MPs. The across-MP variation is the variance across the 12 MP mean scores within a given study.

investing more care and effort in self-presentation makes one appear wealthier in the eyes of voters. For this analysis, a separate set of 80 study participants viewed the same 55 photo pairs used in the spending inference task (presented in Figure 2) and picked which one in the pair “looks richer.” Appendix Figure A4 fits a nonparametric relationship between the proportion of study participants who correctly identified the MP who spent more on development and the proportion of (other) participants who thought that same MP looked like the wealthier one in the pair. The upward sloping line shows a strong positive correlation between the proportion of study participants who correctly identified the less corrupt MP and the proportion of others who thought that the less corrupt MP in the photo pair looked wealthier. This suggests that the rate of correct inference is highest where the less corrupt MP was judged by other voters as looking more favorable.

What additional data and analysis would increase our confidence in the plausibility of this mechanism? Most immediately, one would want to measure conscientiousness directly, and test whether it positively correlates with both effort invested in self-presentation and performance in office. It would further be instructive to parse the dimensions of self-presentation that appeal to voters and trace investments in these specific areas back to high performers. On this point, we can gain some confidence from the fact that several arbitrary differences visible in our sample of photos do not appear to spuriously drive the results. Specifically, Appendix Table A10 shows that differences in backdrop, photo brightness, brand of camera used, presence of facial hair or glasses, Western versus traditional dress, and whether the MP is smiling, do not predict differences in inferred trustworthiness. Lastly, it would be important to account for the null results observed for constituency service. If additional data and analysis showed, for example, that conscientiousness was uncorrelated with the frequency of visits, and that voters value public spending more than visits, then the full pattern of results could be reconciled with this mechanism.

Halo and Socialization Effects

There are, of course, more stable differences in appearance—e.g. the celebrities in Jenkins et al. and my subsample of MPs can be clearly rank ordered by their overall attractiveness and perceived competence score—which raises the possibility of socialization effects.⁸ In an influential meta-analysis, Langlois et al. (2000) summarize the evidence from social psychology that shows how this channel can create a self-fulfilling prophecy for attractiveness: i) observers broadly agree on who is attractive; ii) observers infer more positive traits, including occupational competence, about more attractive targets (or “halo” effects); iii) people treat attractive targets more favorably; and iv) attractive targets in turn behave differently, with impacts on real-world outcomes including their occupational success. Regarding the last link in the chain, there are questions about whether attractive people do in fact behave differently (Maestripieri et al. 2017), and more importantly for our purposes, whether any potential behavior differences affect their occupational *productivity*.

Similar questions have been raised in economics, where an early contribution by Hamermesh and Biddle (1994) documents an attractiveness wage premium across a broad spectrum of occupations, many of which seem unlikely to have any link between attractiveness and productivity. In a clever design, Mobius and Rosenblat (2006) parse this beauty premium into *direct* taste-based discrimination by employers and an *indirect* worker confidence channel. To establish the former, they show that “employers” in an experimental labor market pay attractive “workers” more when exposed to their photo, even though there is no relationship, by construction, between looks and productivity on the “job” (which is solving mazes). The latter indirect channel shows how physical attractiveness lends confidence to workers, a trait they convey even through

⁸ For example, when you compare the trustworthiness ranking of the 12 MPs based on the study 1 photos (the debate stills) to that based on study 2 headshots, the same 4 MPs appear in the bottom third of both distributions.

oral communication. Here they find that attractive “employees” are able to negotiate higher wages even when they communicate with their “employer” by phone, with no visual image. What is interesting about testing this dual channel logic in the realm of politics is that *both* looking and sounding more persuasive could very well enhance one’s productivity in public office.

The third set of tasks in the Sierra Leone study was designed to test just that. Applying the idea of a direct channel linking appearance to persuasion would imply that the “looks of a leader” add legitimacy to one’s policy proposal in the eyes of others. This would explain why participants in the photo-only condition do not always pick their own most preferred sector. It would further suggest that such “looks” could enhance how an MP’s peers evaluate their policy arguments in Parliament. If so, these two facts alone could rationalize the link between ballot photo inferences about persuasion and electoral success. More interesting though, is the second, indirect channel, whereby looking persuasive opens opportunities to build communication skills. If so, this creates a positive correlation between looking persuasive and being persuasive in political dialogue and debate. The finding that photo-based evaluations of which MP is more persuasive predict audio-only evaluations by *other* study participants is consistent with this second, indirect channel.

A few clarifications and caveats are in order. Unlike the literature referenced, this study is not about attractiveness *per se*.⁹ It links observations of persuasion—measuring MP influence over which project a voter picks—across different stimuli (photos, audio) without relying on intervening voter evaluations of MP appearance. This channel also does not require an indelible biomarker of trustworthiness, a concept which the review by Todorov et al. (2015) discredits, and could instead work in combination with the conscientiousness channel above—i.e. effort invested

⁹ For a study that bridges attractiveness and persuasion, see Palmer and Peterson (2016), who find that interviewers give attractive survey respondents a higher subjective political knowledge score, even after controlling for the respondent’s actual knowledge based on responses to survey questions. Attractive respondents are moreover more likely to self-report having tried to persuade others to vote for a particular candidate during the campaign period.

in looking competent opens up avenues for building communication skills. Lastly, consider a potential confound that builds on the party selection argument of Atkinson et al. (2009). Suppose parties invest more resources in recruiting high quality candidates (on all relevant dimensions) where they are more likely to win. Then dominant party candidates—the eventual winners—will be more likely to both look like a leader and have strong oratory skills (as compared to underdog candidates—the eventual losers), even if there is no correlation between these two traits in the general population. Since the Sierra Leone study uses pairs of winning and losing candidates, this rationale would not invalidate the main result that voters make correct inferences about oral communication skills from photos, but it would suggest that this is due to statistical discrimination by voters and not socialization effects over the candidates’ life cycle.

Partisanship

Unlike the channels above, the data clearly reject that inferences about partisanship explain the observed results. Voters in Sierra Leone tend to be strong partisans: for example, 90% of study 1 participants said they “will definitely” vote for the MP candidate from their preferred party in the election that was, at the time, only a few weeks away. If candidate appearance, or the statements they make, reveal information about which party they belong to, study results about electoral forecasting or persuasion may be somehow confounded with these partisan cues. In Western countries, for example, Berggren et al. (2017) show that conservative candidates tend to be more physically attractive, so beauty could function as a low information heuristic for ideology. Benjamin and Shapiro (2009) find that while study participants struggled to infer party from appearance, they did better when they had access to sound clips of what candidates said.

Analysis of the Sierra Leone studies finds little evidence that partisanship plays much of a role. In the video condition of forecasting study 1, 52.7% of participants correctly identified which

candidates in the pairs were from the APC party, which is statistically greater than chance. These inferences, however, did not on net enable participants to select their preferred party: APC supporters in the study were statistically no more likely to “vote today” for the APC candidate after watching the video clips than SLPP supporters (54.4% versus 54.9%). Similarly, in study 2 voters correctly guessed party affiliation based on photos 47.8 percent of the time, and only 35.5 percent based on video clips. Thus, inferences about partisanship do not appear consequential for the electoral forecasting results.

Party identity or the types of policies parties endorse could also play a role in the persuasion tasks, which might manifest in two distinct ways. First, while priority sector (the debate question used in this study) is generally party neutral, the ruling party at the time oversaw a major healthcare reform. If advocating for health during the debate reveals party identity, it could steer voters toward their own party’s candidate. Second, successful parties or candidates might have better information about voter preferences and use them to argue for sectors more in line with what voters want. Indeed, the project endorsed by the winning candidates (in the winner-loser pairs) ranked higher in study participants’ own sectoral rankings 57% of the time, which is greater than chance.

Appendix Table A6 tests whether the observed result that winners are more persuasive than losing candidates (in Figure 3) is robust to controlling for these channels. It presents results from regression analysis that predicts when a study participant chose the first project among the two advocated in a given video pair. The coefficient on the first project being endorsed by the candidate from the voter’s own party is small in magnitude and not statistically distinguishable from zero, providing little evidence for a partisanship channel. By contrast, the positive and significant coefficient on the first project ranking higher in the participant’s own rankings of public goods suggests that sectoral preferences matter for deciding which project to fund. Most

importantly, the positive and significant coefficient on the first project being endorsed by the winning candidate shows that the winner's persuasion advantage is robust to the inclusion of these two alternative mechanisms.

Ethnicity

Inferences about ethnicity also do not appear to explain the forecasting results, which is important in light of the salient role ethnicity plays in politics (Kandeh 1992) and empirical evidence from Uganda that ballot photos can increase the prevalence of ethnicity-based voting (Moehler and Conroy-Krutz 2016). Such concerns about ballot photos appear less relevant in Sierra Leone, for the simple reason that participants were not very accurate at inferring ethnicity from headshots.

