

Questions & Answers on the Harvard Visual Voting Study Benjamin and Shapiro (2006)

How can the study draw reliable conclusions from a sample of 264 Harvard undergraduates? Does this sample's high level of educational attainment skew the results? How is it possible to generalize based on results from this sample about a broader voter population with more variations in race, economic, age, education, and political leaning?

The sample for the experiment is not representative of the American voters: the participants tended to be young, liberal, well-off economically, and highly educated. That makes it *especially* surprising that their reactions to the candidates were able to predict the outcomes so well. It's possible (though we don't yet know for sure) that the reactions of a more representative sample could be even more highly predictive of the election outcomes. The study found no evidence that some types of people are better forecasters than others, which may reflect that the identity of the forecaster is not very important.

Was the sample size big enough?

The sample of 264 students is actually much larger than the typical experiment. However, the more relevant sample size is the number of elections – 58 in all. This was nearly every gubernatorial election for which C-SPAN offered DVD coverage of a debate. It is possible that C-SPAN coverage over-samples competitive elections, but an analysis of the sample found no evidence of this bias.

How were the 10-second clips selected? How did the study account for differences in settings, wardrobe, lighting, physical position (stationary v. moving)? How did it account for the difference between a given 10 seconds (versus the previous or next 10 secs), as well as the other variables of video?

Three random 10-second-long clips were drawn for each candidate. If a clip showed more than one candidate or showed the candidate's name or party or the name of the state on the screen, the clip was dropped and a new 10-second sample was drawn to replace it, until there were three clips for each candidate. Drawing clips from the same debate controlled for setting, lighting, and image quality. Taking several samples of each candidate was meant to help control for differences like when the candidates were stationary vs. moving. Individual clips did of course vary in how well the candidate came across. The study authors estimated that about 60% of the variation in subjects' forecasts of who would win was attributable to differences in which clips the subjects viewed.

Why were 10-second clips used? Why not shorter clips such as those used in other studies from the thin-slicing literature (which can be as short as one second or less)? Or why not use longer, more representative samples?

There is a large psychological literature (some of it summarized in Malcolm Gladwell's book *Blink*) showing that judgments about other people from "thin slices" – exposures to

expressive behavior as brief as a few seconds – tend to be highly predictive of reactions to much longer exposures. That was the basis for the study’s authors’ and sponsors’ expectation that 10-second long clips would contain enough information to be predictive of the election outcomes, and that suggests the feasibility of using this methodology to predict election outcomes in advance. The length of 10 seconds was chosen to be similar to the length of segment used in thin-slice studies such as Ambady and Rosenthal (1993).

Is the visual factor measured in the study is more predictive than campaign spending or incumbency?

In the study’s sample, candidates’ visual self-presentation is less predictive than campaign spending and about as predictive as incumbency. The study may understate the predictive power of candidates’ visual self-presentation relative to these other factors for a number of reasons, however. For one, candidates’ visual self-presentation likely contributes to candidates’ ability to raise funds and to be an incumbent in the first place, creating a double-counting in campaign spending’s predictive value. Some trends in the analysis also suggested that were more subjects added to the study, candidates’ visual self-presentation would have been more predictive than campaign spending. (See the Todorov question below for additional discussion.)

Were there any differences in predictive performance between male and female raters? Conservative or liberal raters? Raters of different races?

The study found no differences. There is no evidence in the data that some individuals are better than others at forecasting, or that certain demographic characteristics are associated with higher accuracy.

Were there any differences with the raters’ predictive effectiveness when viewing male as versus female candidates? Conservative versus liberal candidates? Candidates of different races?

The study found no differences, though at the study’s sample size these questions can be addressed with only limited precision.

How did the better recognized candidates do versus the more obscure ones? What if people had seen candidates before but did not remember it?

When people recognize a candidate, not surprisingly they tend to do better at predicting the winner of the election. But raters who didn’t recognize even better-known candidates were no better at predicting elections with those more recognizable candidates. This suggests that subconscious recognition was not a major factor in people’s judgments.

Were there specific factors or characteristics of elections where the raters were especially good at predicting outcomes? Or especially bad at predicting outcomes (i.e. huge money differential)?

