

# Trust the Polls? Neural and Recall Responses Provide Alternative Predictors of Political Outcomes

Samuel B. Barnett, Northwestern University, USA

Andrés Campero, MIT Media Lab, USA

Ronen Zilberman, MIT Media Lab, USA

Chris Rose, New York University, USA

Aaron Robinson, Northwestern University, USA

Moran Cerf, MIT Media Lab, USA

## ABSTRACT

Given the lack of public trust in opinion polls, we propose novel approaches to supplement existing methods of predicting political outcomes. Neural similarity and recall of candidates in a televised presidential primary debate accurately forecasted election results. Additionally, the neural data predicted changes in the opinion polls following the debate.

## INTRODUCTION

Even though most voters do not trust public opinion polling (Jackson & Sparks 2017), what alternatives do we have? For most of human history, we could only know what others were thinking by asking them to report it. This subjective account, even when given in earnest, suffers from many shortcomings and biases (Nisbett & Wilson 1977, Griffin & Hauser 1993), and these issues recently gained notoriety as political opinion polls predominantly failed to foresee the United Kingdom's vote to leave the European Union (the "Brexit") and the election of Donald Trump as President of the United States. Complicating matters further, scholars have linked political preferences to irrelevant events occurring close to election day (Healy et al. 2010), unconscious feelings about the physical appearance of the candidate (Todorov et al. 2005), and even our genetics (Fowler & Dawes 2008). Thus, decoding political preferences is a challenging task requiring further research (Mauser & Kopel 1992, Bartels 2002, Lakoff 2008, Hall et al. 2013).

Consumer research, neuroscience, and other fields have provided many methods to passively assess preferences, emotions, and decision-making processes (Berka et al. 2007, Levy et al. 2011, Jost & Amodio 2012, Ely et al. 2015). Given that political campaigns spend billions of dollars on messaging, there is growing interest among practitioners in measuring the effectiveness of these techniques (Aron 2007). Recent work on attention and mind-wandering (Mason et al. 2007) has shown that content that can recruit similar brain activity (e.g., activate the same regions or circuits simultaneously) is associated with engagement (Barnett & Cerf 2015, Hasson et al. 2008). Simply put, a measure of engagement can be derived by assessing the similarity between multiple brains (Hasson et al. 2004).

We measured responses to a live televised presidential primary debate in a natural environment (i.e., a large auditorium, as opposed to a traditional laboratory setting). Unlike prior studies on engagement that focused on neutral or positive stimuli (Hasson et al. 2008, Boksem & Smidts 2015, Barnett & Cerf 2017), we studied reactions to candidates with whom the participants generally disagreed. Also, in general, politics can be controversial, generating strong emotional responses among individuals (Haidt 2012). We theorize that elevated neural similarity will reflect collective, engaged disagreement with the content. When participants already predisposed to opposing the candidates exhibit especially strong, aversive engagement, we hypothesize that the population that prefers those candidates will also be more engaged in voting for them. For example, if Trump makes Democrats' brains more alike—perhaps driven by shared aversion—

our hypothesis suggests that Trump will also be more memorable to Republican brains and ultimately increase his likelihood of earning their votes.

We test neural similarity, survey recall and its sentiment, and continuous ratings of subjective agreement as predictors of both public polling changes and population-level voting results. Unlike polls, which typically interview over a thousand individuals, we present these refined techniques that accurately extrapolate from just dozens to predict the behavior of tens of millions. We report the strength of these predictions and also combine the neural and subjective data to form multivariate model estimates.

## METHODS

### Participants

Twenty-nine participants (7 female) viewed the January 14, 2016 Republican presidential primary debate in a large auditorium. Participants were asked to refrain from talking and excessive movement during the viewing and were paid \$60 for their time.

To select participants who would likely disagree with some or all of the presented content, we screened for candidate preference with a pre-survey. Immediately prior to the debate, the three leading Republican presidential candidates, according to opinion polls, were (in descending poll rank) Trump, Cruz, and Rubio, collectively representing nearly two thirds (65.7%) of the sample population; thus, we only accepted participants that preferred other candidates (i.e., not Trump, Cruz, or Rubio).

### Subjective and Neural Data Collection

On the day of the debate, the participants answered a pre-debate survey regarding demographic information as well as their detailed opinions on each individual candidate and a variety of selected topics (e.g., abortion, gun control, taxes).