In study 1, participants correctly guessed candidate ethnicity based on photos only 14% of the time. One way to think about how this number compares to random guessing (which would not be 50% in this case) is to benchmark the correct guess rate against national population shares of each ethnic group. For example, 32% of the national population is from the Mende ethnic group, as are 27% of candidates in the photo sample used in the electoral study. When shown photos of Mende candidates, however, study participants only correctly identified them as such 23% of the time. A hit rate that falls below the relevant population share applies to nearly every ethnic group represented in the candidate sample. To establish another benchmark, in a final study task the enumerator read the name of each candidate aloud and asked the participant to guess that individual's ethnic group. The percent correct based on names was 27%, and the twofold difference in accuracy is highly statistically significant ($p < 0.001$). From a policy perspective, given that including candidate names on the ballot appears unavoidable, the addition of photos seems

unlikely to distort voting behavior due to inferences about ethnicity in this setting.¹⁰

CONCLUSION

There are many reasons to be pessimistic about voter ability to thoughtfully select and discipline politicians. Access to political information is scarce (Delli Carpini and Keefer 1996); those who have it, may ignore it when they vote (Achen and Bartels 2016); and even those who use information sensibly might induce unintended policy consequences (Ashworth and Bueno de Mesquita 2014; Healy and Malhotra 2009). Challenges of poverty, illiteracy and intimidation make accountability's prospects look even bleaker in many low-income countries.

Against this backdrop, findings from Sierra Leone offer a modicum of hope. In a highly constrained information environment, voter responses to candidate photos predict subsequent election results, and analysis establishes a governance-enhancing rationale for why this might be so: candidate appearance helps voters identify better stewards of public money and those who argue more persuasively for government policies. The strength of this counterpoint comes from connecting voter inferences directly to concrete measures of politician traits and skills, which relies on rich and novel datasets. It is worth stressing, however, that a variety of mechanisms could explain these results, which are difficult to parse definitively. Moreover, the empirical relationships between inferences about candidate traits and their productivity in office are based on a particular sample of candidates and voters in a single election and may reflect something exceptional about this group. The extent to which these relationships replicate in other populations and contexts is an important question for future research.

¹⁰ In terms of other demographics, gender also does not drive these results. For all three performance metrics, voter inferences are no more or less accurate for mixed versus same-gender pairs, and accuracy estimates are nearly unchanged (in both magnitude and significance) when mixed pairs are excluded from the analysis (see Appendix Table A11).

Do these observed advantages come with an apparent policy downside in this context? While candidate photos appear to mislead voters about the likely effort MPs put into constituency visits, other metrics that could be found on a politician's vitae do not seem particularly useful either. Voter inferences also do not appear to exacerbate ethnicity-based voting, for the simple reason that ethnicity seems difficult to detect from photos in Sierra Leone. Regarding electoral practices, this study finds little cause for concern about, and indeed potential benefit from, including photos on ballots in light of their informational content and original purpose to aid illiterate voters.

REFERENCES

- Achen, Christopher, Larry Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton: Princeton University Press.
- Ahler, Douglas, Jack Citrin, Michael Dougal, Gabriel Lenz. 2016. "Face Value? Experimental Evidence that Candidate Appearance Influences Electoral Choice," *Political Behavior* 39:77–102.
- Anderson, Michael. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedaian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association* 103(484):1481-95.
- Angrist, Joshua, Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton: Princeton University Press.
- Antonakis, John, Olaf Dalgas. 2009. "Predicting Elections: Child's Play!" *Science* 323:1183.
- Atkinson, Matthew, Ryan Enos, Seth Hill. 2009. "Candidate Faces and Election Outcomes: Is the Face-Vote Correlation Caused by Candidate Selection," *Quarterly Journal of Political*

- Science 4:229-249.
- Ashworth, Scott, Ethan Bueno de Mesquita, Amanda Friedenberg. 2018. "Learning About Voter Rationality," *American Journal of Political Science* 62(1): 37-54.
- Ashworth, Scott, Ethan Bueno de Mesquita. 2014. "Is Voter Competence Good for Voters?: Information, Rationality, and Democratic Performance," *American Political Science Review* 108(3):565-587.
- Ballew, Charles, Alexander Todorov. 2007. "Predicting political elections from rapid and unreflective face judgments," *PNAS* 104(46): 17948–17953.
- Banducci, Susan, Jeffrey Karp, Michael Thrasher, Colin Rallings. 2008. "Ballot Photographs as Cues in Low Information Elections," *Political Psychology* 29:903-917.
- Benjamin, Daniel, Jesse Shapiro. 2009. "Thin-slice Forecasts of Gubernatorial Elections," *Review of Economics and Statistics* 91:523-536.
- Benjamini, Yoav, Abba Krieger, Daniel Yekutieli. 2006. "Adaptive Linear Step-Up Procedures That Control the False Discovery Rate," *Biometrika* 93:491–507.
- Berggren, Niclas, Henrik Jordahl, Panu Poutvaara. 2010. "The Looks of a Winner: Beauty and Electoral Success," *Journal of Public Economics* 94:8-15.
- Berggren, Niclas, Henrik Jordahl, Panu Poutvaara. 2017. "The right look: Conservative politicians look better and voters reward it," *Journal of Public Economics* 146:79-86.
- Bidwell, Kelly, Katherine Casey, Rachel Glennerster. 2020. "Debates: Voting and Expenditure Responses to Political Information," *Journal of Political Economy* 8: 2880-2924.
- Casey, Katherine. 2015. "Crossing Party Lines: The Effects of Information on Redistributive Politics," *American Economic Review*, 105(8): 2410-2448.
- Casey, Katherine. 2021. "Replication Data for Snap Judgments: Predicting Politician Competence

- from Photos,” <https://dataverse.harvard.edu/dataverse/jop>.
- Delli Carpini, Michael, Scott Keeter. 1996. *What Americans Know About Politics and Why It Matters*. New Haven: Yale University Press.
- Fujiwara, Thomas. 2015. “Voting Technology, Political Responsiveness, and Infant Health: Evidence from Brazil,” *Econometrica*, 83(2):423-64.
- Hamermesh, Daniel, Jeff Biddle. 1994. “Beauty and the Labor Market,” *American Economic Review* 84:1174-1194.
- Healy, Andrew, Neil Malhotra 2009. “Myopic Voters and Natural Disaster Policy,” *American Political Science Review*, 103(3), 387-406.
- Healy, Andrew, Neil Malhotra. 2010. “Random Events, Economic Losses, and Retrospective Voting: Implications for Democratic Competence,” *Quarterly Journal of Political Science* 5:193-208.
- Healy, Andrew, Neil Malhotra. 2013. “Retrospective Voting Reconsidered,” *Annual Review of Political Science* 16:285-306.
- Henrich, Joseph, Steven Heine, Ara Norenzayan. 2010. “The weirdest people in the world?” *Behavioral and Brain Sciences* 33:61-135.
- Jenkins, Rob, David White, Xandra Van Montfort, and A. Mike Burton. 2011. “Variability in Photos of the Same Face” *Cognition* 121(3): 313–23.
- Kandeh, Jimmy. 1992. “Politicization of Ethnic Identities in Sierra Leone,” *African Studies Review* 35: 81-99.
- King, Amy, Andrew Leigh. 2009. “Beautiful Politicians,” *Kyklos* 62(4): 579-593.
- Kuklinski, James, Paul Quirk. 2000. “Reconsidering the rational public: Cognition, heuristics, and mass opinion” *Elements of reason: Cognition, choice, and the bounds of rationality* 153-82.

- Langlois, Judith H., Lisa Kalakanis, Adam J. Rubenstein, Andrea Larson, Monica Hallam, Monica Smoot. "Maxims or Myths of Beauty? A Meta-Analytic and Theoretical Review," *Psychological Bulletin* 126(3): 390–423.
- Lau, Richard, David Redlawsk. 2001. "Advantages and Disadvantages of Cognitive Heuristics in Political Decision Making," *American Journal of Political Science* 45: 951-971.
- Lawson, Chappell, Gabriel Lenz, Andy Baker, Michael Myers. 2010. "Looking Like A Winner: Candidate Appearance and Electoral Success in New Democracies," *World Politics* 62:561-93.
- Lenz, Gabriel. 2012. *Follow the Leader? How Voters Respond to Politicians' Policies and Performance*. Chicago: University of Chicago Press.
- Lenz, Gabriel, Chappell Lawson. 2011. "Looking the Part: Television Leads Less Informed Citizens to Vote Based on Candidates' Appearance," *American Journal of Political Science* 55.
- Lupia, Arthur. 1994. "Shortcuts versus Encyclopedias: Information and Voting Behavior in California Insurance Reform Elections," *American Political Science Review* 88(1):63-76.
- Lutz, Georg. 2010. "Electoral Success of Beauties and Beasts," *Swiss Political Science Review* 16:457-480.
- Maestriperi, Dario, Andrea Henry, Nora Nickels. 2017. "Explaining financial and prosocial biases in favor of attractive people: Interdisciplinary perspectives from economics, social psychology, and evolutionary psychology," *Behavioral and Brain Sciences* 40:E19.
- Mobius, Markus, Tanya Rosenblatt. 2006. "Why Beauty Matters," *American Economic Review* 96:222-35.
- Moehler, Devra, Jeffrey Conroy-Krutz. 2016. "Eyes on the ballot: priming effects and ethnic voting in the developing world," *Electoral Studies* 42:99-113.