The study found no evidence of such differences, though it only used 58 elections, and a larger sample might reveal such differences.

Does this study mean the better-looking candidate wins?

No. While physical attractiveness is correlated with the election outcome, physical attractiveness is much less predictive than the study participants' forecasts based on candidates' visual self-presentation. Appearing to be likeable and appearing to be a good leader are also correlated with the outcome (good leadership more than physical attractiveness), but candidates' total visual self-presentation is the strongest predictor. That suggests that other factors besides physical attractiveness play a major role, and that asking subjects to forecast the outcome picks up these other factors.

Did you look at what kinds of body language wins – i.e. smiling, or good posture?

No. Specific characterizations and ratings of the kinds of non-verbal behavior displayed by the candidates were beyond the scope of this study.

Does the taller candidate usually win? Could the raters tell which candidate was taller?

There was a tendency for taller candidates to do better in the study's sample, earning 2 percentage points higher vote share, on average. However, controlling for height has little effect on the core visual measure. This is not surprising, since it is generally difficult for raters to tell which candidate is taller from these clips. None of the clips showed the two candidates standing next to each other; all showed only one candidate at a time.

Does this study measure (relative) charisma? Did the raters tell you what they were looking for in a candidate?

This study measures how well subjects can forecast election outcomes based purely on candidates' appearance and nonverbal cues. Candidates' visual self-presentation may be likened to common understanding of the term "charisma," but the subjects did not report exactly what they were looking for in making their forecasts.

Could they tell which candidate was from which party?

Not when the clips lacked clear sound. In the silent and muddled conditions, respectively, 52% and 48% of participants correctly guessed political party, which are not statistically different from each other or from random guessing (see page 12 of the study).

What elections were you showing to the raters? In the elections you used, did the democrats or the republicans usually win?

The sample was chosen based solely on the availability of C-SPAN DVDs of gubernatorial debates. Democrats win about 40 percent of the elections in our sample.

Todorov et al did a similar study that was published in Science. How does the present study compare?

The study most closely related to this one, Todorov et. al. (2005), has independently shown that ratings of competence based on photographs of House and Senate candidates predict election outcomes and vote shares. This study differs in several important ways, and is complimentary in other ways.

Most importantly, this study's regression framework allowed assessments of the incremental predictive power of candidates' visual self-presentation, after accounting for economic and political predictors of electoral success, and for comparisons the relative predictive power of these factors. An important finding of the present study is that candidates' visual self-presentation has substantial predictive power above and beyond these more typical variables, indicating among other things that the methodology of rating the candidates has practical value for forecasting.

The present study also found that raters are unable to predict election outcomes when the clips contain sound. This suggests the importance of having naïve raters in practical application.

The present study also used a different unit of analysis than Todorov et. al. (2005) and other prior work in the literature. Use of video clips from candidate debates allowed comparison of live, unedited images of the candidates of equal quality, whereas the use of still photographs supplied by the campaigns (as in Todorov et. al.) introduces extra variables under the control of the broader campaign, including image quality, the number of slightly different poses the candidate was photographed in before the final, official photograph was chosen, and the aesthetic and political judgment of whomever made that choice.

A further difference is that Todorov et. al. (2005) asked study participants to make judgments of candidates' competence, which predicted election outcomes, while we ask raters directly who they think is more likely to win (the Todorov study did not ask this question). The relationship between these two measures is not entirely clear.

Interestingly, the judgments of raters in Todorov et. al. (2005) as to which candidate appeared to be more competent were more predictive of the outcomes of the elections in that study than the raters' predictions were of the outcomes of the elections in the present study. Since both studies shared as their core visual stimulus an image of each candidate that showed non-verbal behavior and appearance, this higher prediction rate suggests that a different combination of visual stimulus, length of exposure to raters, and questions might find candidates' visual self-presentation to be even more predictive of election outcomes than has yet been established.

So could this sort of testing be used to see who would do well in an election ahead of time?

The results of the present study provide reason to expect that a similar methodology could be used to help predict election outcomes ahead of time, provided that the raters used to make such predictive judgments are unfamiliar with the candidates evaluated.