

Predicting political elections from rapid and unreflective face judgments

Charles C. Ballew II* and Alexander Todorov*†‡

*Department of Psychology and †Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ 08540

Edited by Edward E. Smith, Columbia University, New York, NY, and approved September 25, 2007 (received for review June 10, 2007)

Here we show that rapid judgments of competence based solely on the facial appearance of candidates predicted the outcomes of gubernatorial elections, the most important elections in the United States next to the presidential elections. In all experiments, participants were presented with the faces of the winner and the runner-up and asked to decide who is more competent. To ensure that competence judgments were based solely on facial appearance and not on prior person knowledge, judgments for races in which the participant recognized any of the faces were excluded from all analyses. Predictions were as accurate after a 100-ms exposure to the faces of the winner and the runner-up as exposure after 250 ms and unlimited time exposure (Experiment 1). Asking participants to deliberate and make a good judgment dramatically increased the response times and reduced the predictive accuracy of judgments relative to both judgments made after 250 ms of exposure to the faces and judgments made within a response deadline of 2 s (Experiment 2). Finally, competence judgments collected before the elections in 2006 predicted 68.6% of the gubernatorial races and 72.4% of the Senate races (Experiment 3). These effects were independent of the incumbency status of the candidates. The findings suggest that rapid, unreflective judgments of competence from faces can affect voting decisions.

face perception | social judgments | voting decisions

With the exception of the president, state governors are arguably among the most powerful elected officials in the United States. American states are significant political and economic entities, with some being larger and economically more powerful than many foreign countries. In addition to wielding power at the state level, governors have been historically well poised to ascend to the presidency. For example, 17 of 43 presidents have been governors, including four of five in the last 30 years. The power and potential of a governorship comes at a cost. In 1998, the 36 gubernatorial races averaged \$14.1 million in expenses each (1). By comparison, the average Senate campaign cost was \$3.3 million in 1996 (2).

Despite the significance of gubernatorial races, we show that rapid, unreflective judgments of competence based solely on facial appearance and made after as little as 100 ms of exposure to the faces of the winner and the runner-up predict election outcomes. In our prior work on forecasting the outcomes of Senate elections (3), we showed that people believe that competence is the most important attribute for a politician and that trait inferences of competence from faces but not other traits (e.g., trustworthiness, attractiveness, likeability, etc.) predict election outcomes. We argued that these inferences are rapid, intuitive, and unreflective, but we did not provide direct evidence for this assumption.

The first objective of the current research was to provide such evidence. The second objective was to replicate the Senate findings for gubernatorial races, which are arguably more important. The third objective was to test whether the effect of competence judgments on prediction of election outcomes is independent of the incumbency status of the candidates. In the Senate and House of Representatives elections, there are no terms limits, and incumbents have overwhelming odds of being reelected (4). In contrast,

many states have term limits for governors, and, correspondingly, there are fewer incumbents in gubernatorial races.

Faces are a rich source of social information, and trait judgments from faces can be made after minimal time exposure (5). For example, we have shown that 100 ms of exposure to a face is sufficient for people to make a variety of trait judgments, including competence, and that additional time only increases confidence in judgments (6). In Experiment 1, we tested whether competence judgments made after 100 ms of exposure to the faces of the candidates predict the outcomes of gubernatorial elections better than chance and whether additional time exposure (250 ms and unlimited time) improves the accuracy of prediction.

In our previous work (3) and Experiment 1, participants were asked to rely on their “gut” feeling or first impression when making the judgment. In Experiment 2, we studied the effect of deliberation on judgments. Deliberating about judgments that are unreflective and not easy to articulate can interfere with the quality (7) and consistency (8) of the judgments. If trait judgments from faces are unreflective, instructions to deliberate and make a good judgment should not improve the predictive accuracy of judgments. In Experiment 2, we tested whether deliberation judgments are less accurate in predicting the election outcomes than judgments made after 250 ms of exposure to the faces and judgments made under response deadline of 2 s, forcing participants to rely on quick judgments.

In Experiment 3, we collected competence judgments for both gubernatorial and Senate races in 2006 before the actual election. We tested whether these judgments based solely on facial appearance would predict the election outcomes better than chance, as we did in our previous work on predicting the Senate elections prospectively in 2004 (3).

Experiment 1

Participants were presented with the faces of the winner and the runner-up for 89 gubernatorial races and asked to judge which person was more competent. In two of the experimental conditions, the pair of faces was presented for 100 and 250 ms, respectively (Fig. 1). In the third condition the faces were presented until the participant responded, with no time constraints. For each race, participants made three consecutive judgments: a binary choice of who was more competent, a continuous judgment (on a nine-point scale) of how much more competent the chosen person was[§], and a recognition judgment. If participants recognized either of the faces for a given race, their responses for that race were not

Author contributions: C.C.B. and A.T. designed research; C.C.B. performed research; A.T. analyzed data; and A.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

†To whom correspondence should be addressed. E-mail: atodorov@princeton.edu.