- Open Science Collaboration. 2015. "Estimating the reproducibility of psychological science," *Science* 349: aac4716.
- Palmer, Carl L., Rolfe D. Peterson. 2016. "Halo Effects and the Attractiveness Premium in Perceptions of Political Expertise," *American Politics Research* 44(2): 353–82.
- Popkin, Samuel. 1991. *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. Chicago: University of Chicago Press.
- Reynolds, Andrew, Marco Steenbergen. 2006. "How the world votes: The political consequences of ballot design, innovation and manipulation," *Electoral Studies*, 25:570-598.
- Todorov, Alexander, Anesu Mandisodza, Amir Goren, Crystal Hall. 2005. "Inferences of Competence from Faces Predict Election Outcomes," *Science* 308:1623-1626.
- Todorov, Alexander, Christopher Olivola, Ron Dotsch and Peter Mende-Siedlecki. 2015. "Social Attributions from Faces: Determinants, Consequences, Accuracy, and Functional Significance," *Annual Review of Psychology* 66: 519-545.
- Todorov, Alexander, and Jenny M. Porter. 2014. "Misleading First Impressions: Different for Different Facial Images of the Same Person," *Psychological Science* 25(7): 1404–17.

Table 1: Photo inferences better predict public spending than professional qualifications

Dependent variable:	Coeffi-	<i>p</i> -value	Coeffi-	<i>p</i> -value	Coeffi-	<i>p</i> -value
Proportion of CFF	cient		cient		cient	
verified on development	(1)	(2)	(3)	(4)	(5)	(6)
Photo inferences	3.531**	[0.035]			5.436**	[0.011]
Elected office experience			0.514	[0.345]	0.054	[0.911]
Management experience			0.102	[0.795]	-0.029	[0.933]
Years of education			-0.113	[0.300]	-0.212**	[0.042]
Incumbency			-0.674	[0.305]	-0.315	[0.581]
Age			0.008	[0.700]	0.009	[0.583]
Female			0.299	[0.618]	0.022	[0.966]
<i>F</i> -statistic on photo inferences	4.98				7.97	
<i>p</i> -value	[0.04]				[0.01]	
<i>F</i> -statistic on observables			0.38		1.02	
<i>p</i> -value			[0.88]		[0.45]	
R ²	0.26		0.20		0.45	
Observations	27 MPs		27 MPs		27 MPs	

*Notes: This table shows that inferences about how much money an MP will spend on development projects predict actual spending more accurately than alternative heuristics that could be found on an MP's vitae. Coefficient estimates are from ordinary least squares regressions with *p*-values reported in brackets that correspond to the maximum of (unadjusted OLS, HC2 corrected) standard errors to accommodate the small number of races (Angrist and Pischke 2009), the larger OLS values are reported here. Significance levels indicated by * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.001$. The outcome variable is the proportion of the annual CFF allotment that was verified by the audit as spent on development projects for each MP. Photo inferences capture the share of study participants who thought the MP looked less corrupt than peers, averaged across multiple photo pairings. The other explanatory variables were collected in surveys of MPs. The number of MPs varies between the spending tests here (27) and visit tests in Table 2 (28) since one MP took office late (due to a dispute over the electoral result) and missed receiving the first annual CFF allotment. Regressions include an assignment indicator from the field experiment in Bidwell et al. (2020).*

Table 2: Neither photo inferences nor professional qualifications reliably predict constituency visits

Dependent variable:	Coeffi-	<i>p</i> -value	Coeffi-	<i>p</i> -value	Coeffi-	<i>p</i> -value
Actual number of MP visits to home constituency	cient		cient		cient	
	(1)	(2)	(3)	(4)	(5)	(6)
Photo inferences	-10.355**	[0.015]			-7.838	[0.122]
Elected office experience			-0.384	[0.665]	0.022	[0.980]
Management experience			0.271	[0.711]	0.366	[0.605]
Years of education			-0.316	[0.124]	-0.204	[0.324]
Incumbency			0.149	[0.890]	-0.175	[0.868]
Age			-0.070**	[0.049]	-0.068**	[0.047]
Female			0.048	[0.965]	0.519	[0.638]
<i>F</i> -statistic on photo inference	6.88				2.62	
<i>p</i> -value	0.01				0.12	
<i>F</i> -statistic on observables			1.65		1.06	
<i>p</i> -value			0.18		0.42	
R ²	0.26		0.42		0.49	
Observations	28 MPs		28 MPs		28 MPs	

*Notes: This table shows that neither photo-based inferences nor alternative heuristics that could be found on an MP's vitae reliably predict how many visits an MP made to his or her home constituency. Coefficient estimates are from ordinary least squares regressions with *p*-values reported in brackets that correspond to the maximum of (unadjusted OLS, HC2 corrected) standard errors to accommodate the small number of races (Angrist and Pischke 2009), the larger OLS values are reported here. Significance levels indicated by * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.001$. The outcome variable is the average number of visits reported across the multiple respondents surveyed in the MP's home constituency. Photo inferences capture the share of study participants who thought the MP made more visits than peers, averaged across multiple photo pairings. The other explanatory variables were collected in surveys of MPs. Regressions include an assignment indicator from the field experiment in Bidwell et al. (2020).*

Table 3. Photo inferences predict candidates' oral communication skills

Dependent variable: Participant chose first sector in pair, audio condition	Coefficient (1)	p-value (2)	Coefficient (3)	p-value (4)
Mean proportion chose first sector in photos	0.584***	[<0.001]	0.401***	[<0.001]
First sector higher in own preferences			0.267***	[<0.001]
First sector endorsed by the eventual winner			0.023	[0.394]

Observations

1,413 evaluations, audio condition

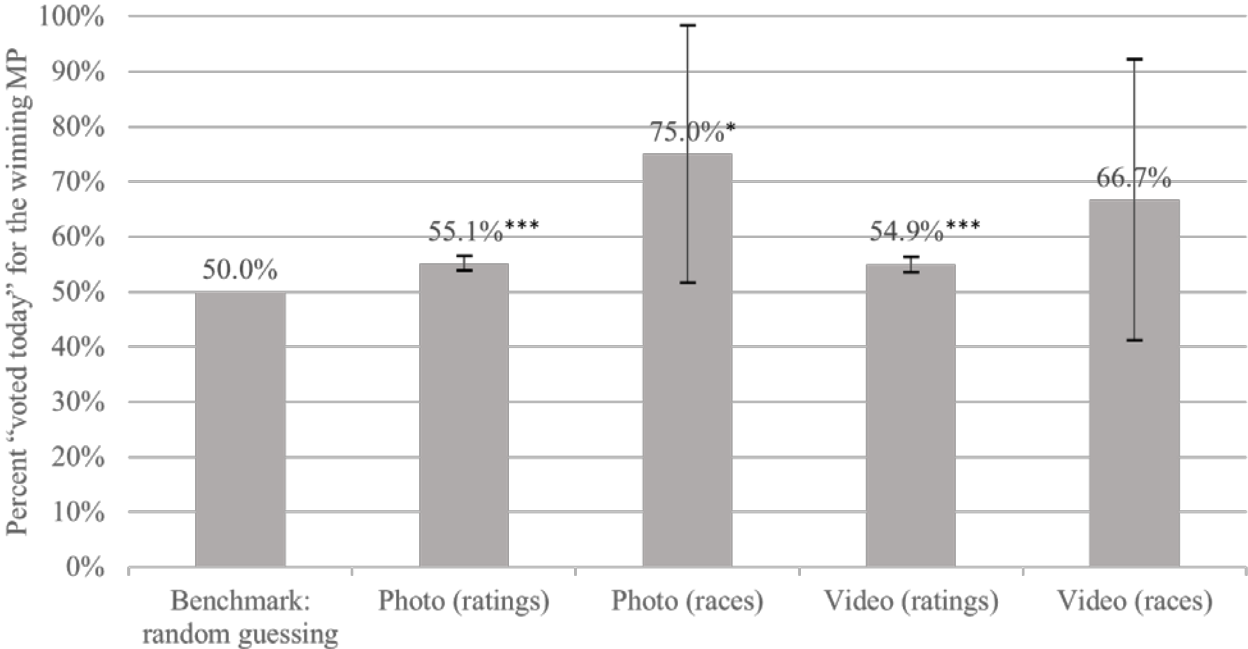
*Note: This table shows that photo inferences predict oral communication skills. Estimates are from ordinary least squares regressions (corresponding estimates from logistic regressions are comparable in sign and significance) and the p-values reported in brackets are associated with robust standard errors clustered at the level of the 72 unique MP candidate pairings. Significance levels indicated by * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.001$. The dependent variable is whether a participant picked the first sector in a pair after listening to candidate debates about why their sectors should be prioritized for greater government funding. The main explanatory variable is the share of other study participants who picked the first sector based on candidate photo pairs. The next two explanatory variables control for whether the first sector in a pair ranked higher in the participant's own sectoral preferences and whether it was endorsed by the eventual winner.*

Table 4: Photo inferences about candidate traits correlate with electoral returns

Dependent variable:	Candidate's actual vote share				
	(1)	(2)	(3)	(4)	(5)
Inferred spending performance (winner-winner photo pairs)	0.590** [0.022]		0.744** [0.017]		
Inferred visit performance (winner-winner photo pairs)		-0.442 [0.170]	-0.647* [0.073]		
Inferred persuasiveness (winner-loser photo pairs)				0.643*** [0.002]	0.501** [0.013]
Elected office experience			0.008 [0.910]		0.095 [0.700]
Management experience			-0.047 [0.353]		0.214** [0.031]
Years of education			-0.016 [0.310]		0.042** [0.048]
Incumbency			-0.100 [0.241]		0.216 [0.439]
Age			-0.000 [0.941]		-0.003 [0.516]
Female			0.015 [0.847]		0.024 [0.858]
Observations		27 winning MPs		34 candidates	

