

Science

 AAAS

**Inferences of Competence from Faces Predict
Election Outcomes**

Alexander Todorov, *et al.*
Science **308**, 1623 (2005);
DOI: 10.1126/science.11110589

***The following resources related to this article are available online at
www.sciencemag.org (this information is current as of July 30, 2008):***

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/308/5728/1623>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/308/5728/1623/DC1>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/308/5728/1623#related-content>

This article **cites 16 articles**, 2 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/308/5728/1623#otherarticles>

This article has been **cited by** 25 article(s) on the ISI Web of Science.

This article has been **cited by** 11 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/308/5728/1623#otherarticles>

This article appears in the following **subject collections**:

Psychology

<http://www.sciencemag.org/cgi/collection/psychology>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

We suspect that this system is not unique. Several cod stocks, inhabiting similar oceanographic regimes (north of 44°N latitude) in the northwest Atlantic where they were the dominant predators, collapsed in the early 1990s (decline by >95% of maximum historical biomass) and failed to respond to complete cessation of fishing [there was one exceptional stock (table S1)]. For example, the current biomass of these stocks has increased only slightly, ranging from 0.4 to 7.0% during the past 10+ years (table S1). Reciprocal relationships between macroinvertebrate biomass and cod abundance in these areas (12) suggest that the processes that we document for the Scotian Shelf may have occurred there. On the other hand, the three major cod stocks resident south of 44° N, though reaching historical minimum levels at about the same time as the northerly stocks and experiencing similar intensive fishing pressure, declined by only 50 to 70%; current biomass has increased from 10 to 44% of historical minimum levels. These stocks inhabit different oceanographic regimes with respect to temperature and stratification and do not show the inverse relationship between the biomass of macroinvertebrates and cod found by Worm and Myers (12). These geographic differences in cod population dynamics merit additional study.

The changes in top-predator abundance and the cascading effects on lower trophic levels that we report reflect a major pertur-

bation of the eastern Scotian Shelf ecosystem. This perturbation has produced a new fishery regime in which the inflation-adjusted, monetary value of the combined shrimp and crab landings alone now far exceed that of the groundfish fishery it replaced (13). From an economic perspective, this may be a more attractive situation. However, one cannot ignore the fundamental importance of biological and functional diversity as a stabilizing force in ecosystems, and indeed in individual populations (20), in the face of possible future perturbations (whether natural or human-made). One must acknowledge the ecological risks inherent in “fishing down the food web” (21), as is currently occurring on the Scotian Shelf, or the ramifications associated with indirect effects reverberating across levels throughout the food web, such as altered primary production and nutrient cycling.

References and Notes

1. M. L. Pace, J. J. Cole, S. R. Carpenter, J. F. Kitchell, *Trends Ecol. Evol.* **14**, 483 (1999).
2. G. A. Polis, A. L. W. Sears, G. R. Huxel, D. R. Strong, J. Maron, *Trends Ecol. Evol.* **15**, 473 (2000).
3. J. B. C. Jackson, E. Sala, *Sci. Mar.* **65**, 273 (2001).
4. M. Scheffer, S. Carpenter, J. A. Foley, C. Folke, B. Walker, *Nature* **413**, 591 (2001).
5. P. C. Reid, E. J. V. Battle, S. D. Batten, K. M. Brander, *ICES J. Mar. Sci.* **57**, 495 (2000).
6. D. R. Strong, *Ecology* **73**, 747 (1992).
7. J. B. Shurin et al., *Ecol. Lett.* **5**, 785 (2002).
8. J. Terborgh et al., *Science* **294**, 1923 (2001).
9. J. A. Estes, M. T. Tinker, T. M. Williams, D. F. Doak, *Science* **282**, 473 (1998).

10. J. H. Steele, J. S. Collie, in *The Global Coastal Ocean: Multiscale Interdisciplinary Processes*, A. R. Robinson, K. Brink, Eds. (Harvard Univ. Press, Cambridge, MA, 2004), vol. 13, chap. 21.
11. F. Micheli, *Science* **285**, 1396 (1999).
12. B. Worm, R. A. Myers, *Ecology* **84**, 162 (2003).
13. Materials and methods are available as supporting material on Science Online.
14. J. B. Jackson et al., *Science* **293**, 629 (2001).
15. L. P. Fanning, R. K. Mohn, W. J. MacEachern, *Canadian Science Advisory Secretariat Research Document 27* (2003).
16. W. D. Bowen, J. McMillan, R. Mohn, *ICES J. Mar. Sci.* **60**, 1265 (2003).
17. J. S. Choi, K. T. Frank, B. D. Petrie, W. C. Leggett, *Oceanogr. Mar. Biol. Annu. Rev.* **43**, 47 (2005).
18. K. T. Frank, N. L. Shackell, J. E. Simon, *ICES J. Mar. Sci.* **57**, 1023 (2000).
19. J. S. Choi, K. T. Frank, W. C. Leggett, K. Drinkwater, *Can. J. Fish. Aquat. Sci.* **61**, 505 (2004).
20. K. T. Frank, D. Brickman, *Can. J. Fish. Aquat. Sci.* **57**, 513 (2000).
21. D. Pauly, V. Christensen, J. Dalsgaard, R. Froese, F. C. Torres Jr., *Science* **279**, 860 (1998).
22. We thank the Department of Fisheries and Oceans staff who collected and maintained the data with care and thoroughness, and M. Pace, N. L. Shackell, J. E. Carscadden, and two anonymous reviewers for helpful criticisms. This research was supported by Fisheries and Oceans Canada and a grant from the Natural Sciences and Engineering Research Council of Canada Discovery (to K.T.F. and W.C.L.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5728/1621/DC1

Materials and Methods
SOM Text
Table S1
References

4 April 2005; accepted 7 April 2005
10.1126/science.1113075

Inferences of Competence from Faces Predict Election Outcomes

Alexander Todorov,^{1,2*} Anesu N. Mandisodza,^{1†} Amir Goren,¹ Crystal C. Hall¹

We show that inferences of competence based solely on facial appearance predicted the outcomes of U.S. congressional elections better than chance (e.g., 68.8% of the Senate races in 2004) and also were linearly related to the margin of victory. These inferences were specific to competence and occurred within a 1-second exposure to the faces of the candidates. The findings suggest that rapid, unreflective trait inferences can contribute to voting choices, which are widely assumed to be based primarily on rational and deliberative considerations.