[§]This measure did not contribute any additional information over the information gained from the binary competence judgments. Details are provided in *SI Text*.

This article contains supporting information online at www.pnas.org/cgi/content/full/0705435104/DC1.

© 2007 by The National Academy of Sciences of the USA

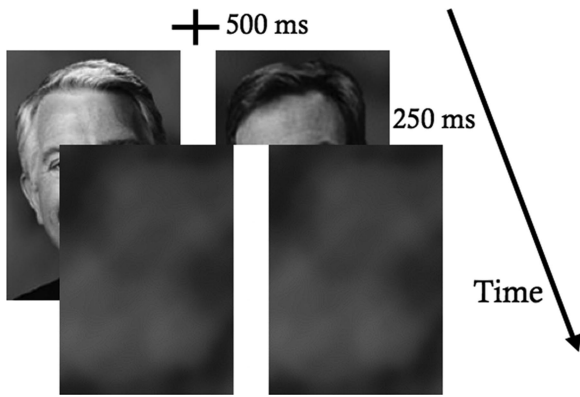


Fig. 1. An example of an experimental trial in the 250-ms presentation condition. Participants decided who was more competent.

analyzed. Thus, all judgments of competence were based only on facial appearance and not on other knowledge.

Results. Analysis at the level of participants. Participants in all three conditions were more likely to choose the winner than the runner-up as more competent, $P < 0.001$ against the chance prediction of 0.50. The judgments in the different conditions did not differ (Fig. 2A) ($F < 1$). Across conditions, the mean judgment was 0.57 [SE = 0.006; $t(119) = 11.22$, $P < 0.001$, $d = 2.05$]. Although the judgments did not differ across conditions, the response times in the unlimited time condition were twice as long as the response times in the 100-ms and 250-ms conditions (Fig. 3A). The response times in the latter two conditions did not differ from each other and were significantly shorter than the response times in the unlimited time condition [$t(117) = 13.16$, $P < 0.001$, and $t(44.84) = 10.22$, $P < 0.001$ (not assuming equal variance), respectively].

Analysis at the level of the races. Aggregating across participants, the percentage of correctly predicted races (i.e., races for which $>50\%$ of participants judged the winner as more competent) was higher in the 250-ms condition than in the other two conditions (Table 1), although this difference was not significant. Aggregating across the three experimental conditions, the binary competence judgments predicted 64.0% of the outcomes of the gubernatorial races, which was significantly higher than chance [$\chi^2(1) = 7.02$, $P < 0.008$].

We also tested whether the difference in competence between the two candidates was linearly related to the difference in votes between them. As shown in Table 1, in all conditions the average competence of the candidate correlated positively and significantly with the proportion of votes won by this candidate. The more competent the candidate was perceived to be relative to the other candidate, the higher the proportion of votes for this candidate.

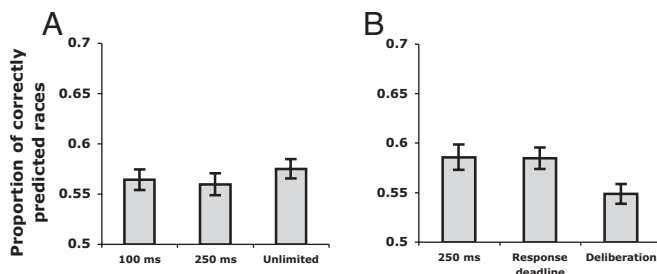


Fig. 2. Proportion of correctly predicted gubernatorial races in which the winner was judged as more competent than the runner-up. (A) As a function of time exposure to faces in Experiment 1. (B) As a function of experimental condition in Experiment 2: 250-ms exposure to faces, response deadline of 2 s, and deliberation. Error bars show the SEM.

Averaging across the three experimental conditions, the mean competence judgments for the candidates correlated 0.27 ($P < 0.011$), with the proportion of votes [supporting information (SI) Fig. 5]. Thus, snap judgments of competence from facial appearance accounted for 7.2% of the variance of vote share.

Discussion. These findings suggest that simple, fast, binary judgments of competence aggregated across a relatively small sample size of raters are sufficient to predict the outcomes of gubernatorial elections. Judgments made after as little as 100 ms of exposure to the faces of the candidates predicted the election outcomes better than chance. Additional time exposure to the faces did not improve these predictions, although the response times for the judgments substantially increased when the time exposure was unconstrained.