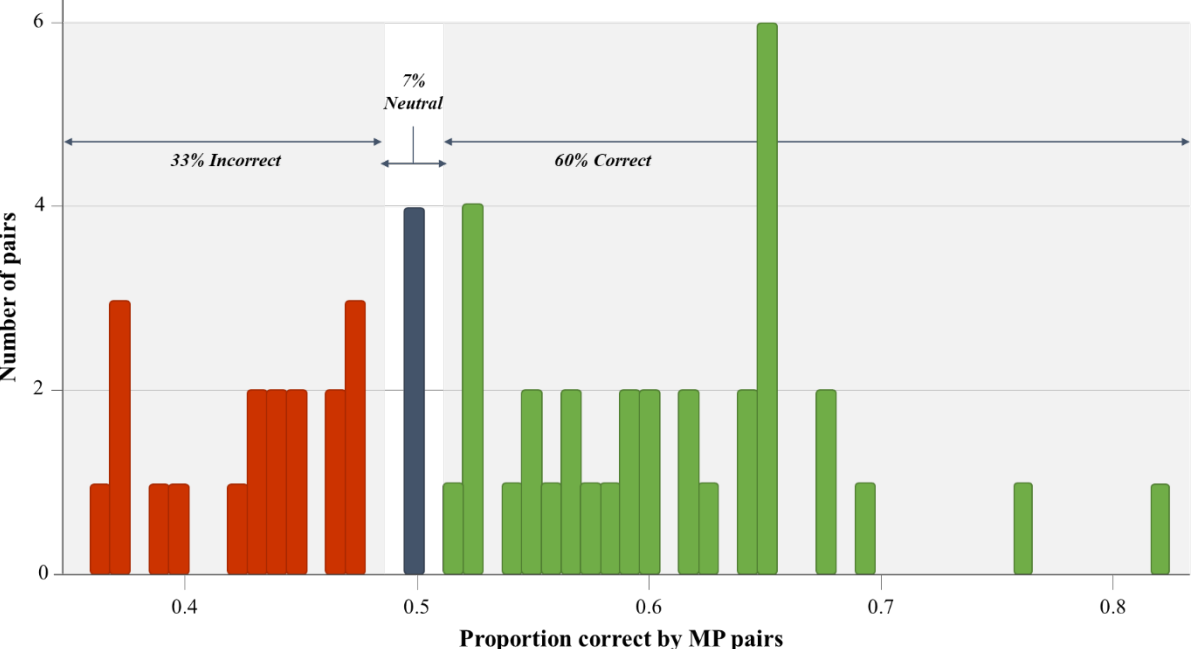
*Notes: This table shows that photo-based inferences about which candidates are less corrupt and more persuasive predict higher vote shares in the 2012 electoral returns. Coefficient estimates are from ordinary least squares regressions with p-values reported in brackets that correspond to the maximum of (unadjusted OLS, HC2 corrected) standard errors to accommodate the small number of races (Angrist and Pischke 2009). Significance levels indicated by * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.001$. The outcome variable is the vote share the politician received in the 2012 Parliamentary elections. The inferred performance measures capture the share of study participants who thought the MP performed better than peers on spending the CFF on development and making constituency visits, respectively, averaged across multiple photo pairings among elected MPs. Inferred persuasiveness captures the share of study participants who chose to prioritize the candidate's sectoral project for greater government funding, averaged across multiple photo pairings between winning and losing candidates. The other explanatory variables were collected in surveys of MPs. Regressions include an assignment indicator from the field experiment in Bidwell et al. (2020).*

Figure 1: Pre-election snap judgments predict subsequent winners of Parliamentary races



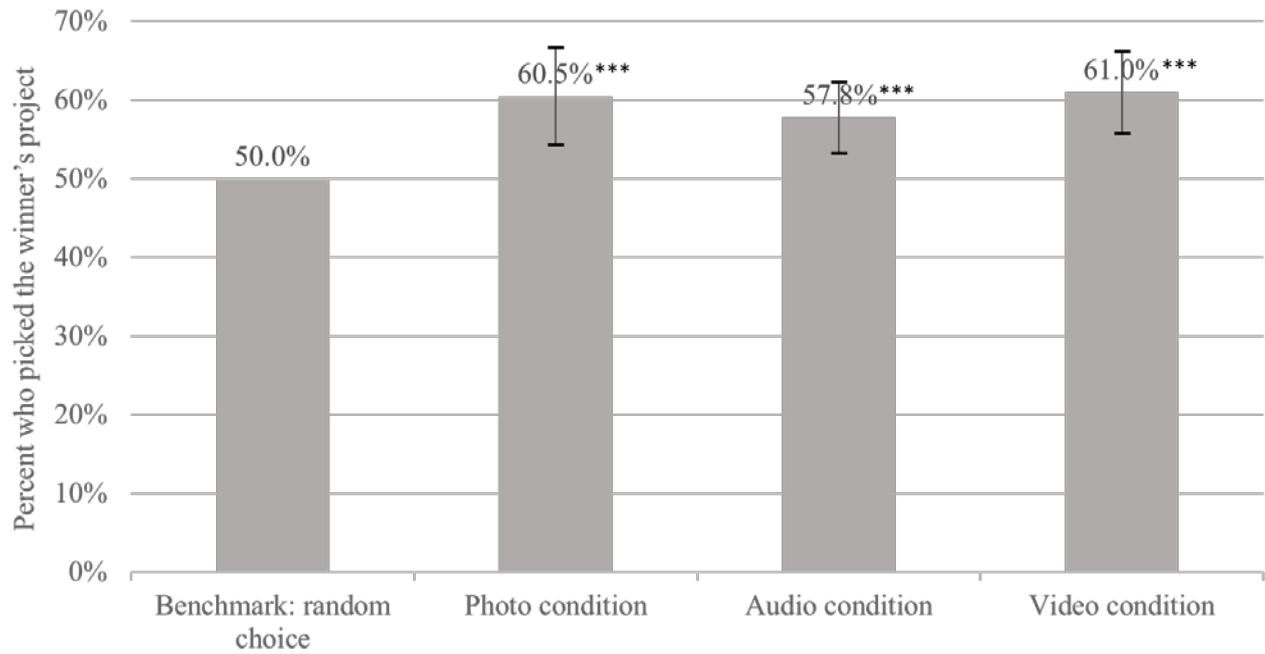
Notes: this figure shows that study choices over which candidate a participant would “vote today” for based on photo and video pairs predict which candidate later won the Parliamentary race. The bars denote the percent of participants who chose the eventual winner, accompanied by 90% confidence intervals, at both the level of participant ratings (N=3,581) and aggregated up to races (N=12). Stars denote the significance level (p < 0.10, ** p < 0.05 and *** p < 0.001) at which a two-sided t-test rejects that the percentage is statistically equivalent to chance (i.e. the benchmark of 50.0% which would be the percent expected under random guessing).*

Figure 2: Distribution of correct guesses about which MP in photo pairs actually spent more public money on development



Notes: This figure shows that photo-based inferences correctly identify the true better public spending performer in a majority (60%) of the politician photo pairings used in the post-election Study 2. Recall that each of the 55 MP photo pairs in the CFF spending task was evaluated by approximately 40 different study participants. This histogram orders the photo pairs by the proportion of these participants who correctly guessed which MP in the pair actually spent more money on development. Arrows demarcate the percentage of pairs that were on average correctly assessed by the study (in green, where the share of participants who chose the true better performer is greater than 0.5), incorrectly assessed (in red, for correct shares below 0.5) and neutral (in dark gray, for study shares equal to 0.5 which would be the share expected for random guessing).

Figure 3: Winners are more persuasive than losing candidates under all study conditions



Notes: this figure shows that when asked to choose one of two projects to receive greater government funding, study participants were more likely to pick the project advocated for by candidates who eventually won the election compared to candidates who eventually lost. This implies that the eventual winners were more persuasive than losers based solely on their looks (in the photo condition), their oral communication skills (in the audio condition), and the combination of the two (in the video condition). The bars denote the percentage of study participants who chose the project endorsed by the eventual winner under each condition, accompanied by 95% confidence intervals (N=72 candidate pairs for all conditions). Stars denote the significance level ($p < 0.10$, ** $p < 0.05$ and *** $p < 0.001$) at which a two-sided t-test rejects that the percentage is statistically equivalent to chance (i.e. the benchmark of 50.0% which would be expected if study participants chose a project at random).*

ONLINE APPENDIX

For “Snap Judgments: Predicting Politician Competence from Photos”

Contents:

- Supplemental Information
 - Study 1 Participants and protocols
 - Study 2 Double blind protections
- Supplemental Figures
 - Figure A1: Examples of study media
 - Figure A2: Spending scatterplot
 - Figure A3: Persuasion scatterplot
 - Figure A4: Spending and inferred wealth scatterplot
- Supplemental Tables
 - Table A1: Regression analysis of electoral forecasting
 - Table A2: Correlation matrix of vote shares and inferences
 - Table A3: Map of study 2 task apportionment
 - Table A4: Two-way comparisons for inferred spending and observables
 - Table A5: Two-way comparisons for inferred effort and observables
 - Table A6: Persuasion robustness check
 - Table A7: Corrections for multiple inference across studies and tasks
 - Table A8: Photo inferences and vote margins
 - Table A9: Variation in evaluations within versus across MPs
 - Table A10: Arbitrary image differences do not predict voter inferences
 - Table A11: Robustness check on mixed gender pairs

SUPPLEMENTAL INFORMATION

Study 1 Participants and protocols

The first pre-election study recruited 407 participants, sampled from an earlier household listing of registered voters, living near 40 polling centers that are located in 8 of the 12 constituencies where debates occurred. Overall, 54% of participants are female, with an average age of 41, and 70% have no formal schooling. Analysis excludes participant evaluations of candidates from their own constituency and any they recognized.