Eleven participants were fitted with high performance EEG systems (BrainVision, Morrisville, NC) to measure their neural activity. Sampling was performed at a rate of 250 Hz from sixteen channels, high-pass filtered at 1 Hz, low-pass filtered at 50 Hz, and processed via Independent Component Analysis (ICA) (Delorme et al. 2007, Hyvärinen & Oja 2000) using EEGLAB in MATLAB (Delorme & Makeig 2004). Additionally, for redundancy, headband EEG systems (Muse, Toronto, Canada) with a lower sampling rate and fewer channels were assigned to ten additional participants; however, none of the main acquisition systems failed, so only the BrainVision data were ultimately used.

All of the participants used an online application to continuously report their agreement/disagreement with the content throughout the debate ("How much do you agree with what is being said?" presented on a sliding scale as in many focus groups; cf., Kamberelis & Dimitriadis 2005). Immediately following the debate, a post-debate survey was administered, which assessed political views and candidate recall. Two research assistants independently coded the recall sentiments for each candidate (e.g., favorable/unfavorable).

**Calculation of Neural Similarity**

We calculated neural similarity across every pairwise combination of participants as performed in related techniques with both fMRI (Hasson et al. 2004, 2008) and EEG data (Dmochowski et al. 2012, 2015; Barnett & Cerf 2015, 2017). We used a Fast Fourier Transform (FFT) to select neural data in the alpha frequency spectrum (7.5-12.5 Hz), which has been linked to attentional modulation and engagement (Barnett & Cerf 2015, Boksem & Smitsds 2015, Falk et al. 2012, Ki et al. 2016). We determined the maximally correlated pair of components for each pair of participants (Dmochowski et al. 2012, 2015), and then averaged the correlation levels across all participant pairs to obtain a single value for a 30-second window. We repeated the process for each successive second (i.e., rolling 30-second windows) to obtain a time series of neural similarity throughout the debate.

An underlying assumption with measures of neural similarity is that the relevant correlations are due the simultaneous exposure of a stimulus to multiple brains. If the assumption holds, then data correctly synchronized across subjects would be statistically different than randomly aligned data. We tested this assumption with bootstrapping; we correlated 100-second segments of raw EEG data, again pairwise across participants, but with a random starting point for each subject (i.e., misaligning the data). We found that indeed the assumption holds: the real neural similarity value was significantly higher (1.2 standard deviations, SDs; Z-test:  $p < .01$ ) than the mean of the values computed for 1,000 randomly shuffled segments.

The neural similarity that we attribute to a given candidate (or the moderators or commercials) is assessed in a related way. We compared the correctly aligned data with 1,000 random shuffles (matched in duration); higher z-scores reflected greater neural similarity generated by that candidate.

**RESULTS**

Subjective and neural responses varied significantly over the course of the debate (144 minutes). We analyzed moment-to-moment responses to each of the seven candidates, the two moderators grouped together, and the television commercials. We also studied survey data before and after the debate as well as public opinion polling (aggregated by HuffPost Pollster, 2016 National Republican Primary) and population-level vote totals in the primary elections.

**Engagement with Candidates vs. Other Stimuli**

Since a political debate centers around the candidates, we expected that they would induce greater engagement than the neutral moderators. Indeed, each of the candidates generated higher neural similarity (by .15-2.54 SDs) than the moderators. However, the candidates produced significantly less neural similarity (by 2.42-4.81 SDs) than the comparatively non-controversial and widely engaging television commercials, which were 4.96 SDs more engaging than the moderators.

**Forecasting Changes in Public Opinion Polls**

From the day before the debate (January 13) through the first primary contest (February 1), three of the seven candidates rose in the public polls ( $1.57\% \pm 1.42\%$ ), while the other four fell slightly ( $-.65\% \pm .44\%$ ). On average, a candidate's poll support changed  $0.30\% \pm 1.47\%$ .

Neural similarity during each candidate's debate answers forecasted these changes in popular opinion levels ( $r = .67, p = .10$ ). Additionally, the two candidates yielding the highest neural similarity (Rubio and Cruz) showed greatest improvement in public support (3.20% and 0.70%, respectively). However, unlike neural similarity, none of the subjective responses were correlated with with subsequent changes in the candidates' poll standings ( $p > .70$ ; see Table 1).