To our knowledge, this study is the first demonstration that judgments made after minimal time exposure to the faces of the candidates predict election outcomes. In our previous work (3), the minimum time exposure to the faces was 1 s. Clearly, much less exposure is needed for these judgments. The current findings are consistent with the ideas that trait judgments from faces can be characterized as rapid, unreflective, intuitive (“system 1”) judgments (e.g., refs. 9 and 10) and that, because of these properties, their influence on voting decisions may not be easily recognized by voters (3).

Experiment 2

In Experiment 2, we tested how instructions to deliberate and make a good judgment (rather than relying on a gut feeling or first impression) affect competence judgments. Participants were randomly assigned to one of three conditions: a deliberation condition in which they were asked to think carefully about their choice and make a good judgment, a response deadline condition in which they had to decide within 2 s, and a 250-ms replication condition.

We included a response deadline condition in which the faces were presented until the participants responded, but they had to respond within 2 s. As shown in Experiment 1 (Fig. 3A), this time was longer than the average response time for the 100- and 250-ms conditions but substantially shorter than the average response time for the unlimited time condition. Thus, the response deadline procedure should force participants to rely on quick judgments. If, as we argue, trait judgments from faces are rapid and unreflective, participants’ judgments in this condition should predict the outcomes of the elections better than chance. However, the judgments in the deliberation condition should be less predictive of the election outcomes than the judgments in the 250-ms and response deadline conditions.

From a psychological point of view, races in which the candidates are of the same gender and ethnicity are more interesting because differences in judgments of competence cannot be attributed to differences in gender and ethnicity. Moreover, the salience of the latter factors can activate gender and race stereotypes and, accordingly, change participants’ responses. In fact, when the analysis was limited to the 55 gubernatorial races in which the winner and the runner-up were of the same gender and ethnicity, the predictions improved, just as they did in our previous work on Senate elections (3). Averaging across the three conditions, the percentage of correctly predicted races was 69.1% [$\chi^2(1) = 8.02$, $P < 0.005$], and the linear correlation between the perceived competence of the candidates and the vote share was 0.32 ($P < 0.017$). Thus, in Experiment 2, participants made judgments only for the 55 races in which the candidates were of the same sex and ethnicity.

Results. Analysis at the level of participants. As in Experiment 1, participants in all three conditions were more likely to choose the winner than the runner-up as more competent ($P < 0.001$). However, the effect was smaller in the deliberation condition than in the 250-ms and response deadline conditions (Fig. 2B) [$F(2, 107) = 3.51$, $P < 0.033$ for the omnibus test]. Follow-up contrast

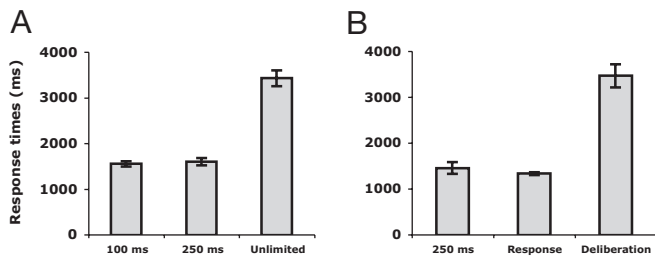


Fig. 3. Response times for competence judgments. (A) As a function of time exposure to faces in Experiment 1. (B) As a function of experimental condition in Experiment 2: 250-ms exposure to faces, response deadline of 2 s, and deliberation. Error bars show the SEM.

tests showed that, although the judgments in the latter two conditions were not significantly different ($t < 1$), they were significantly better than the judgments in the deliberation condition [$t(107) = 2.65, P < 0.009, d = 0.51$].

The response times in the deliberation condition were substantially longer than the response times in the 250-ms and response deadline conditions (Fig. 3B) [$t(107) = 10.34, P < 0.001, d = 2.0$ and $t(39.69) = 7.90, P < 0.001$ (not assuming equal variance), respectively]. The response times in the latter two conditions were not significantly different ($t < 1$).

Analysis at the level of the races. Aggregating across participants, the judgments in the 250-ms and response deadline conditions predicted a higher percentage of races than the judgments in the deliberation condition (Table 1), although these differences were not significant. The percentage of correctly predicted races in the deliberation condition was not significantly different from chance. Aggregating across the 250-ms and the response deadline conditions, the binary competence judgments predicted 70.9% of the gubernatorial races, which was significantly higher than chance [$\chi^2(1) = 9.62, P < 0.002$]. It should be noted that this prediction was better than the predictions in each of the conditions, 250 ms and response deadline (see Table 1), demonstrating that aggregating across more judges improves the prediction (see the supporting online material for ref. 3 for bootstrapping simulations).

As shown in Table 1, in all conditions the average competence of the candidate correlated positively with the proportion of votes won by the candidate, although the only correlation that reached significance was in the 250-ms condition. Aggregating across the 250-ms and response deadline conditions, the correlation between competence judgments and vote share was 0.32 ($P < 0.018$). Thus, rapid, unreflective judgments of competence from facial appearance accounted for 10.2% of the variance of vote share.