Regarding protocols, for the photo condition participants first evaluated each candidate on leadership, corruption and attractiveness, and guessed their ethnicity. The corruption prompt explains that: “every year, Honorables receive 44 million Leones [USD 11,000] in the constituency facilitation fund. How much do you think this candidate would put in his own pocket and not use for development or for trips to his constituency?” (The second performance study uses audit data on how elected MPs actually spent this money.) Participants were then shown the two photos together and asked “if you had to vote today, which candidate would you choose?”

Similarly, for the video condition, participants watched 10 to 20 second clips of candidates speaking about which sector they would prioritize for additional government spending. Participants rated each candidate on leadership, corruption, likeability and articulateness; and then compared the two and made “vote today” selections. Participants then guessed party affiliation.

Study 2 Double blind protections

Enumerators who administered the study to participants were blinded to the three performance outcomes of interest for the MPs included in the photo pairs. This protection was ensured through several steps. First, note that the outcome for persuasion—namely whether participant photo selections predict the audio only selections of other study participants—was not knowable until after the study field work was complete and the data analyzed. Second, enumerators did not have access to identifiable performance data for the remaining two outcomes, namely constituency visits and CFF spending, which were collected via field audits. This data is publicly available only in fully anonymized form. Third, the enumerators were not involved in designing the study protocols and specifically had no role in forming the high and low performing MP photo pairs. Fourth, only one enumerator of 12 total who administered this study also worked on the field audit team.

Estimates for CFF spending and visits are robust to excluding participant responses who were administered tasks for this study by this particular enumerator (which affects 9.75% of study participants).

Specifically, the estimate reported in the main text for the accuracy of photo inferences predicting CFF spending using responses from all study participants is 54.2 percent (p -value <0.001 , $N = 2,158$). The estimate dropping all responses collected by this enumerator is 54.6 percent (p -value <0.001 , $N = 1,942$). Similarly, the estimate reported in the main text for the accuracy of photo inferences predicting constituency visits using responses from all study participants is 45.7 percent (p -value <0.001 , $N = 2,158$). The estimate dropping all responses collected by this enumerator is 45.5 percent (p -value <0.001 , $N = 1,972$).

SUPPLEMENTAL FIGURES

Figure A1: Examples of study media

Panel A: Example of photo still pair of rival candidates taken from pre-election debates



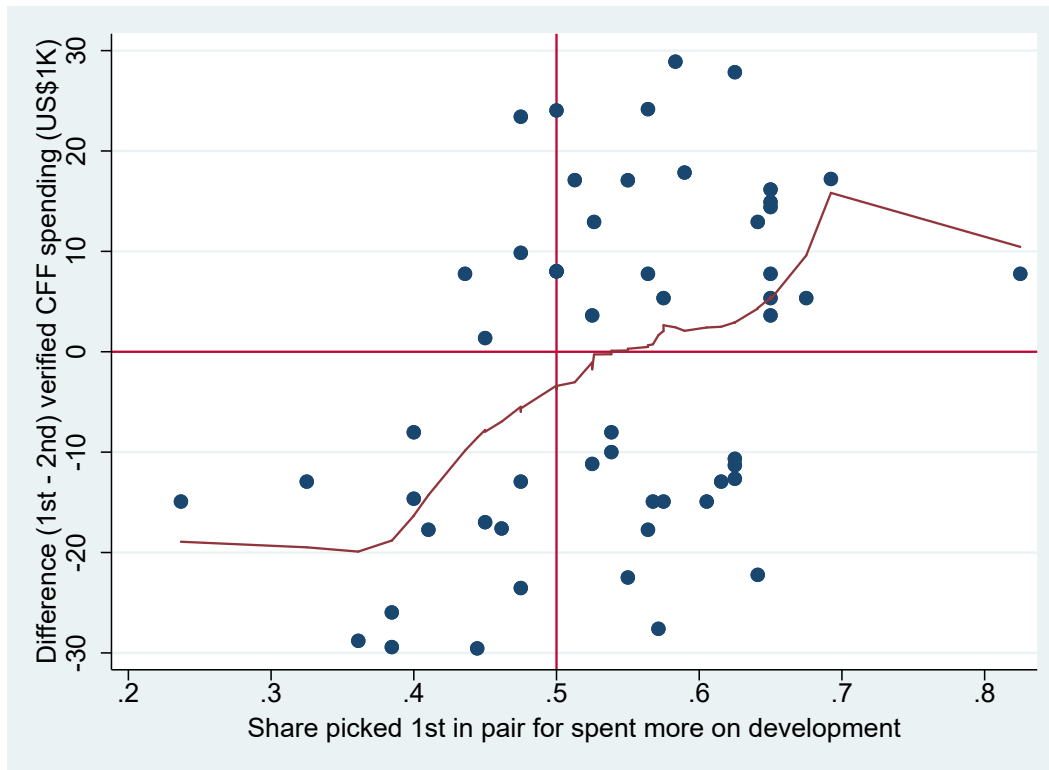
If you had to vote today, which candidate would you choose?

Panel B: Example of headshot pair of elected MPs from the official website of Parliament



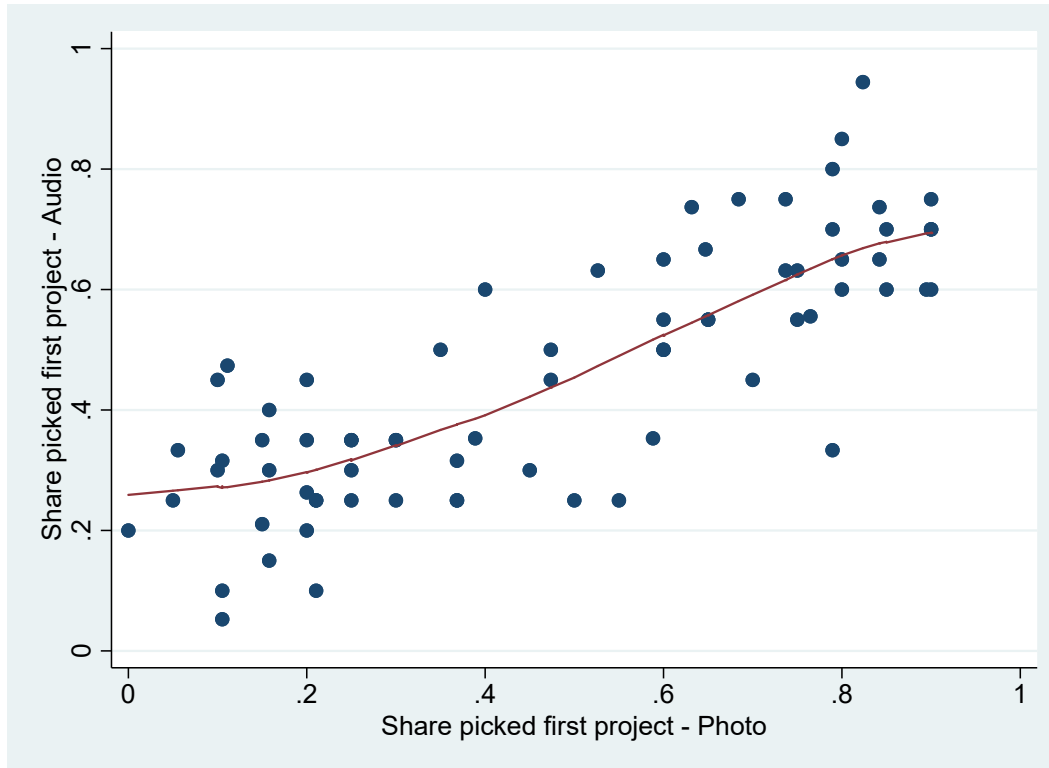
Do your best to guess which MP spent more of the CFF on development and put less in his own pocket

Figure A2: Spending scatterplot that shows how study participant guesses of which politician in a photo pair spent more on development predict actual CFF spending



Notes: this figure is a two-way scatterplot version of the relationship between looking less corrupt in photos and spending more public funds on development projects as presented in main text Figure 2. In this graph, each dot represents a politician pair, each of whose public spending under the constituency facilitation fund (CFF) was verified via field audits (N=55 pairs). The horizontal axis orders these pairs by the proportion of study participants who guessed that the first MP in the photo pair spent more on development. The vertical axis shows the actual difference in development spending of the first minus the second MP in the pair as measured in the audits. The upward sloping (lowess fitted) line shows that photo assessments strongly predict differences in actual spending. To see this, note that a majority of observations (33 of 55 pairs or 60%) are correct on average: these are located in the bottom left quadrant (where the first MP spent less and fewer than 50% of study participants thought the first MP spent more) and upper right quadrant (where the first MP spent more and more than 50% of study participants thought the first MP spent more). The remaining minority of observations are incorrect on average (18 pairs or 33% located in the remaining two quadrants) or neutral (4 pairs or 7% located on the vertical 0.50 share line). In the corresponding bivariate regression, the coefficient on pick first photo share is a highly significant 63.69 (robust standard error 14.98), which implies that moving from a neutral (0.50) photo share to the 75th percentile of the photo share distribution (or 0.625) corresponds to an average difference of 7.96, or US\$7,961, in verified development spending between the first and second MP in the pair.