Faces are a major source of information about other people. The rapid recognition of familiar individuals and communication cues (such as expressions of emotion) is critical for successful social interaction (1). Howev-

er, people go beyond the inferences afforded by a person's facial appearance to make inferences about personal dispositions (2, 3). Here, we argue that rapid, unreflective trait inferences from faces influence consequential decisions. Specifically, we show that inferences of competence, based solely on the facial appearance of political candidates and with no prior knowledge about the person, predict the outcomes of elections for the U.S. Congress.

In each election cycle, millions of dollars are spent on campaigns to disseminate infor-

mation about candidates for the U.S. House of Representatives and Senate and to convince citizens to vote for these candidates. Is it possible that quick, unreflective judgments based solely on facial appearance can predict the outcomes of these elections? There are many reasons why inferences from facial appearance should not play an important role in voting decisions. From a rational perspective, information about the candidates should override any fleeting initial impressions. From an ideological perspective, party affiliation should sway such impressions. Party affiliation is one of the most important predictors of voting decisions in congressional elections (4). From a voter's subjective perspective, voting decisions are justified not in terms of the candidate's looks but in terms of the candidate's position on issues important to the voter.

Yet, from a psychological perspective, rapid automatic inferences from the facial appearance of political candidates can influence processing of subsequent information about these candidates. Recent models of social cognition and decision-making (5, 6) posit a qualitative distinction between fast, unreflective, effortless “system 1” processes and slow, deliberate, effortful “system 2” processes. Many inferences about other people, including inferences from facial appearance,

¹Department of Psychology, ²Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ 08544, USA.

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

†Present address: Department of Psychology, New York University, New York, NY 10003, USA.

can be characterized as system 1 processes (7, 8). The implications of the dual-process perspective are that person impressions can be formed “on-line” in the very first encounter with the person and can have subtle and often subjectively unrecognized effects on subsequent deliberate judgments.

Competence emerges as one of the most important trait attributes on which people evaluate politicians (9–11). If voters evaluate political candidates on competence, inferences of competence from facial appearance could influence their voting decisions. To test this hypothesis, we asked naïve participants to evaluate candidates for the U.S. Senate (2000, 2002, and 2004) and House (2002 and 2004) on competence (12). In all studies, participants were presented with pairs of black-and-white head-shot photographs of the winners and the runners-up (Fig. 1A) from the election races.

As shown in Table 1, the candidate who was perceived as more competent won in 71.6% of the Senate races and in 66.8% of the House races (13). Although the data for the 2004 elections were collected before the actual elections (14), there were no differences between the accuracy of the prospective predictions for these elections and the accuracy of the retrospective predictions for the 2000 and 2002 elections (15). Inferences of competence not only predicted the winner but also were linearly related to the margin of victory. To model the relation between inferred competence and actual votes, we computed for each race the difference in the proportion of votes (16). As shown in Fig. 1B, competence judgments were positively correlated with the differences in votes between the candidates for Senate [$r(95) = 0.44, P < 0.001$] (17, 18). Similarly, the correlation was 0.37 ($P < 0.001$) for the 2002 House races and 0.44 ($P < 0.001$) for the 2004 races. Across 2002 and 2004, the correlation was 0.40 ($P < 0.001$).

In the previous studies, there were no time constraints on the participants’ judgments. However, system 1 processes are fast and efficient. Thus, minimal time exposure to the faces should be sufficient for participants to make inferences of competence. We conducted an experiment in which 40 participants (19) were exposed to the faces of the candidates for 1 s (per pair of faces) and were then asked to make a competence judgment. The average response time for the judgment was about 1 s (mean = 1051.60 ms, SD = 135.59). These rapid judgments based on minimal time exposure to faces predicted 67.6% of the actual Senate races ($P <$

0.004) (20). The correlation between competence judgments and differences in votes was 0.46 ($P < 0.001$).

The findings show that 1-s judgments of competence suffice to predict the outcomes of actual elections, but perhaps people are making global inferences of likability rather than specific inferences of competence. To address this alternative hypothesis, we asked participants to make judgments on seven different trait dimensions: competence, intelligence, leadership, honesty, trustworthiness, charisma, and likability (21). From a simple halo-effect perspective (22), participants should evaluate the candidates in the same manner across traits. However, the trait judgments were highly differentiated. Factor anal-

ysis showed that the judgments clustered in three distinctive factors: competence (competence, intelligence, leadership), trust (honesty, trustworthiness), and likability (charisma, likability), each accounting for more than 30% of the variance in the data (table S1). More important, only the judgments forming the competence factor predicted the outcomes of the elections. The correlation between the mean score across the three judgments (competence, intelligence, leadership) and differences in votes was 0.58 ($P < 0.001$). In contrast to competence-related inferences, neither the trust-related inferences ($r = -0.09, P = 0.65$) nor the likability-related inferences ($r = -0.17, P = 0.38$) predicted differences in votes. The correlation between the competence judgment