Deliberation judgments and unreflective judgments—judgments aggregated across the 250 ms and response deadline conditions—

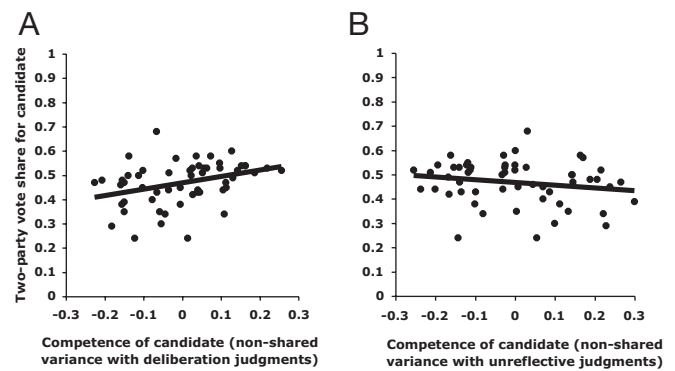


Fig. 4. Scatter plots of the two-party vote share for the candidates and nonshared variance of unreflective judgments of competence of the candidates (the x axis plots the regression residuals of unreflective judgments regressed on deliberation judgments) (A) and nonshared variance of deliberation judgments of competence of the candidates (the x axis plots the regression residuals of deliberation judgments regressed on unreflective judgments) (B). Each point represents a gubernatorial race. The line represents the best fitting linear curve.

were highly correlated ($r = 0.78, P < 0.001$). This shared variance is consistent with the possibility that deliberation judgments were anchored on rapid, immediate impressions from the faces. If this is the case, removing the shared variance between deliberation and unreflective judgments should not affect the correlation with the vote share for unreflective judgments, but it should affect this correlation for deliberation judgments. Partial correlation analysis confirmed this hypothesis. The partial correlation between unreflective judgments and vote share controlling for deliberation judgments was 0.34 ($P < 0.011$) (Fig. 4A). In contrast, the partial correlation between deliberation judgments and vote share controlling for unreflective judgments was -0.19 ($P = 0.18$) (Fig. 4B). **Analysis across both experiments.** Although the response times for the judgments in the unlimited time (Experiment 1) and deliberation (Experiment 2) conditions were almost identical (Fig. 3) (see also *SI Text and SI Fig. 6*), the predictive accuracy of judgments was only affected in the latter condition (Fig. 2). This finding suggests that intuitive judgments are affected by the deliberative mind set rather than by the additional time for judgments. Additional time does not necessarily lead to changes in judgments, although it may lead to increased confidence in judgments (6). Although judgments in the unlimited time condition and deliberation judgments share variance ($r = 0.82, P < 0.001$), perhaps reflecting controlled processing, this shared variance should not predict the outcome of the races to the extent that these predictions are based on rapid, intuitive judgments. Because Experiment 2 included a subset of the races used

Table 1. Percentage of correctly predicted gubernatorial races and correlations between perceived competence of candidates and their vote share as a function of experimental conditions in Experiments 1–3

| Experimental condition | Experiment 1 (89 races) | | Experiment 2 (55 races) | | Experiment 3 (35 races) | |
|--------------------------|----------------------------|-------------------|----------------------------|-------------------|----------------------------|----------|
| | % | <i>r</i> | % | <i>r</i> | % | <i>r</i> |
| 100-ms exposure to faces | 59.6* | 0.21 [†] | | | | |
| 250-ms exposure to faces | 68.5 [†] | 0.23 [†] | 67.3 [†] | 0.38 [†] | | |
| Unconstrained exposure | 62.9 [†] | 0.27 [†] | | | 68.6 [†] | 0.29* |
| Response deadline (2 s) | | | 65.5 [†] | 0.22 | | |
| Deliberation | | | 60.0 | 0.14 | | |

The percentages indicate the races in which the candidate who was perceived as more competent by the majority of participants won the race. The significance is tested against the chance prediction of 50%. *, $P < 0.10$; [†], $P < 0.05$.

in Experiment 1 and the analysis was conducted at the level of the races, we could test this hypothesis. Controlling for the shared variance did not affect the correlation between vote share and judgments in the unlimited time condition. The partial correlation (0.28, $P < 0.05$) was practically the same as the zero order correlation (0.27, $P < 0.05$). Thus, what predicted the outcomes of the races in the unlimited time condition in Experiment 1 was the nonshared variance with deliberation judgments. Finally, the predictive accuracy of the judgments in the unlimited time condition was eliminated when the analysis controlled for unreflective judgments. The correlation between vote share and the corrected time unconstrained judgments was reduced from 0.28 to -0.02 (see also *SI Text and SI Table 3*).