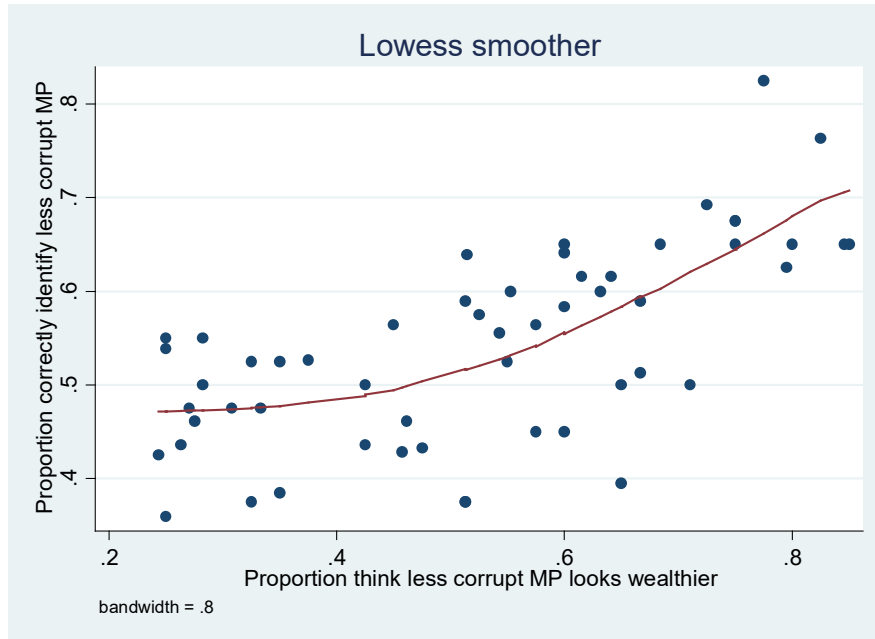
Figure A3: Persuasion scatterplot that shows how looking more persuasive in photos predicts better oral communication skills for pairs of politicians



Notes: this figure is a two-way scatterplot version of the relationship between looking persuasive in photos and having better oral communication skills presented in main text Table 3. In this graph, each dot represents a politician pair, each of whom argued for a specific sector to receive greater government funding ($N=72$ pairs). The horizontal axis orders these pairs by the proportion of study participants who selected the sector backed by the first politician in the pair based on seeing their photos. The vertical axis shows the corresponding share of other study participants who picked the sector backed by the first politician in the pair based on hearing their audio arguments. The upward sloping (lowess fitted) line shows that the photo assessments strongly positively predict the audio assessments. To provide a sense of magnitude, the estimated coefficient on photo share in the corresponding bivariate regression is a highly significant 0.58 (robust standard error 0.05).¹ This implies that a percentage point increase (+0.01) in the photo share is associated on average with roughly a half point increase (+0.0058) in the audio share.

¹ Note that this bivariate regression operates at the pair level ($N=72$) while that in Table 3 operates at the level of study participant evaluation ($N=1,413$) with standard errors clustered by pair. As expected, results are nearly identical.

Figure A4: Photo inferences about trustworthiness and wealth, compared to actual CFF spending performance



Notes: this figure fits a nonparametric relationship between the proportion of study participants who correctly identified the MP who spent more on development (labeled less corrupt) and the proportion of (other) participants who thought that same MP looked like the wealthier one in the pair. In this graph, each dot represents a politician pair, each of whose public spending under the constituency facilitation fund (CFF) was verified via field audits (N=55 pairs). The horizontal axis orders these pairs by the proportion of study participants who thought that the MP who was in fact less corrupt (e.g. had higher development spending verified in the audits) looked like the wealthier MP in the pair. The vertical axis shows the corresponding share of other study participants who correctly guessed which MP in the pair spent more of the CFF money on development.

The upward sloping line in Figure A5 shows a strong positive correlation between the proportion of study participants who correctly identified the less corrupt MP and the proportion of others who thought that the less corrupt MP in the photo pair looked wealthier. This suggests that the rate of correct inference is highest where the less corrupt MP appeared wealthier. In fact, inferences in the bottom third of the relative wealth appearance pairs, where the less corrupt MP looked poorer, are marginally worse than random guessing (47.3% correct, p-value = 0.075, N=18 pairs); while those in the top third of the distribution, where the less corrupt MP looked wealthier, are substantially more accurate than random guessing (62.3% correct, p-value = 0.001, N=18 pairs). Inference accuracy in the middle third, where the two MPs appeared comparable in wealth terms, is greater than 50% but not significantly so (52.3%, p-value 0.318, N=17 pairs).

As one final point about potential mechanisms behind spending inferences, the U.S. labs experimented with how content delivery affects the accuracy of voter forecasts. For example, Ballew and Todorov (2007) show that voter inferences about competence are as accurate in predicting gubernatorial winners based on only 100 milliseconds of exposure as under unlimited time exposure to candidate photos. Benjamin and Shapiro (2009) varied whether lab participants saw gubernatorial debate clips with or without the sound on. They find that turning the sound on increased the accuracy of inferences about party identity but decreased the accuracy of predictions about who won. In this vein, the Sierra Leone study compared the discriminatory power of photo inferences to those formed in response to audio and video snippets of MPs describing their plans for the CFF (recorded in the pre-election debates). Results suggest that if anything, the richer information conditions decrease the accuracy of voter ability to select which MP in fact spent more money on development. The percent correct guesses for both the audio and video condition are indistinguishable from random guessing at conventional levels: 47.0% correct (p -value = 0.054) for audio and 50.1% correct (p -value = 0.926) for video. Yet note that since only 13 of the elected MPs in the CFF audit sample participated in these debates, the audio and video tasks cover only 27 distinct pairs of high and low performers, so have less statistical power than the photo condition.

SUPPLEMENTAL TABLES

Table A1: Regression estimates of snap judgment forecast accuracy based on photos

Dependent variable:	APC vote share (1)	Difference APC- SLPP vote shares (2)
Share of study participants who would vote today for APC	1.021** [0.038]	
Mean difference in APC-SLPP leadership assessments		-1.569 [0.572]
Mean difference in APC-SLPP (not) corrupt assessments		5.222* [0.074]
Mean difference in APC-SLPP attractiveness assessments		-1.609* [0.052]
R ²	0.36	0.63
Observations	12 Parliamentary races	

*Notes: Coefficient estimates are from ordinary least squares regressions with p-values reported in brackets that correspond to the maximum of (unadjusted OLS, HC2 corrected) standard errors to accommodate the small number of races (Angrist and Pischke 2009), the larger OLS values are reported here. Significance levels indicated by * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.001$. The dependent variable is the actual share of the two party vote going to the All People’s Congress (APC) candidate in the 2012 Parliamentary Elections in column 1 and the difference in vote shares between the two major parties in column 2 (the other major party is the Sierra Leone People’s Party, SLPP).*

Table A1 complements Figure 1 by using regression analysis to link study inferences to electoral returns at the race-level. Column 1 shows that “vote today” study choices in favor of the All People’s Congress (APC) party candidate predict the subsequent share of the two-party vote that the APC candidate received in the actual election. The R² indicates that these inferences alone explain 36% of the total variation in vote shares. Column 2 explores mechanisms, with one important caveat: the electoral study was designed to establish whether the main forecasting result holds with individual voters in Sierra Leone (so the number of voter evaluations is large, at 3,581), and not to parse the mechanisms behind the result, which requires a large number of races (more

than the 12 covered here). Thus, statistical power to parse race-level attributes is limited. With that said, assessments of which candidates voters thought would spend more of the CFF on development and keep less for themselves (i.e. looked less corrupt) mildly positively correlate with vote shares. In contrast to several existing studies (Banducci et al. 2008; Berggren et al. 2010, 2017; King and Leigh 2009; Lutz 2010; Rosar et al. 2008), column 2 shows that physical attractiveness, if anything, carries a mild electoral penalty in this context. To express this in a real-world shorthand, note that the 12 candidates judged more attractive in the photo pairs won fewer than half of the races studied.

Note that column 1 of Table A1 follows the approach of Benjamin and Shapiro (2009) by testing whether “vote today” choices in the study predict vote choices in the real world. Column 2 follows Todorov et al. (2005) by testing whether inferences about candidate traits in the study predict vote choices in the real world. Table A2 below explores how these two approaches relate to one another, by presenting the correlation matrix for the vote choice outcome of interest and all four potential explanatory variables. Note first that differences in “vote today” study shares have the largest correlation with differences in actual vote shares (0.59), followed closely by inferences about which candidate appears less corrupt (0.56). These two explanatory measures are also highly correlated with each other (0.98). By contrast, differences in leadership inferences and attractiveness assessments are somewhat less strongly correlated with actual vote shares (0.50, 0.28 respectively) and also less tightly correlated with “vote today” shares (0.89 and 0.87 respectively). While all four potential explanatory variables are highly positively correlated with each other, this constellation of relationships is consistent with the idea that inferences about corruption are a key component of the predictive face-vote relationship in this context.