**Forecasting Election Outcomes**

Over the course of the Republican presidential primaries (February-May 2016), over 30 million votes were cast for the seven candidates analyzed in this study. Trump's rival candidates suspended their campaigns (effectively dropping out of the race) at various times. To normalize by campaign length, we divided a candidate's total votes received by the number of days after the debate (January 14) to the date he suspended his bid for the nomination (or, in Trump's case, when he had no more rivals after Kasich suspended his campaign on May 4). Christie received the fewest average votes per day of active campaigning (2,135) whereas Trump received the most (126,270). The typical candidate received  $45,813 \pm 43,748$  votes per active campaign day.

*Neural Similarity and Voting Results*

Neural similarity during the debate was strongly predictive ( $r = .79, p = .03$ ; see Table 1) of votes received in the first primary contest (Iowa caucuses, February 1). Additionally, neural similarity was positively linked to candidates' overall future campaign performance ( $r = .37, p = .41$ ). Statistical significance was reduced by Trump's vote total being a distinct outlier. Excluding Trump, neural similarity was highly correlated with average votes per active campaign day ( $r = .78, p = .07$ ; see Table 1) over the entire primary season.

*Subjective Responses and Voting Results*

Participant recall of candidates following the debate was closely linked ( $r = .83, p = .02$ ) with eventual voting results throughout the primary campaign. Partitioning the recall data by subjective valence, we observe that this result is predominantly driven by unfavorable recall ( $r = .91, p < .01$ ; see Table 1) as opposed to favorable memory of the candidates ( $r = -.09, p > .80$ ; see Table 1). To wit, aversive memorability among participants who oppose the candidates reliably forecasts population-level voting outcomes.

Furthermore, recall was positively correlated with votes received in the first primary contest ( $r = .69, p = .08$ ), albeit to a lesser degree than its correlation with the aggregate results over all pri-

**Table 1: Correlations of Various Predictors and Political Outcomes.**

Significant predictions were asterisked: \*, \*\*, and \*\*\* denote  $p$  less than .10, .05, and .01, respectively. Trump's data point was removed as an outlier in one correlation analysis (†).

Predictor	Collection Period	Correlations with Predictor		
		Change in Poll Standings	Results of Iowa Caucuses	Votes Per Active Campaign Day
Neural Similarity	During Debate	.67*	.79**	.78*†
Subjective Disapproval		.06	.60	.44
Total Recall	Following Debate	.00	.69*	.83**
Favorable Recall		.17	-.23	.03
Unfavorable Recall		-.13	.87**	.91***

maries. Again, we attribute this finding to the level of unfavorable recall ( $r = .87, p = .01$ ; see Table 1) versus favorable recall, which was not correlated ( $r = -.23, p > .60$ ; see Table 1) with the results of the Iowa caucuses. Similarly, real-time subjective disapproval of the candidates was associated with higher vote totals on February 1 ( $r = .60, p = .15$ ; see Table 1) and on average throughout the campaign ( $r = .44, p = .33$ ; see Table 1).

**Combining Metrics for Improved Forecasts**

Multivariate models achieved even greater predictive power over voting outcomes (see Figure 1). We performed linear regressions of this form:

$$[Expected\ Votes] = \beta_0 + \beta_1 \times [Unfavorable\ Recall\ (\%)] \dots + \beta_2 \times [Relative\ Neural\ Similarity\ vs.\ Moderators\ (SDs)]$$

This model, which combines subjective and neural data, predicted the results of the Iowa caucuses with notable accuracy ( $r = .98, p < .01$ ;  $\beta_0 \approx 18,388, \beta_1 \approx 1,706, \beta_2 \approx 11,364$ ; see red circles in Figure 1). The regression of the same form was also predictive of votes per day during the primary season ( $r = .91, p < .01$ ;  $\beta_0 \approx 5,425, \beta_1 \approx 4,896, \beta_2 \approx -1,490$ ; see blue diamonds in Figure 1). Note that the model for the Iowa caucuses is weighted heavily by neural similarity whereas the model for the full primary season is more dependent on unfavorable recall.

**DISCUSSION**

We found alternative predictors of three examples of political outcomes: (1) changes in poll standings following the debate, (2) votes received in the Iowa caucuses, and (3) average votes received per active campaign day, which each had a different time scale (on the order of days, weeks, and months, respectively). The recall measures, in particular, improved with longer time; aggregate recall was not at all correlated ( $r = .00, p \approx 1.00$ ) with poll changes, positively associated ( $r = .69, p = .08$ ) with Iowa results, and highly predictive of daily votes over months ( $r = .83, p = .02$ ). Since neural similarity was closely linked ( $r > .67, p < .10$ ; see Table 1) with all three of the political outcomes, its relative effectiveness over the self-report data was greatest in predictions immediately following the debate (e.g., the changes in public opinion polls).