Discussion. As in Experiment 1, judgments made after 250 ms of exposure to the faces of the candidates predicted the outcomes of gubernatorial elections. This result was also obtained for judgments that were made within a response deadline of 2 s, forcing participants to rely on rapid, unreflective judgments. The judgments of participants who were asked to deliberate and make a good judgment were less accurate in predicting the election outcomes and substantially slower than the judgments of participants in the other two conditions.

Deliberation judgments shared a substantial amount of variance with unreflective judgments. Removing this variance did not affect the relation between vote share and unreflective judgments, but it did affect the relation between vote share and deliberation judgments. Specifically, whereas the corrected unreflective judgments predicted vote share, the corrected deliberation judgments did not predict vote share. If anything, the correlation between corrected deliberation judgments and vote share was negative. These findings are consistent with the hypothesis that deliberation judgments are anchored on rapid, automatic trait impressions from faces and that any positive relation between deliberation judgments and vote share can be accounted for by the shared variance of deliberation judgments with these automatic impressions. That is, what predicts the outcomes of elections is the automatic component of trait judgments. Deliberation instructions add noise to automatic trait judgments and, consequently, reduce the accuracy of prediction. This hypothesis is also consistent with the analyses across Experiments 1 and 2. What predicted the outcomes of the races in the unlimited time condition in Experiment 1 was not the variance that was shared with deliberation judgments, but the variance that was shared with rapid, intuitive judgments.

Experiment 3

In this experiment, we collected competence judgments 2 weeks before the gubernatorial elections in 2006 to demonstrate that these judgments can predict elections prospectively. Participants were presented with the pictures of the Democratic and Republican candidates for each gubernatorial race and asked to choose the more competent person by using their gut feeling. We also included the 2006 Senate races in this experiment.

Participants were more likely to choose the winner than the runner-up as more competent for both the gubernatorial [$M = 0.57$, $SE = 0.01$; $t(63) = 6.50$, $P < 0.001$, $d = 1.64$] and Senate [$M = 0.55$, $SE = 0.01$; $t(63) = 3.94$, $P < 0.001$, $d = 0.99$] races. Aggregating across participants, the judgments predicted 68.6% of the gubernatorial races [$\chi^2(1) = 4.83$, $P < 0.028$] against the chance prediction of 50%, and 72.4% of the Senate races [$\chi^2(1) = 5.83$, $P < 0.016$].

The correlation between the perceived competence of the candidates and their vote share was 0.47 ($P < 0.011$) for the Senate races and 0.29 ($P = 0.09$) for the gubernatorial races. Although the latter correlation was not significant, it was comparable in size to the correlations obtained in the other experiments (see Table 1). The small number of races in this experiment makes the rejection of the null hypothesis more difficult. Because in this experiment we

Table 2. Percentage of correctly predicted gubernatorial races by competence judgments as a function of incumbency status

| Experiment | Incumbent won | Incumbent lost | No incumbent |
|------------|--------------------|-------------------|--------------------|
| 1 | 65.8% ($n = 38$) | 85.7% ($n = 7$) | 59.1% ($n = 44$) |
| 3 | 70.8% ($n = 24$) | 100% ($n = 1$) | 60.0% ($n = 10$) |

For Experiment 1, the competence judgments were aggregated across the three experimental conditions.

used the same procedures as those used in the unlimited time condition in Experiment 1, we combined the mean judgments for the 35 gubernatorial races from 2006 and the 89 gubernatorial races from Experiment 1. For these 124 races, the correlation between the perceived competence of candidates and their vote share was 0.27 and highly significant ($P < 0.003$).

Replicating our prior findings of prospectively predicting the outcomes of the Senate races in 2004 (3), judgments of competence based solely on the facial appearance of the candidates and collected before the actual elections in 2006 predicted the outcomes of both gubernatorial and Senate elections.

Incumbency Status and Competence Judgments

Gubernatorial races are not only more important than House and Senate races, but also more interesting with respect to addressing the effects of incumbency and perceived competence of candidates on predictions of the election outcomes. It is a well known fact that incumbents have a distinctive advantage in American politics (4, 11). For example, in the House races studied by Todorov *et al.* (3), incumbents won in 89% of the races. In the Senate races, incumbents won in 74% of the races. If incumbents appear more competent than challengers and participants are choosing the incumbent more frequently than the challenger, the predictive effect of competence judgments may be explained as an artifact of incumbency status. That is, according to this account, competence judgments will predict the winner only in races in which the incumbents win. Gubernatorial races are particularly interesting for the test of this hypothesis because many states have term limits for governors and, correspondingly, there are fewer incumbents in these races. However, as shown in Table 2, there was no support for this hypothesis. Competence judgments were independent of incumbency status in predicting the outcome of the elections.