Table A2: Correlation matrix for actual vote shares and study inferences used in Table A1

Difference (APC- SLPP) in candidates’:					
	Actual vote shares	“Vote today” shares	Corruption inferences	Leadership inferences	Attractiveness ratings
Actual vote shares	1.00				
“Vote today” shares	0.59	1.00			
Corruption inferences	0.56	0.93	1.00		
Leadership inferences	0.50	0.89	0.98	1.00	
Attractiveness ratings	0.28	0.87	0.91	0.91	1.00

Table A3: Map of Study 2 task apportionment

V	N	Section 0	Section 1	Section 2	Section 3	Section 4
1	20	Candidate recognition	1A- 14/28 generic photo evaluations	3C- 13/27 spending video pairs	4A- 12/72 persuasion photo pairs	Party affiliation
2	19		1A- 14/28 generic photo evaluations	3C- 13/27 spending video pairs	4A- 12/72 persuasion photo pairs	
3	20		1A- 14/28 generic photo evaluations	3D- 14/27 spending audio pairs	4A- 12/72 persuasion photo pairs	
4	20		1A- 14/28 generic photo evaluations	3D- 14/27 spending audio pairs	4A- 12/72 persuasion photo pairs	
5	20		1B- 18/36 generic video evaluations	1C- generic voter preferences		
6	20		1B- 18/36 generic video evaluations	1C- generic voter preferences		
7	20		2- 28/56 effort photo pairs	3D- 13/27 spending audio pairs	4A- 12/72 persuasion photo pairs	
8	20		2- 28/56 effort photo pairs	3D- 13/27 spending audio pairs	4A- 12/72 persuasion photo pairs	
9	20		2- 28/56 effort photo pairs	4B- 12/72 persuasion audio pairs		
10	20		2- 28/56 effort photo pairs	4B- 12/72 persuasion audio pairs		
11	20		3A- 28/56 spending photo pairs	4B- 12/72 persuasion audio pairs		
12	20		3A- 28/56 spending photo pairs	4B- 12/72 persuasion audio pairs		
13	20		3A- 28/56 spending photo pairs	4C- 12/72 persuasion video pairs		
14	20		3A- 28/56 spending photo pairs	4C- 12/72 persuasion video pairs		
15	20		3B- 28/56 affluence photo pairs	4C- 12/72 persuasion video pairs		
16	20		3B- 28/56 affluence photo pairs	4C- 12/72 persuasion video pairs		
17	20		3B- 28/56 affluence photo pairs	4C- 12/72 persuasion video pairs		
18	20		3B- 28/56 affluence photo pairs	4C- 12/72 persuasion video pairs		
19	20		3C- 14/27 spending video pairs	4B- 12/72 persuasion audio pairs		
20	20		3C- 14/27 spending video pairs	4B- 12/72 persuasion audio pairs		

Notes: this table charts the assignment of tasks across the 399 participants in the performance study. Each row represents one of the 20 distinct versions of the study that were administered to 20 unique participants per version during home visits. Tasks were allocated to maintain a reasonable workload for participants and to reduce the interdependence among statistical tests. The 20 versions were randomly assigned across the 20 participants recruited from each unique voter registration center, for maximum geographic spread for all tasks. The order in which politician pairs were presented to participants within any given task was randomized across survey versions to mitigate order effects.

Table A4: Two-way comparisons of inferred trust from photos versus observable politician characteristics in predicting actual MP development spending

First explanatory variable	Coefficient (std. error)	Second explanatory variable	Coefficient (std. error)	Difference [<i>p</i> -value]
Trust inference	3.516** (1.596)	Elected office experience	0.216 (0.282)	3.300* [0.054]
Trust inference	3.542** (1.615)	Management experience	0.074 (0.280)	3.468** [0.044]
Trust inference	5.487*** (1.608)	Years of Education	-0.186** (0.072)	5.674*** [0.002]
Trust inference	3.773** (1.639)	Incumbency	0.211 (0.307)	3.563** [0.036]
Trust inference	3.609** (1.601)	Age	0.011 (0.014)	3.598** [0.034]
Trust inference	3.736*** (1.662)	Female	-0.193 (0.393)	3.929** [0.040]
Observations	27 MPs			

*Notes: Each row corresponds to a separate ordinary least squares regression where the dependent variable is actual CFF spending on development projects, which is regressed on the first and second explanatory variables listed. The final column presents the difference in the two estimated coefficients and the associated *p*-value rejecting the null hypothesis of zero difference. Significance levels * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are the maximum of (unadjusted OLS, HC2 corrected) to accommodate the small number of races, following Angrist and Pischke (2009), the larger OLS values are reported here. These MPs participated in a field experiment reported in Bidwell et al. 2020, so all photo pairings are formed within that treatment assignment and regressions include an assignment indicator.*

Table A5: Two-way comparisons of inferred effort from photos versus observable politician characteristics in predicting effort convening public meetings

First explanatory variable	Coefficient (std. error)	Second explanatory variable	Coefficient (std. error)	Difference [<i>p</i> -value]
Effort inference	-10.269** (3.987)	Elected office experience	-0.413 (0.567)	-9.856** [0.023]
Effort inference	-10.050** (4.019)	Management experience	0.387 (0.582)	-10.437** [0.015]
Effort inference	-9.255* (4.562)	Years of Education	-0.086 (0.171)	-9.168* [0.060]
Effort inference	-10.880** (4.065)	Incumbency	-0.434 (0.638)	-10.446** [0.015]
Effort inference	-9.632** (3.625)	Age	-0.062** (0.026)	-9.570** [0.014]
Effort inference	-9.568** (4.299)	Female	-0.431 (0.850)	-9.137* [0.062]
Observations	28 MPs			

*Notes: Each row corresponds to a separate ordinary least squares regression where the dependent variable is the actual number of meetings the MP convened with constituents, which is regressed on the first and second explanatory variables listed. The final column presents the difference in the two estimated coefficients and the associated *p*-value rejecting the null hypothesis of zero difference. Significance levels * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are the maximum of (unadjusted OLS, HC2 corrected) to accommodate the small number of races, following Angrist and Pischke (2009), the larger OLS values are reported here. These MPs participated in a field experiment reported in Bidwell et al. 2020, so all pairings are formed within that treatment assignment and regressions include an assignment indicator.*

Table A6: Winners' persuasion advantage controlling for party and sector preferences

Dependent variable:	Photo condition		Audio condition		Video condition	
Participant picked first project in pair	(1)	(2)	(3)	(4)	(5)	(6)
First project endorsed by winning candidate	0.215***	0.120**	0.161***	0.090**	0.221***	0.136***
	[0.001]	[0.014]	[0.000]	[0.013]	[0.000]	[0.001]
First project endorsed by own party's candidate		-0.001		0.03		0.037
		[0.985]		[0.340]		[0.192]
First project ranks higher in own preferences		0.451***		0.326***		0.402***
		[0.000]		[0.000]		[0.000]
Constant term	0.362***	0.199***	0.372***	0.248***	0.367***	0.192***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Observations	1,397		1,413		1,422	

*Notes: this table shows that the results discussed in the main text that winning candidates argue more persuasively than losers in promoting a sector for greater government funding are robust to controlling for party identity and sectoral preferences, under all three study conditions (photo, oral, video). Coefficient estimates are from ordinary least squares regressions with p-values reported in brackets that correspond to robust standard errors clustered by 72 photo pairs (note that corresponding estimates from logistic regressions are comparable in sign and significance, results not shown). Significance levels indicated by * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.001$. The outcome variable is whether a participant picked the first sector in a pair after being exposed to study stimuli, which is photo pairs in columns 1 and 2, listening to pairs of oral arguments about why a candidate's sector should be prioritized for greater government funding in columns 3 and 4, and watching a video of the same arguments in columns 5 and 6. The first (and main) explanatory variable is whether the first project was endorsed by the eventual winner. The other two explanatory variables control for whether the first sector in a pair was endorsed by the candidate from the participant's own preferred party, and whether it ranked higher in the participant's own ranking of sectoral preferences.*

Table A7: Corrections for multiple inference across studies and tasks

	Voter predictions				Race-level forecasts			
	<i>p</i> -value	<i>N</i>	FDR <i>q</i> -val	FWER <i>p</i> -val	<i>p</i> -value	<i>N</i>	FDR <i>q</i> -val	FWER <i>p</i> -val
Outcome – condition	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Winner – photo	0.0001	3,581	0.001	0.001	0.0819	12	0.058	0.737
Winner – video	0.0001	3,581	0.001	0.001	0.2657	12	0.141	>0.99
Visits – photo	0.0001	2,158	0.001	0.001	0.0005	56	0.003	0.005
Spending – photo	0.0001	2,158	0.001	0.001	0.0042	55	0.005	0.038
Spending – audio	0.0536	1,032	0.014	0.482	0.4472	27	0.202	>0.99
Spending – video	0.9255	1,029	0.115	>0.99	0.9575	27	0.469	>0.99
Persuasion – photo	0.0001	1,397	0.001	0.001	0.0011	72	0.003	0.010
Persuasion – audio	0.0001	1,413	0.001	0.001	0.0011	72	0.003	0.010
Persuasion – video	0.0001	1,422	0.001	0.001	0.0001	72	0.001	0.001