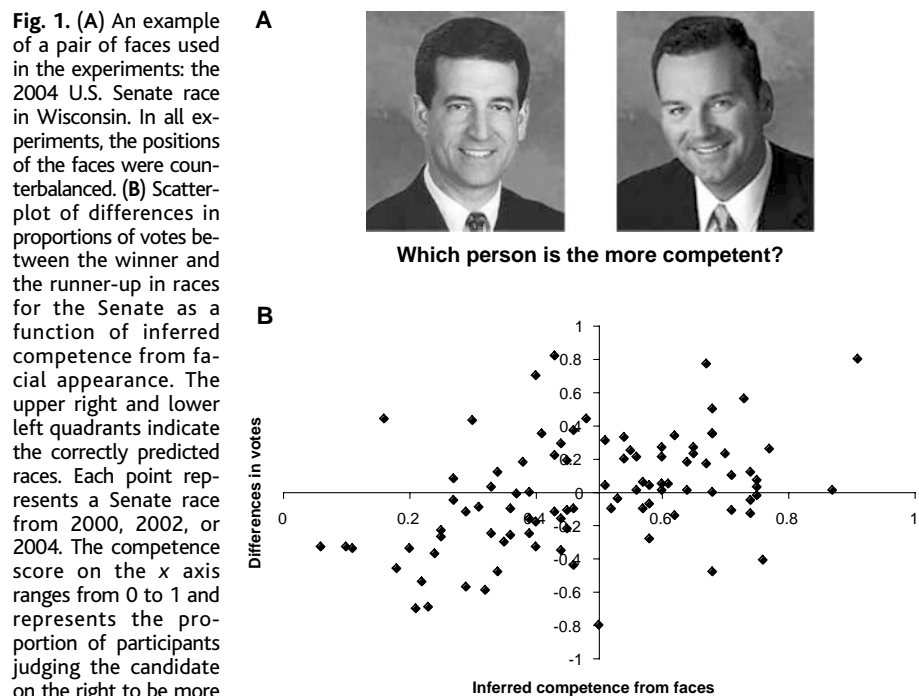


Fig. 1. (A) An example of a pair of faces used in the experiments: the 2004 U.S. Senate race in Wisconsin. In all experiments, the positions of the faces were counterbalanced. (B) Scatterplot of differences in proportions of votes between the winner and the runner-up in races for the Senate as a function of inferred competence from facial appearance. The upper right and lower left quadrants indicate the correctly predicted races. Each point represents a Senate race from 2000, 2002, or 2004. The competence score on the x axis ranges from 0 to 1 and represents the proportion of participants judging the candidate on the right to be more competent than the one on the left. The midpoint score of 0.50 indicates that the candidates were judged as equally competent. The difference in votes on the y axis ranges from -1 to $+1$ [(votes of candidate on the right – votes of candidate on the left)/(sum of votes)]. Scores below 0 indicate that the candidate on the left won the election; scores above 0 indicate that the candidate on the right won the election. [Photos in (A): Capitol Advantage]

Table 1. Percentage of correctly predicted races for the U.S. Senate and House of Representatives as a function of the perceived competence of the candidates. The percentages indicate the races in which the candidate who was perceived as more competent won the race. The χ^2 statistic tests the proportion of correctly predicted races against the chance level of 50%.

Election	Correctly predicted	χ^2
	<i>U.S. Senate</i>	
2000 ($n = 30$)	73.3%	6.53 ($P < 0.011$)
2002 ($n = 33$)	72.7%	6.82 ($P < 0.009$)
2004 ($n = 32$)	68.8%	4.50 ($P < 0.034$)
Total ($n = 95$)	71.6%	17.70 ($P < 0.001$)
	<i>U.S. House of Representatives</i>	
2002 ($n = 321$)	66.0%	33.05 ($P < 0.001$)
2004 ($n = 279$)	67.7%	35.13 ($P < 0.001$)
Total ($n = 600$)	66.8%	68.01 ($P < 0.001$)

alone and differences in votes was 0.55 ($P < 0.002$), and this judgment correctly predicted 70% of the Senate races ($P < 0.028$). These findings show that people make highly differentiated trait inferences from facial appearance and that these inferences have selective effects on decisions.

We also ruled out the possibility that the age, attractiveness, and/or familiarity with the faces of the candidates could account for the relation between inferences of competence and election outcomes. For example, older candidates can be judged as more competent (23) and be more likely to win. Similarly, more attractive candidates can be judged more favorably and be more likely to win (24). In the case of face familiarity, though unrecognized by our participants, incumbents might be more familiar than challengers, and participants might have misattributed this familiarity to competence (25). However, a regression analysis controlling for all judgments showed that the only significant predictor of differences in votes was competence (Table 2). Competence alone accounted for 30.2% of the variance for the analyses of all Senate races and 45.0% of the variance for the races in which candidates were of the same sex and ethnicity. Thus, all other judgments combined contrib-

uted only 4.7% of the variance in the former analysis and less than 1.0% in the latter analysis.

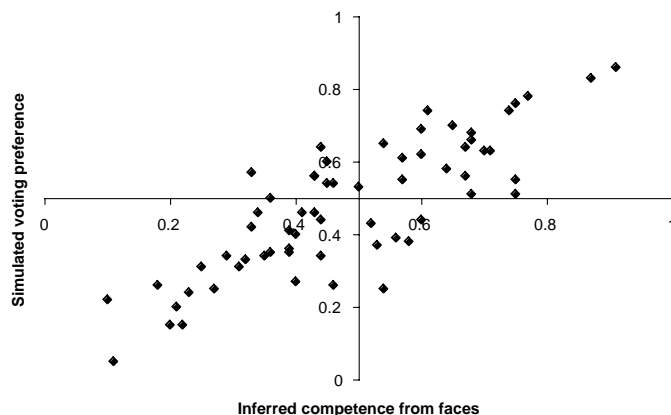
Actual voting decisions are certainly based on multiple sources of information other than inferences from facial appearance. Voters can use this additional information to modify initial impressions of political candidates. However, from a dual-system perspective, correction of intuitive system 1 judgments is a prerogative of system 2 processes that are attention-dependent and are often anchored on intuitive system 1 judgments. Thus, correction of initial impressions may be insufficient (26). In the case of voting decisions, these decisions can be anchored on initial inferences of competence from facial appearance. From this perspective, in the absence of any other information, voting preferences should be closely related to such inferences. In real-life voting decisions, additional information may weaken the relation between inferences from faces and decisions but may not change the nature of the relation.

To test this hypothesis, we conducted simulated voting studies in which participants were asked to choose the person they would have voted for in a political election (27). If voting preferences based on facial appearance

derive from inferences of competence, the revealed preferences should be highly correlated with competence judgments. As shown in Fig. 2, the correlation was 0.83 ($P < 0.001$) (28). By comparison, the correlation between competence judgments and actual differences in votes was 0.56 ($P < 0.001$). These findings suggest that the additional information that voters had about the candidates diluted the effect of initial impressions on voting decisions. The simulated votes were also correlated with the actual votes [$r(63) = 0.46, P < 0.001$] (29, 30). However, when controlling for inferences of competence, this correlation dropped to 0.01 ($P = 0.95$), which suggests that both simulated and actual voting preferences were anchored on inferences of competence from facial appearance.

Our findings have challenging implications for the rationality of voting preferences, adding to other findings that consequential decisions can be more “shallow” than we would like to believe (31, 32). Of course, if trait inferences from facial appearance are correlated with the underlying traits, the effects of facial appearance on voting decisions can be normatively justified. This is certainly an empirical question that needs to be addressed. Although research has shown that inferences from thin slices of nonverbal behaviors can be surprisingly accurate (33), there is no good evidence that trait inferences from facial appearance are accurate (34–39). As Darwin recollected in his autobiography (40), he was almost denied the chance to take the historic Beagle voyage—the one that enabled the main observations of his theory of evolution—on account of his nose. Apparently, the captain did not believe that a person with such a nose would “possess sufficient energy and determination.”