Because the races used in Experiment 1 included the races used in Experiment 2, we report the analysis only for Experiment 1. For simplicity of presentation, for Experiment 1, we used the competence judgments aggregated across the three experimental conditions. In all of these conditions, participants were instructed to rely on their gut feeling when making the judgments, and the results were identical when the analysis was performed separately for each condition. The test for dependence of judgments and incumbency status was not significant [$\chi^2(2) = 1.95$, $P = 0.38$] (see Table 2 for the relevant proportions). Collapsing across the races in which the incumbent lost and the races with no incumbent, the candidate who was perceived as more competent won in 62.7% of the races. The corresponding percentage for the races in which the incumbent won was 65.9% [$\chi^2(1) < 1$, $P = 0.77$]. For Experiment 3, as in the case of Experiment 1, the test for dependence of judgments and incumbency status was not significant [$\chi^2(2) < 1$, $P = 0.65$] (Table 2). Combining the races from both experiments ($n = 124$ races) to increase statistical power did not change the results. The candidate who was perceived as more competent won in 67.7% of the races in which the incumbent won and in 62.9% of the races in which the incumbent lost or there was no incumbent [$\chi^2(1) < 1$, $P = 0.57$]. Thus, incumbency status and perceived competence were independent predictors of the election outcomes.

Discussion

Extending our prior work on forecasting the outcomes of Senate elections (3), we have shown that rapid, unreflective judgments of competence based solely on facial appearance predicted the outcomes of gubernatorial elections. Even after 100 ms of exposure to the faces of the winner and the runner-up, participants were more likely to choose the winner as more competent. In addition to showing that people rapidly extract trait information from faces (5, 6), we also show that instructions to deliberate and make a good judgment led to less accurate predictions of the election outcomes. These findings are consistent with research showing that deliberation can interfere with the quality of unreflective judgments (7) and even with judgments that can be characterized on simple quantitative dimensions (12). For example, in a classic study Wilson and Schooler (7) showed that judgments of the quality of jams were worse after people were asked to reflect on the jams. Evaluating the quality of jams and making trait judgments from faces are quite different, but both processes rely on mechanisms that are most likely inaccessible to awareness (13). In a more apt analogy, verbally describing a face can interfere with face recognition (14, 15), and thinking about the reasons for liking faces can reduce the consistency of liking judgments (8).

The current findings contribute to a growing body of evidence that the outcomes of important elections can be predicted from person judgments (refs. 3 and 16, and D. Benjamin and J. Shapiro, personal communication). In the research of Benjamin and Shapiro, participants predicted the outcomes of gubernatorial races after observing 10 s of gubernatorial debates. When the sound of the debate was off or muffled, these judgments predicted the outcomes better than chance. The judgments remained a significant predictor of the vote share after controlling for incumbency, campaign spending, and a number of economic indicators. Interestingly, when the sound was on, predictions were at chance, suggesting that the useful information in terms of prediction was nonverbal and that inferring the party affiliation of the candidates and policy positions led to worse predictions. These findings are consistent with a large body of evidence in social psychology that “thin slices” of nonverbal behaviors can provide sufficient information for accurate social judgments (17–23). The current findings show that in the case of elections, even 100 ms of exposure to the faces of the candidates can provide sufficient information to predict the election outcomes.

A recent study suggests that presidential elections can be predicted by face judgments too. Using a morphing technique, Little *et al.* (ref. 16, study 1) created faces based on the shape differences between the candidates for the highest posts in the United States, United Kingdom, Australia, and New Zealand. These novel pairs of faces, although derived from the politicians’ faces, are not recognizable by participants. Remarkably, participants were more likely to choose the winner than the runner-up in a simulated voting procedure. We have shown that simulated voting decisions are highly correlated with judgments of competence (3), suggesting that the same mechanisms are operating when people are asked to make competence judgments and cast hypothetical votes for faces. Most likely, when faced with a voting choice between two faces, participants make a rapid judgment of competence and base their voting decision on this judgment.

How do effects of facial appearance play out in the real world? Certainly, having a competent face is not sufficient for electoral success. If a politician does not have the backing of one of the two major parties in the United States, his or her face would not make much of a difference. In almost all of the races that we have studied, the candidates represented these parties. Having the support of a major party, a politician with competent appearance can have higher chances of electoral success. However, competence as assessed in our studies is always relative. Thus, in some races a politician may appear more competent relative to the challenger,

and in others they may appear less competent. Finally, there are multiple routes through which competent appearance can affect electoral outcomes. For example, party leaders can promote competent-appearing candidates for key positions even though these candidates may not be that competent after all. At the level of voting decisions, competent appearance most likely would not affect strongly identified partisans but can affect voters who are not strongly identified with a party. In many cases, these are precisely the voters who can swing an election. Appearance can also affect decisions to vote. For example, competent-looking incumbents may deter undecided voters, who have a mild preference for the challengers, from voting for the challenger. Studies on actual voting decision processes will be critical to delineate the causal influences of appearance on electoral success. Our findings suggest that, in many cases, the effects of appearance on voting decisions may be subtle and not easily recognized by voters (cf. ref. 24).