*Note: This table adjusts standard errors to account for multiple inference across different study tasks, pooling Studys 1 and 2 together. Columns 1 through 4 are for voter-level predictions, and columns 5 through 8 are estimates aggregated up to the race-level. Columns 1 and 5 display the unadjusted or naïve *p*-value; columns 2 and 6 the number of individual or race-level observations; columns 3 and 7 display false discovery rate (FDR) adjustments, and columns 4 and 8 display conservative familywise error rate (FWER) adjustments. Note that FDR adjusted *q*-values can at times be smaller than (large) naïve *p*-values when there are many true rejections in the test set, as is the case here.*

Additional information on multiple inference adjustments

Multiple inference adjustments inflate standard errors to account for interdependencies between tests conducted on the same dataset. For Study 1, this is a typical application, since analysis tests the accuracy of forecasts under two conditions on the same participant dataset (all participants completed both tasks). Study 2 is more nuanced, as it uses 20 different versions of the study to both apportion the work across participants and to create independence between tests (see Table

A3). For example, one version asks 20 participants to evaluate 28 of the 56 total MP *photo* pairs about *spending* performance, and then to evaluate the *persuasiveness* of 12 of the 72 *video* pairs of winning and losing candidates. A “sister” version asks 20 (other) participants to rate 14 of 27 total MP *video* pairs about *spending*, and then rate the *persuasiveness* of 12 of the 72 candidate *audio* pairs. This means that the associated corrections are likely more conservative than needed.

Note that there is no overlap in the participant pool across studies: e.g. none of the 407 participants in Study 1 are among the 399 in Study 2, which was conducted in a wholly separate set of communities. There is, however, partial overlap in the politician pool across studies and also across tasks within: e.g. the 12 winning candidates from the electoral forecasting study appear in the CFF audit dataset used in the performance study, and again in the candidate persuasion videos. Table A6 thus pursues a prudent approach and adjusts across all 9 outcome predictions collected in both studies taken together. Column 3 implements a false discovery rate (FDR) correction (following Anderson 2008 and Benjamini et al. 2006), which controls the proportion of rejections that are type I errors. Column 4 presents a Bonferroni correction, which controls the family-wise error rate (FWER) or risk of any single type I error. While coarse and conservative, the Bonferroni approach has the advantage of being straightforward to implement across different datasets and non-overlapping subsamples of study participants. I adjust separately across voter-level predictions (in columns 1 to 4) and race-level predictions (in columns 5 to 8).

Results are reassuring. For the central question of what voters might be able to infer from ballot photos of candidates, the four relevant photo predictions are robust to false discovery rate (FDR) adjustments at both the individual voter- and race-level (in columns 3 and 7), and 7 of these 8 total estimates are robust to conservative family-wise error rate (FWER) adjustments (in columns 4 and 8). The one rejection is for forecasting winners from photos at the race-level under FWER (first row, column 8). Broadening consideration to the auxiliary conditions deployed to explore mechanisms, only one that was at least marginally significant on a naïve basis (in column 1 or 5) loses statistical significance under multiple inference adjustment: this is for spending under the audio condition (row 5, column 4), which is discussed in the appendix with reference to Figure A5. For others that are not significant under adjustments, like picking the winner under the video condition at the race-level (in row 2, column 8), we already knew they were weak on a naïve basis (from column 5).

Table A8: Photo inferences about candidate traits correlate with vote margins

Dependent variable:	Winner's 1 st versus 2 nd place vote margin		
	(1)	(2)	(3)
Inferred trustworthiness	1.012**		1.357**
(winner-winner photo pairs)	[0.019]		[0.014]
Inferred effort		-0.457	-0.630
(winner-winner photo pairs)		[0.404]	[0.404]
Elected office experience			-0.056
			[0.668]
Management experience			-0.074
			[0.404]
Years of education			-0.032
			[0.249]
Incumbency			-0.128
			[0.393]
Age			0.002
			[0.650]
Female			-0.024
			[0.864]
Observations	27 winning MPs		

*Notes: This table shows that photo-based inferences about which candidates are less corrupt and more persuasive predict higher winning margins in the 2012 electoral returns. Coefficient estimates are from ordinary least squares regressions with p-values reported in brackets that correspond to the maximum of (unadjusted OLS, HC2 corrected) standard errors to accommodate the small number of races (Angrist and Pischke 2009), the larger OLS values are reported here. Significance levels indicated by * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.001$. The outcome variable is the first versus second place finisher vote margin in the 2012 Parliamentary elections. The inferred performance measures capture the share of study participants who thought the MP performed better than peers on spending the CFF on development and making constituency visits, respectively, averaged across multiple photo pairings among elected MPs. The other explanatory variables were collected in surveys of MPs. Regressions include an assignment indicator from Bidwell et al.'s (2020) field experiment.*

Table A9: Variation in participant evaluations within versus across MPs

Dependent Variable:	Corruption	Leadership	Attractiveness
	(1)	(2)	(3)
<i>Panel A: Within-MP Variance</i>			
Average Over MPs	0.0294	0.0496	0.0484
Observations		12 MPs	
<i>Panel B: Across-MP Variance</i>			
Study 1	0.0229	0.0106	0.0424
Study 2	0.0256	0.0444	0.1566
Observations		12 MPs	

Notes: This table shows that there is as much variation in participant evaluations of MP appearance within MPs as there is across MPs. Variance estimates are based on the subset of 12 MPs whose appearance was assessed twice, using two different photos (one from study 1 and another from study 2). Panel A reports the variance in the mean scores for a given MP across the two studies, averaged over the 12 MPs. Panel B reports, for each study respectively, the average variance in mean score across the 12 MPs. Corruption evaluations range from 1 to 4 and are based on the response to the question, “How much of the CFF would this candidate keep for himself and not spend on community development.” Leadership evaluations range from 1 to 4 and are based on the response to the question “Do you think this candidate would be a good leader?” Attractiveness evaluations range from 1 to 4 and are based on the response to the question, “How good looking (fine, handsome) do you think this candidate is?”

Table A10: Arbitrary image differences do not predict voter inferences about competence

Dependent variable:	Trust Inferences	Effort Inferences
	(1)	(2)
Wearing Glasses	0.029 [0.646]	0.045 [0.228]
Smiling	0.043 [0.293]	-0.012 [0.651]
Brightness	0.001 [0.680]	0.000 [0.641]
Formal Dress	-0.001 [0.987]	-0.003 [0.904]
Traditional Dress	0.035 [0.505]	0.067* [0.068]
Facial Hair	-0.016 [0.703]	-0.022 [0.432]
Background	0.037 [0.622]	0.065 [0.212]
Camera 2	0.029 [0.756]	0.110* [0.092]
Camera 3	.	0.002 [0.977]
<i>F</i> -statistic on all coefficients	0.41	1.72
<i>p</i> -value	0.90	0.16
Observations	27 MPs	28 MPs

*Notes: This table shows that arbitrary differences visible in the sample of MP photos do not predict participant inferences about public spending performance (trustworthiness) or constituency visits (effort). Reported in brackets are *p*-values corresponding to unadjusted OLS standard errors. Significance levels indicated by * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.001$. The outcome variable in column 1 is trust inferences, which captures the share of study participants who thought the MP looked less corrupt than peers, averaged across multiple photo pairings. The outcome variable in column 2 is effort inferences, which captures the share of study participants who thought the MP made more visits than peers, averaged across multiple photo pairings. Explanatory variables are manually coded dummy variables reflecting characteristics of MP photos. Brightness refers to relative luminance, calculated as the weighted mean of all RGB color values composing the photograph. “Background” is an indicator to distinguish the external brick wall of Parliament from the internal white wall as the omitted category. “Camera” refers to the brand of camera the photograph was taken with. The *F*-statistic and associated *p*-values suggest that we cannot reject the null hypothesis that all of these arbitrary differences are jointly equal to zero for either column.*

Table A11: Photo inference accuracy is robust to excluding mixed gender pairs

	All photo pairs (1)	Same-gender only pairs (2)
<i>Panel A: Trustworthiness Inferences</i>		
Proportion correct	54.2*** (1.07) [0.000]	54.4*** (1.27) [0.001]
Observations	2,158	1,532
<i>Panel B: Effort Inferences</i>		
Proportion correct	45.7*** (1.07) [0.000]	44.1*** (1.40) [0.000]
Observations	2,185	1,253
<i>Panel C: Persuasion</i>		
Coefficient on photo selections	0.600*** (0.050) [0.000]	0.600*** (0.060) [0.000]
Observations	1,413	1,010

*Notes: This table shows that the accuracy of participant inferences about which MP in a photo pair performed better are robust to dropping evaluations of mixed-gender pairs. Panel A shows the proportion of respondents who correctly guess which MP spent more of the constituency facilitation fund (CFF) on development and kept less for themselves. Panel B shows the proportion of respondents who correctly guess which MP made more visits to their home constituency. Panel C reports regression estimates on whether photo selections (independent variable) predict the audio selections (dependent variable) of other study participants in the persuasiveness study. Reported in parentheses are standard errors, and in brackets are p-values associated with a t-test that the estimate is not statistically equivalent to chance guessing (which would be 0.50). Significance levels indicated by * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.001$.*