Fig. 2. Scatterplot of simulated voting preferences as a function of inferred competence from facial appearance. Each point represents a U.S. Senate race from 2000 or 2002. One group of participants was asked to cast hypothetical votes and another group was asked to judge the competence of candidates. Both the competence score and the voting preference score range from 0 to 1.



the competence score represents the proportion of participants judging the candidate on the right to be more competent than the one on the left. The preference score represents the proportion of participants choosing the candidate on the right over the one on the left. The midpoint score of 0.50 on the x axis indicates that the candidates were judged as equally competent. The midpoint score of 0.50 on the y axis indicates lack of preference for either of the candidates.

Table 2. Standardized regression coefficients of competence, age, attractiveness, and face familiarity judgments as predictors of differences in proportions of votes between the winner and the runner-up in races for the U.S. Senate in 2000 and 2002. Matched races are those in which both candidates were of the same sex and ethnicity.

Predictor	Differences in votes between winner and runner-up	
	All races	Matched races
Competence judgments	0.49 ($P < 0.002$)	0.58 ($P < 0.002$)
Age judgments	0.26 ($P < 0.061$)	0.07 ($P = 0.62$)
Attractiveness judgments	0.07 ($P = 0.63$)	0.08 ($P = 0.62$)
Face familiarity judgments	-0.05 ($P = 0.76$)	0.03 ($P = 0.86$)
Accounted variance (R^2)	34.9%	45.8%
Number of races	63	47

References and Notes

1. J. V. Haxby, E. A. Hoffman, M. I. Gobbini, *Trends Cognit. Sci.* **4**, 223 (2000).
2. R. Hassin, Y. Trope, *J. Pers. Soc. Psychol.* **78**, 837 (2000).
3. L. A. Zebrowitz, *Reading Faces: Window to the Soul?* (Westview, Boulder, CO, 1999).
4. L. M. Bartels, *Am. J. Polit. Sci.* **44**, 35 (2000).
5. S. Chaiken, Y. Trope, Eds., *Dual Process Theories in Social Psychology* (Guilford, New York, 1999).
6. D. Kahneman, *Am. Psychol.* **58**, 697 (2003).
7. A. Todorov, J. S. Uleman, *J. Exp. Soc. Psychol.* **39**, 549 (2003).
8. J. S. Winston, B. A. Strange, J. O’Doherty, R. J. Dolan, *Nat. Neurosci.* **5**, 277 (2002).
9. D. R. Kinder, M. D. Peters, R. P. Abelson, S. T. Fiske, *Polit. Behav.* **2**, 315 (1980).
10. In one of our studies, 143 participants were asked to rate the importance of 13 different traits in considering a person for public office. These traits included competence, trustworthiness, likability, and 10 additional traits mapping into five trait dimensions that are generally believed by personality psychologists to explain the structure of personality: extraversion, neuroticism, conscientiousness, agreeableness, and openness to experience (11). Competence was rated as the most important trait. The mean importance assigned to competence was 6.65 (SD = 0.69) on a scale

- ranging from 1 (not at all important) to 7 (extremely important). The importance assigned to competence was significantly higher than the importance assigned to any of the other 12 traits ($P_s < 0.005$).
11. S. D. Gosling, P. J. Rentfrow, W. B. Swan Jr., *J. Res. Pers.* **37**, 504 (2003).
 12. See supporting data on Science Online.
 13. For the House races in 2002, we were able to obtain pictures of both the winner and the runner-up for 321 of the 435 races. For the House races in 2004, we were able to obtain pictures for 279 of the 435 races (12).
 14. In the studies involving these races, we used photographs of the Democratic and Republican candidates (12).
 15. In addition, the accuracy of the predictions was not affected by the race and sex of the candidates. This is important because participants might have used race and sex stereotypes to make competence judgments for contests in which the candidates were of different sexes and races. For example, in such contests Caucasian male candidates were more likely to win. However, if anything, competence judgments predicted the outcomes of elections in which the candidates were of the same sex and race (73.1% for the Senate and 68.5% for the House) more accurately than elections in which they were of different sexes and races (67.9% and 64.3%, respectively). This difference possibly reflects participants' social desirability concerns when judging people of different race and sex.
 16. For races with more than two candidates, we standardized this difference so that it was comparable to the difference in races with two candidates. Specifically, the difference between the votes of the winner and those of the runner-up was divided by the sum of their votes.
 17. From the scatterplot showing the relation between competence judgments and votes for Senate (Fig. 1B), seven races (three in the lower right quadrant and four in the upper left quadrant) could be identified as deviating from the linear trend. It is a well-known fact that incumbents have an advantage in U.S. elections (18). In six of the seven races, the incumbent won but was judged as less competent. In the seventh race (Illinois, 2004) there was no incumbent, but the person who won, Barack Obama, was the favorite long before the election. Excluding these seven races, the correlation between competence judgments and differences in votes increased to 0.64 ($P < 0.001$). Although incumbent status seemed to affect the strength of the linear relation between inferences of competence and the margin of victory, it did not affect the prediction of the outcome. Competence judgments predicted the outcome in 72.9% of the races in which the incumbent won, in 66.7% of the races in which the incumbent lost, and in 68.8% of the cases in which there was no incumbent ($\chi^2 < 1.0$ for the difference between these percentages; $P = 0.89$).
 18. A. D. Cover, *Am. J. Polit. Sci.* **21**, 523 (1977).
 19. A bootstrapping data simulation showed that increasing the sample size to more than 40 participants does not improve the accuracy of prediction substantially (12) (fig. S1).
 20. Given the time constraints in this study, to avoid judgments based on salient differences such as race and sex, we used only Senate races (2000, 2002, and 2004) in which the candidates were of the same sex and race.
 21. For this study, we used the 2002 Senate races. The judgments in this and the subsequent studies were performed in the absence of time constraints (12).
 22. H. H. Kelley, *J. Pers.* **18**, 431 (1950).
 23. J. M. Montepare, L. A. Zebrowitz, *Adv. Exp. Soc. Psychol.* **30**, 93 (1998).
 24. T. L. Budenheim, S. J. DePaola, *Pers. Soc. Psychol. Bull.* **20**, 339 (1994).
 25. C. M. Kelley, L. L. Jacoby, *Acta Psychol. (Amsterdam)* **98**, 127 (1998).
 26. D. T. Gilbert, in *Unintended Thought*, J. S. Uleman, J. A. Bargh, Eds. (Prentice-Hall, Englewood Cliffs, NJ, 1989), pp. 189–211.
 27. For these studies, we used the 2000 and 2002 Senate races (12).
 28. An additional analysis from a study in which participants made judgments of the candidates for the Senate (2000 and 2002) on 13 different traits [see (10) for the list of traits] provided additional evidence that inferences of competence were the key determinants of voting preferences in this situation. We regressed voting preferences on the 13 trait judgments. The only significant predictor of these preferences was the judgment of competence [$\beta = 0.67$, $t(49) = 4.46$, $P < 0.001$].
 29. A similar finding was obtained in an early study conducted in Australia (30). Hypothetical votes based on newspaper photographs of 11 politicians were closely related to the actual votes in a local government election. Moreover, both hypothetical and actual votes correlated with inferences of competence.
 30. D. S. Martin, *Aust. J. Psychol.* **30**, 255 (1978).
 31. G. A. Quattrone, A. Tversky, *Am. Polit. Sci. Rev.* **82**, 719 (1988).
 32. J. R. Zaller, *The Nature and Origins of Mass Opinion* (Cambridge Univ. Press, New York, 1992).
 33. N. Ambady, F. J. Bernieri, J. A. Richeson, *Adv. Exp. Soc. Psychol.* **32**, 201 (2000).
 34. There is some evidence that judgments of intelligence from facial appearance correlate modestly with IQ scores (35). However, these correlations tend to be small [e.g., < 0.18 in (35)], they seem to be limited to judgments of people from specific age groups (e.g., puberty), and the correlation is accounted for by the judges' reliance on physical attractiveness. That is, attractive people are perceived as more intelligent, and physical attractiveness is modestly correlated with IQ scores.
 35. L. A. Zebrowitz, J. A. Hall, N. A. Murphy, G. Rhodes, *Pers. Soc. Psychol. Bull.* **28**, 238 (2002).
 36. Mueller and Mazur (37) found that judgments of dominance from facial appearance of cadets predicted military rank attainment. However, these judgments did not correlate with a relatively objective measure of performance based on academic grades, peer and instructor ratings of leadership, military aptitude, and physical education grades.
 37. U. Mueller, A. Mazur, *Soc. Forces* **74**, 823 (1996).
 38. There is evidence that trait inferences from facial appearance can be wrong. Collins and Zebrowitz [cited in (23), p. 136] showed that baby-faced individuals who are judged as less competent than mature-faced individuals actually tend to be more intelligent. There is also evidence that subtle alterations of facial features can influence the trait impressions of highly familiar presidents such as Reagan and Clinton (39).
 39. C. F. Keating, D. Randall, T. Kendrick, *Polit. Psychol.* **20**, 593 (1999).
 40. F. Darwin, Ed., *Charles Darwin's Autobiography* (Henry Schuman, New York, 1950), p. 36.
 41. Supported by the Department of Psychology and the Woodrow Wilson School of Public and International Affairs at Princeton University. We thank M. Savard, R. Hackell, M. Gerbasi, E. Smith, B. Padilla, M. Pakrashi, J. Wey, and R. G.-L. Tan for their help with this project and E. Shafir, D. Prentice, S. Fiske, A. Conway, L. Bartels, M. Prior, D. Lewis, and two anonymous reviewers for their comments on previous drafts of this paper.