We focused on judgments of competence because of our prior work, which showed that people believe that competence is one of the most important traits for a politician and that competence judgments predict election outcomes (3). However, the context of election can change the relative importance of traits and, consequently, voters’ preferences. For example, Little *et al.* (ref. 16, study 2), using the morphing procedure described above, showed that participants preferred the morphed George W. Bush face “in a time of war” context but preferred the morphed John Kerry face “in a time of peace” context. The former face was perceived as more masculine and dominant but less intelligent and forgiving. This finding suggests that “fitting the face to the context” may be a more important factor in elections than having a competent appearance.

Methods

Experiment 1. Participants. One hundred and twenty Princeton University undergraduate students participated in the study in exchange for \$5. Participants were randomly assigned to one of six experimental conditions: 3 (presentation time: 100 ms vs. 250 ms vs. unlimited time) \times 2 (counterbalancing of the position of the images: left vs. right). In prior bootstrapping simulations (see supporting online material for ref. 3), we have shown that reliable estimates of the perceived competence of the candidates can be obtained from a sample of \approx 40 participants. Thus, we recruited 40 participants for each of the presentation time conditions.

Gubernatorial races. Using the *Almanac of American Politics* (25), we compiled a list of all gubernatorial races from 1995 to 2002, excluding races with highly familiar politicians (e.g., Howard Dean). Pictures of the winner and the runner-up were collected from various Internet sources (e.g., CNN, Wikipedia, and local media sources). Seven races were unusable because standardized pictures of both major candidates could not be found. For the remaining 89 races, the image of each politician was cropped to 150×215 pixels, placed on a standard background, and converted to grayscale. Each race pair was combined into a single image with 30 pixels of white space separating the images. The winner of each race was placed on the right for half of the races (selected randomly) and on the left for the other half. The position of the images was counterbalanced across participants. In Experiment 2, we used only those races in which the candidates were of the same sex and ethnicity ($n = 55$). In Experiment 3, we used the same procedure to standardize the images of the candidates in the 2006 election.

Procedures. The instructions in all conditions emphasized that participants should rely on their gut reactions. Neither elections nor candidates were mentioned at any point.

The order of the 89 races was randomized for each participant. For each race, participants made three consecutive judgments: a binary competence judgment, a nine-point scale competence judgment for the person selected as more competent, and a recognition judgment. The intertrial interval was 1 s. Each trial started with a fixation cross (+) presented for 500 ms. The race pair image was displayed with the letter “A” under the face on the left and the letter



“B” under the face on the right. In the unlimited time condition, the faces were displayed along with a binary competence judgment measure (“Which person is more competent?”). In the timed conditions, the faces and letters were displayed for 100 or 250 ms and then replaced with a perceptual mask. The mask was a grayscale cloud filter that occupied the same area as the images (Fig. 1). The mask remained up along with the A/B letters and the binary measure prompt until the participant responded. Large neon A and B tabs were placed over the “w” and “p” keys on the keyboard, respectively. Thus, pressing the A tab always corresponded with choosing the candidate on the left (marked with an A) as more competent, and vice versa.

The binary competence judgment was followed by another blank screen (1,000 ms) and fixation cross (500 ms). The faces were presented again, as above, with the unlimited time condition simply displaying the faces with a scaled continuous competence measure presented below the faces: “On a scale of 1 to 9, how much more competent is this person?” Participants responded about the person whom they chose as more competent on the preceding trial using the 1 through 9 keys on the top of the keyboard. In the timed conditions, the faces were presented for 100 or 250 ms and replaced with masks when the question was displayed.

Finally, the faces were presented again and participants were asked whether they recognized either of the faces from outside the study. Large neon “NO” and “YES” tabs were placed over the “z” and “/” keys, respectively, to collect this response. In all conditions, faces were presented for an unlimited time to ensure the most conservative measure of recognition.

Preliminary analyses. To ensure that competence judgments were based solely on facial appearance and not on prior person knowledge, judgments for races in which the participant recognized any of the faces were excluded from all analyses. This procedure resulted in the exclusion of 4.4% of the judgments.