Supporting Online Material
www.sciencemag.org/cgi/content/full/308/5728/1623/DC1
 Materials and Methods
 SOM Text
 Fig. S1
 Table S1
 References

2 February 2005; accepted 7 April 2005
 10.1126/science.1110589

TLR11 Activation of Dendritic Cells by a Protozoan Profilin-Like Protein

Felix Yarovsky,^{1*} Dekai Zhang,³ John F. Andersen,² Gerard L. Bannenberg,^{4†} Charles N. Serhan,⁴ Matthew S. Hayden,³ Sara Hieny,¹ Fayyaz S. Sutterwala,³ Richard A. Flavell,³ Sankar Ghosh,³ Alan Sher^{1*}

Mammalian Toll-like receptors (TLRs) play an important role in the innate recognition of pathogens by dendritic cells (DCs). Although TLRs are clearly involved in the detection of bacteria and viruses, relatively little is known about their function in the innate response to eukaryotic microorganisms. Here we identify a profilin-like molecule from the protozoan parasite *Toxoplasma gondii* that generates a potent interleukin-12 (IL-12) response in murine DCs that is dependent on myeloid differentiation factor 88. *T. gondii* profilin activates DCs through TLR11 and is the first chemically defined ligand for this TLR. Moreover, TLR11 is required in vivo for parasite-induced IL-12 production and optimal resistance to infection, thereby establishing a role for the receptor in host recognition of protozoan pathogens.

Mammalian Toll-like receptors (TLRs) play a fundamental role in the initiation of immune responses to infectious agents through their recognition of conserved microbial molecular patterns (1). TLR signaling in antigen-presenting cells, such as dendritic cells (DCs), results in the production of cytokines and costimulatory molecules that are required for initiation of the adaptive immune response (2, 3). Human and mouse TLR family mem-

bers have been shown to have distinct ligand specificities, recognizing molecular structures such as lipopeptide (TLR2) (4), lipopolysaccharide (TLR4) (5, 6), flagellin (TLR5) (7), double- and single-stranded RNA (TLR3 and TLR7) (8–11), and CpG motifs of DNA (TLR9) (12). Although several TLRs have been shown to be important for immune responses to microbial products in vitro, their role in host resistance to infection appears to be complex and not

Materials and Methods

Participants

Participants were undergraduate or graduate students at Princeton University. They participated for partial fulfillment of course credit or for payment. The total number of participants across all studies was 843.

Selection of photographs

All initial photographs were obtained from the website of the Cable News Network (CNN). There are 33 to 34 races for the US Senate and 435 races for the US House of Representatives every two years. For the Senate races in 2000 and 2002, and the House races in 2002, we obtained photographs of the winner and the runner-up in the race. For the Senate and House races in 2004, we obtained photographs of the Republican and Democratic candidates. If the quality of a photograph was poor, a web search for a new photograph was undertaken and if a better quality photograph of the candidate was found, it replaced the old photograph.

Because we were interested in inferences from facial appearance, uncontaminated by prior knowledge, races involving highly familiar individuals were excluded from the studies. For the Senate elections in 2000, we excluded the races for New York (Hillary Clinton), Connecticut (Joe Lieberman), New Jersey (John Corzine), and Missouri (John Aschcroft). The New Jersey race was excluded because all studies were conducted in Princeton, NJ, and John Corzine's campaign was extensively covered in the media. For the Senate elections in 2002, we excluded the race for Massachusetts (John Kerry). For the Senate elections in 2004, we excluded the race for Arizona (John McCain). The race

for Idaho was excluded too, because Senator Mike Crapo ran unopposed. Thus, we used 30 pairs of faces for the 2000 Senate races, 33 pairs for the 2002 races, and 32 pairs for the 2004 races.

For many of the House of Representatives races, the photograph of one of the two major candidates was missing and we were unable to include these races. We also excluded races that were uncontested (i.e., there was only one candidate). Finally, we excluded races involving highly familiar individuals. For the 2002 elections, these were the race for Ohio (Dennis Kucinich) and the race for Missouri (Dick Gephardt), both candidates highly familiar from the Democratic presidential primary. In total, we were able to obtain photographs of the winner and the runner-up for 321 House races in 2002, and photographs of the Republican and Democratic candidates for 279 House races in 2004.

All photographs were transformed to black-and-white bitmap files and standardized in size. Any conspicuous background (e.g., the Capital or a U.S. flag) was removed and replaced with gray background. For the photographs of candidates for the 2004 election, we created a standard gray background. Then, we removed all faces from their original background and superimposed them on the standard gray background. In all studies, each critical trial consisted of a pair of standardized black-and-white photographs of the major contenders for a congressional race (Fig. 1a). The photographs were presented either in a questionnaire format or on a computer screen.

General experimental procedures

In all studies, the position of the photographs was counterbalanced across participants. We created two versions for each pair of faces - one in which the photograph of the winner was positioned on the right side, and another in which the same photograph was positioned on the left side. For the election in 2004, because the data were collected before the outcome of the election was known, we used the party affiliation of the candidates to counterbalance the positions of the photographs. That is, in one version the photograph of the Democratic candidate was positioned on the right side, and in another it was positioned on the left side. To avoid confounding the position of the photograph and the election outcome, within each election – Senate 2000, Senate 2002, and House 2002 – for half of the races the photograph of the winner was positioned on the right side and for the other half it was positioned on the left side. Within the 2004 elections, for half of the races the photograph of the Democratic candidate was positioned on the right side and for the other half it was positioned on the left side. Given the counterbalancing of the positions of the photographs, two main experimental versions were created for each study. Participants were randomly assigned to one of the versions. We also randomized the order of presentation of the races. This manipulation is described below.

Procedures for questionnaire studies

Most of the studies, involving the elections for the Senate, were administered in questionnaire sessions. Participants were asked to respond to a number of different and unrelated questionnaires. The questionnaire with the Senate races was embedded in this set of questionnaires. Participants worked individually and were paid \$8 for their participation in the questionnaire session. In all studies, participants were kept naïve with

respect to the objectives of the study. The studies were described as studies on face perception and participants were encouraged to work as quickly as possible and to rely on their “gut instincts” when responding. In addition to the main person judgments, participants were always asked whether they recognized any of the faces. For all analyses reported in the paper, judgments for races in which the participant recognized any of the faces were excluded.

We also manipulated the order of presentation of the pairs of faces. For each election – Senate 2000, Senate 2002, and Senate 2004 – we generated four different random orders. With the counterbalancing of the positions of the pictures, this created eight versions of the questionnaires: 2 (position) X 4 (random order). Participants were randomly assigned to one of the eight versions.

Competence and other trait judgments. In the first study, conducted in May 2003, we used only the Senate races from 2002. The races were divided into two sets, and 114 participants were randomly assigned to one of the sets. For each pair of faces, participants were asked to make seven trait judgments: competence, trustworthiness, honesty, leadership, intelligence, charisma, and likeability. Specifically, they selected the person who they perceived as better on the respective trait (e.g., the more competent person).

We collected competence judgments for the Senate races in both 2000 and 2002 in three different waves. The first wave of data was collected in Fall 2003. One hundred participants were presented with the photographs of the candidates for the Senate and asked to select the person who was more competent. Participants made competence judgments either for the races in 2000 ($n = 50$) or for the races in 2002 ($n = 50$).

Participants recognized very few of the politicians. The maximum number of recognitions was 8 (out of 50 responses) for the Massachusetts race in 2000 (Senator Kennedy was the incumbent). The mean recognition per race was 1.35 (SD = 1.59). Because individual competence judgments for races in which participants recognized any of the faces were excluded, the aggregated competence judgment for each race was based on 42 to 50 individual judgments. In the first wave of data collection, another 100 participants were asked to make attractiveness, face familiarity, or age judgments. These judgments are described below.

The second wave of data was collected in the beginning of 2004. In this study, 127 participants were asked to make thirteen trait judgments per pair of faces. The first judgment was the competence judgment. In addition to this judgment, participants were asked to decide who was more 2) honest, trustworthy; 3) likeable; 4) extraverted, enthusiastic; 5) reserved, quiet; 6) calm, emotionally stable; 7) anxious, easily upset; 8) dependable, self-disciplined; 9) disorganized, careless; 10) sympathetic, warm; 11) critical, quarrelsome; 12) open to new experience, complex; and 13) conventional, uncreative. Competence, trustworthiness, and likeability represented the three main traits identified in the factor solution of the first study with multiple trait judgments per pair of faces. There is a general consensus among personality psychologists that personality can be explained in terms of five global factors: extraversion, neuroticism, conscientiousness, agreeableness, and openness to experience (*S1*). We used a validated scale of 10 traits to measure these five trait dimensions (*S2*): 4 and 5 (reversely scored) for extraversion, 6 (reversely scored) and 7 for neuroticism, 8 and 9 (reversely scored) for conscientiousness,

10 and 11 (reversely scored) for agreeableness, 12 and 13 (reversely scored) for openness to experience.

To reduce the response burden on participants, within each Senate election – 2000 and 2002 - the races were divided into two sets. Thus, we created 4 sets of pairs of faces: two sets for the 2000 races with 15 pairs of faces per set, and two sets for 2002 with 16 and 17 pairs per set respectively. Participants were randomly assigned to one of the four sets (32, 31, 31, and 33 participants per set respectively). The maximum number of recognitions per race was 6 and the mean recognition was 0.56 (SD = 1.33). The aggregated competence judgment for each race was based on 25 to 33 individual judgments.