To test whether the difference in competence between the two candidates was linearly related to the difference in votes between them, we used a measure of the two-party vote share. In this analysis, both vote share (e.g., the vote for the Democratic candidate out of the total votes for Republican and Democratic candidates) and competence (e.g., the perceived competence of the Democratic candidate relative to the Republican candidate) are conditioned on the candidate’s party. The analysis is the same whether it is conditioned on the Republican or Democratic candi-

date, because the measures for the candidates are perfectly negatively correlated.

Experiment 2. Participants. One hundred and ten Princeton University undergraduate students participated in the studies in exchange for payment or partial course credit. Participants were randomly assigned to one of six experimental conditions: 3 (condition: 250 ms vs. response deadline vs. deliberation) × 2 (counterbalancing of the position of the images).

Procedures. In both the 250-ms and the response deadline conditions, the instructions were the same as those in Experiment 1. In the deliberation condition, participants were told that we were interested in thoughtful reactions and that they should think carefully and make a good choice. The order of the 55 race pairs was randomized for each participant. The procedures were the same as those in Experiment 1 except that we did not collect the continuous competence judgments, because these judgments did not contribute any additional information over the binary competence judgments in Experiment 1. The faces in the deliberation and the response deadline conditions were presented until the participant responded. However, in the latter condition participants had only 2 s to respond. After 2 s, the images were replaced by a blank screen (1 s) and a fixation point (500 ms) signaling the beginning of the next trial.

Experiment 3. Sixty-four Princeton University undergraduate students participated in the studies in exchange for partial course credit. Participants were randomly assigned to one of two experimental conditions (counterbalancing of the position of the images of Republican and Democratic candidates). The procedures were the same as in the unconstrained time condition in Experiment 1. Participants made judgments for 35 gubernatorial races and 29 Senate races. The order of the races was randomized for each participant. We excluded one gubernatorial race, because the incumbent was famous (Arnold Schwarzenegger in California) and would have been recognized by most participants; we also excluded four Senate races, because two races included famous incumbents (Hillary Clinton in New York and Joe Lieberman in Connecticut), and two included challengers that were unknown at the time of data collection, and pictures were unavailable.

We thank Amir Goren and Crystal Hall for comments on an earlier version of this paper, and Manish Pakrashi and Valerie Loehr for their help in running the experiments.

1. Moore JL (2003) *Elections A to Z* (CQ Press, Washington, DC), 2nd Ed.
2. Cantor JE (2001) *Campaign Financing*, CRS Issue Brief for Congress, No. IB87020 (Congressional Research Service, Library of Congress, Washington, DC).
3. Todorov A, Mandisodza AN, Goren A, Hall CC (2005) *Science* 308:1623–1626.
4. Gelman A, King G (1990) *Am J Polit Sci* 34:1142–1164.
5. Bar M, Neta M, Linz H (2006) *Emotion* 6:269–278.
6. Willis J, Todorov A (2006) *Psychol Sci* 17:592–598.
7. Wilson TD, Schooler JW (1991) *J Pers Soc Psychol* 60:181–192.
8. Levine GM, Halberstadt JB, Goldstone RL (1996) *J Pers Soc Psychol* 70:230–240.
9. Chaiken S, Trope Y, eds (1999) *Dual Process Theories in Social Psychology* (Guilford, New York).
10. Kahneman D (2003) *Am Psychol* 58:697–720.
11. Cover AD (1977) *Am J Polit Sci* 21:523–541.
12. Dijksterhuis A, Bos MW, Nordgren LF, van Baaren RB (2006) *Science* 311:1005–1007.
13. Nisbett RE, Wilson TD (1977) *Psychol Rev* 84:231–259.
14. Dodson CS, Johnson MK, Schooler JW (1997) *Mem Cognit* 25:129–139.
15. Schooler JW, Engstler-Schooler TY (1990) *Cognit Psychol* 22:36–71.
16. Little AC, Burriss RP, Jones BC, Roberts SC (2007) *Evol Hum Behav* 28:18–27.
17. Albright L, Kenny DA, Malloy TE (1988) *J Pers Soc Psychol* 55:387–395.
18. Ambady N, Hallahan M, Rosenthal R (1995) *J Pers Soc Psychol* 69:518–529.
19. Ambady N, Rosenthal R (1992) *Psychol Bull* 111:256–274.
20. Borkenau P, Liebler A (1992) *J Pers Soc Psychol* 65:645–657.
21. Kenny DA, Horner C, Kashy DA, Chu LC (1992) *J Pers Soc Psychol* 62:88–97.
22. Park B, Judd CM (1989) *J Pers Soc Psychol* 56:493–505.
23. Watson D (1989) *J Pers Soc Psychol* 57:120–128.
24. Hassin R, Trope Y (2000) *J Pers Soc Psychol* 78:837–852.
25. Barone M, Cohen R, eds (2004) *Almanac of American Politics* (National Journal Group, Washington, DC).