The third wave of data was collected in May 2004. Seventy-four participants were presented with all races for the Senate in 2000 and 2002 and asked to make a single competence judgment for each of the 63 pairs of faces. The maximum number of recognitions per race was 9 and the mean recognition was 2.79 (SD = 2.50). The aggregated competence judgment for each race was based on 65 to 74 individual judgments. Another 73 participants were asked to express a voting preference for each of the 63 pairs of faces. These judgments are described below.

The competence judgments across the three waves of data collection were highly correlated. The pair-wise correlations were greater than .76, $p < .001$, and a measure of the internal consistency of the judgments, Cronbach's alpha, indicated high reliability of the combined competence judgment, $\alpha = .91$. For these reasons, we computed a mean competence judgment weighted according to the number of individual competence

judgments per pair of faces for each wave of data collection. The predictions for the outcomes of the Senate races in 2000 and 2002 (Table 1) are based on this mean competence judgment. However, it should be noted that each of the three competence judgments predicted the outcomes of the elections for the Senate. Both the judgments from wave 1 and wave 2 data collection predicted the outcomes of 66.7% of the races, $\chi^2 = 7.00$, $p < .008$. The competence judgment from wave 3 predicted 74.6% of the races, $\chi^2 = 15.25$, $p < .001$. The correlation of the judgment with the differences in votes between the candidates was .50 for wave 1, .49 for wave 2, and .55 for wave 3, all $ps < .001$.

The competence judgments for the 2004 Senate elections were collected two weeks before the elections on November 2, 2004. One hundred and twenty-seven participants were asked to make single competence judgments for each of the 32 pairs of faces. The maximum number of recognitions per race was 36 for the Illinois race. The mean recognition was 7.13 (SD = 7.05). The aggregated competence judgment for each race was based on 91 to 127 individual judgments.

Attractiveness, age, and face familiarity judgments. In addition to trait judgments of the candidates for the Senate in 2000 and 2002, we collected judgments of age, attractiveness, and face familiarity. A total of 100 participants were asked to make these judgments. Participants were randomly assigned either to the races in 2000 or the races in 2002. For each pair of faces, 34 participants decided who was more attractive, 24 participants decided who was older, and 42 participants decided whose face was more familiar.

Simulated voting preferences. Another 73 participants were asked to cast hypothetical votes for each of the Senate races in 2000 and 2002. Participants were asked to imagine that this was a political election and to choose the person for whom they would vote. As with all other judgments, if participants recognized any of the candidates in a race, the data for this race were not used in the analyses.

Procedures for computer studies

The studies for the House of Representatives races involved a large number of faces (321 pairs of faces for 2000 and 279 for 2002) and the studies were computerized. We also conducted a computerized study for the Senate races in order to control the rate of presentation of faces and to measure response times for competence judgments. In all three studies, participants were asked to make single competence judgments. As with the questionnaire studies, the position of the faces was counterbalanced and participants were randomly assigned to one of the two experimental versions. In all three studies, the order of the pairs of faces was randomized for each participant by the computer. Participants were kept naïve with respect to the objective of the experiment. The studies were described as studies on face perception and how people assess personality from faces alone. At no point in the study was any mention made of candidates or elections. Participants worked in individual computer booths.

Forty-seven participants made judgments of the candidates competing for the House in 2002, and 41 participants made judgments of the candidates competing in 2004. Each experimental trial consisted of a pair of faces. The photographs had a standard size of 3.2 cm (width) X 4.5 cm (height). The faces were positioned at the center of the screen

with a distance of 4 cm between them (2 cm to the central point of the screen). Each pair of faces was presented on the screen until the participant selected the face that they perceived as more competent. The next trial was presented immediately after the participant's response. At the end of the computer experiment, the experimenter showed a printout of all photographs to the participant and asked them to circle faces that they recognized.

Forty participants participated in the Senate study. In this study, we used all races from 2000, 2002, and 2004 in which the candidates were of the same gender and race. The total number of these races was 68 (24 from 2000, 24 from 2002, and 20 from 2004). Each experimental trial started with a fixation point presented for 500 ms at the center of the screen. Then, the pair of faces was presented for 1 second. The presentation was immediately followed by the participant's judgment of competence. After the competence judgment, participants were asked whether they recognized any of the faces in the pair. The inter-trial interval was 1 second.

Analyses

The unit of analysis in all studies was the individual election race – a state race in the case of the elections for the Senate and a district race in the case of the elections for the House of Representatives. For each race, the individual judgments were aggregated, excluding judgments where participants recognized any of the faces for the respective race. The final judgment could range from 0 to 1, reflecting the proportion of participants who expressed a preference for one of the candidates. For example, if 45 out of 50 participants perceived one of the candidates as more competent, the competence score

was .90, assuming that none of the participants recognized any of the candidates. If more than 50% of participants perceived the winner of the race as more competent, the race was recorded as correctly predicted.

The competence score was also used in correlation and regression analyses as a predictor of the differences in votes between the candidates in a given race. Although only two candidates competed in most of the races, in many races there were more than two candidates. To create a standardized difference score, we used only the votes for the winner and the runner-up. This score was computed as the difference between the votes for the person positioned on the right side in the studies and the votes for the other person divided by the sum of the votes. Because the winner was presented on the right side for half of the trials and on the left side for the other half, the difference score could range from -1 to 1 .

The response times for the competence judgments in the Senate study described above were positively skewed. To remove outliers, we used a liberal procedure of not using response times that were more than 10 standard deviations above the mean. The mean response time across participants was 1181.56 milliseconds with a standard deviation of 284.71 ms. Excluding response times longer than 4028.66 ms (mean + 10 SD) removed response times for only 2.02% of the trials.

Supporting text: Simulation of accuracy of prediction

In order to estimate how the accuracy of prediction of the outcomes of the Senate elections changes as a function of the number of participants making competence judgments from facial appearance, we performed bootstrapping data simulations using two different samples of participants. The first sample consisted of 74 participants who were asked to make single competence judgments for each of the 63 races for the Senate in 2000 and 2002. The second sample consisted of 127 participants who made single competence judgments for each of the 32 Senate races in 2004. In both cases, we repeatedly drew random samples of participants with a fixed size. Starting with a sample size of 10, we drew 50 random samples. For each of the 50 samples, we recorded the percentage of correctly predicted Senate races. We repeated this procedure for larger sample sizes, increasing the sample size by 10 participants at each step (Fig. S1).

The average individual accuracy was better than the chance prediction of 50%, ($M = 59\%$, $SD = 7\%$), $t(73) = 10.58$, $p < .001$, for the predictions for the 2000 and 2002 Senate races, and ($M = 53\%$, $SD = 10\%$), $t(126) = 3.58$, $p < .001$, for the predictions for the 2004 races. As shown in Fig. S1, increasing the sample size to 40 participants substantially increased the accuracy of prediction over the average individual accuracy, $t(49) = 26.29$, $p < .001$, for the 2000 and 2002 races, and $t(49) = 19.46$, $p < .001$, for the 2004 races. However, the benefit of increased sample size diminished after this point. In fact, for both simulations – 2000/02 and 2004 races - the accuracy of prediction was significantly improved with the increase of the sample size from 30 to 40 participants, $t(49) = 2.82$, $p < .007$, and $t(49) = 2.09$, $p < .042$, respectively; but did not improve

significantly with the increase of the sample size from 40 to 50 participants, $t = 1.05$, $p = .30$, and $t < 1$ respectively.

To capture the diminishing benefit of the increased sample size for the accuracy of prediction, we modeled the accuracy as an inverse function of the sample size. For the 2000 and 2002 data, this function (1) accounted for 98.9% of the variance. For the 2004 data, the function (2) accounted for 94.1% of the variance.

$$(1) \text{ Accuracy (\%)} = 74.65 - 140.28 / n$$

$$(2) \text{ Accuracy (\%)} = 67.17 - 145.14 / n$$

Although the coefficients were slightly different, in both models increasing the sample size from 10 to 40 participants increases the accuracy by more than 10%. At the same time, increasing the sample size from 40 to 100 participants increases the accuracy by slightly more than 2%.

Figure S1. Data simulation of accuracy of prediction of outcomes of the races for the US Senate as a function of sample size for **A)** 2000 and 2002 Senate races and **B)** 2004 Senate races. The plots show the means of the 50 sample means for each sample size with their corresponding standard errors.

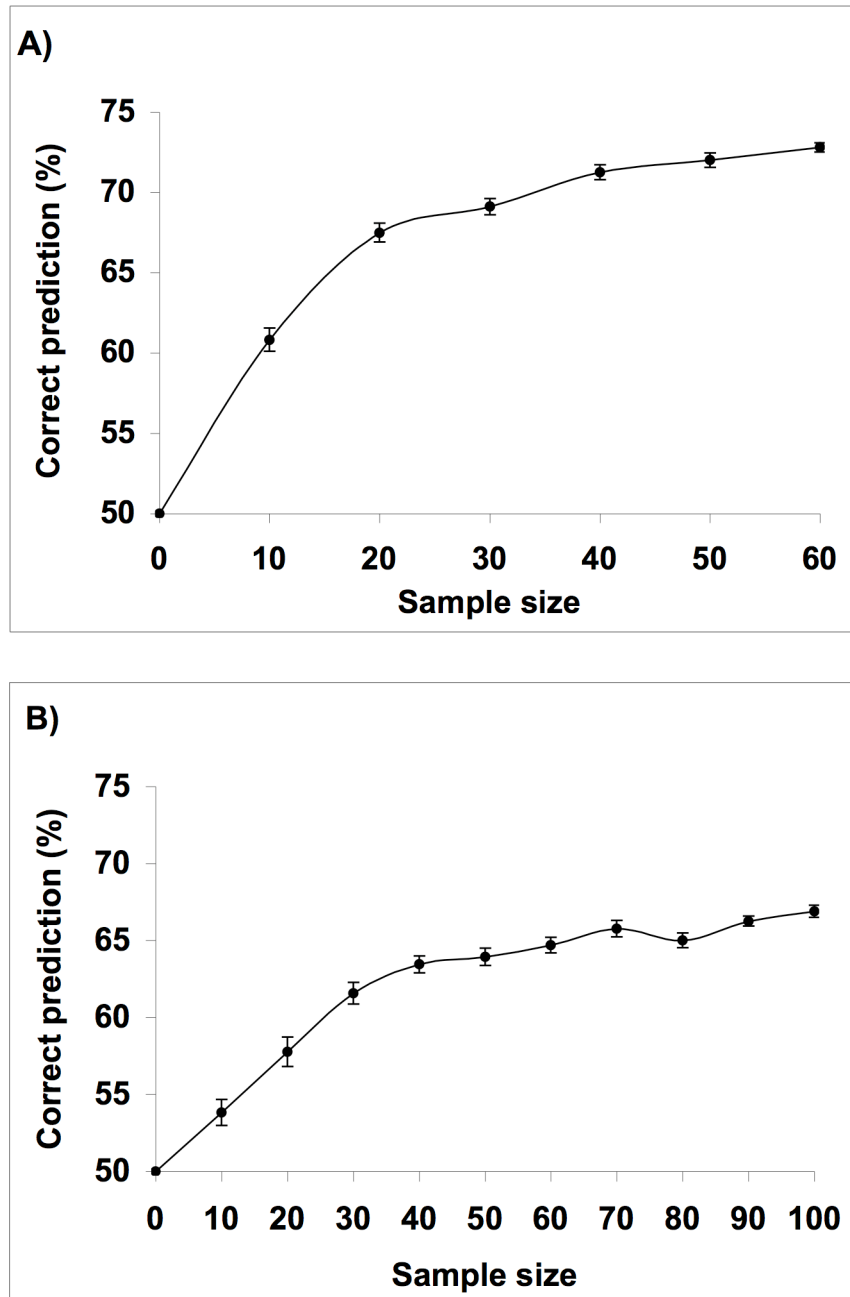


Table S1. Factor loadings of trait judgments of candidates competing for the US Senate in 2002 on factors identified in a principal components analysis with a Varimax rotation. The factor analysis was performed on the aggregated judgments for each trait at the level of the Senate races.

Traits	Factor solution		
	Competence	Trustworthiness	Likeability
Competence	.90	.33	-.20
Intelligence	.89	.21	-.26
Leadership	.82	-.34	.36
Honesty	.05	.98	.03
Trustworthiness	.16	.96	-.01
Charisma	-.05	-.14	.98
Likeability	-.12	.17	.96
Accounted variance	33.0%	31.4%	30.3%

References

- S1. O. P. John, S. Srivastava, in *Handbook of Personality: Theory and Research*, L. A. Pervin, O. P. John, Eds. (Guilford, New York, 1999), pp. 102-138.
- S2. S. D. Gosling, P. J. Rentfrow, W. B. Swan, Jr., *J. Res. Pers.*, **37**, 504 (2003).