Most of our analysis focuses on candidate-versus-candidate election predictions, but the observation that commercials generated significantly higher neural similarity (more than 2 SDs above the candidates and moderators) is worthy of future study. Of course, commercials are designed to appeal to common denominators and promote consensus around an advertised product, whereas a debate, by definition, presents conflicting viewpoints.

Additionally, in prior studies, neural similarity was linked to collective engagement with neutral or pleasing stimuli, such as movies and advertisements, but our results with contentious content suggest that, regardless of valence, neural similarity is a meaningful measure of engagement. In particular, strong, shared brain responses and self-reported feelings on one side of the political spectrum corresponded with similarly strong behavioral responses on the other side in the form of polling support and votes.

**Limitations**

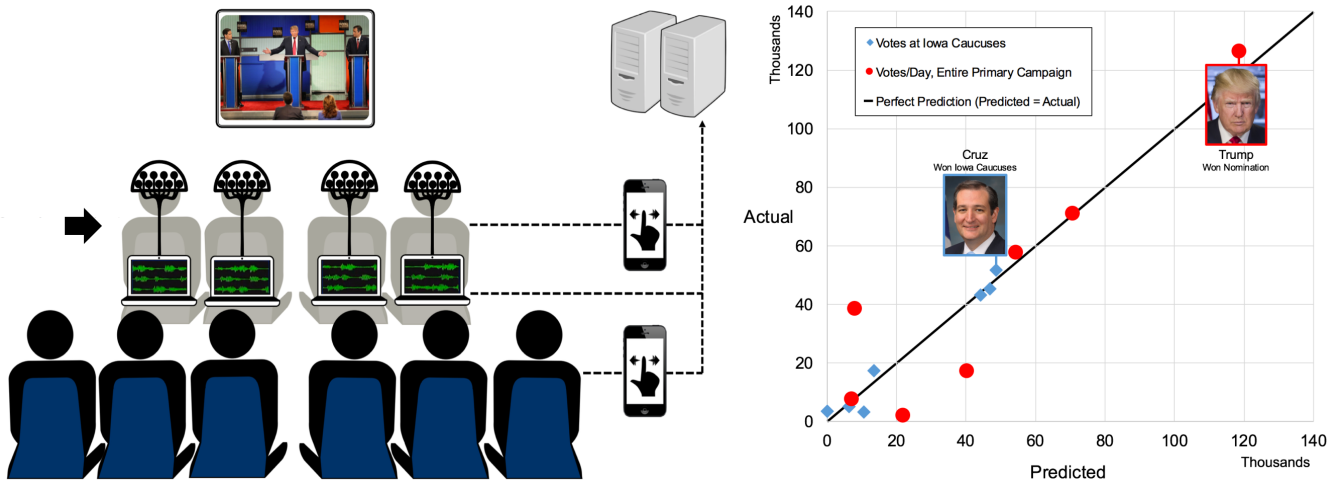
Our use of a live debate enhanced the naturalistic aspects of the participants' experience, but it also meant that we had no control of the stimulus. While seven candidates in a primary debate is a large number in its political context, it is a small number of data points. Similarly, free recall responses are true-to-life, but vary with effort and memory abilities. Furthermore, these responses required manual coding, and although this was performed independently by two individuals, the process introduces subjectivity.

**CONCLUSION**

Elections impact us profoundly, which is why we still pay attention to opinion polls despite not trusting their predictive power. This work demonstrates that certain neural and recall responses can serve as additional predictors of political outcomes. Voters, candidates, policymakers, and consumer researchers all stand to benefit from an increasingly robust set of election forecasting tools, and our proposed techniques contribute to that end.

**REFERENCES**

Aron, A. et al., 2007. *Politics and the brain*, s.l.: New York Times.  
 Barnett, S. B. & Cerf, M., 2015. Connecting on movie night? Neural measures of engagement differ by gender. *Advances in Consumer Research*, Volume 43.



**Figure 1: Neural Similarity and Unfavorable Recall During Televised Debate Collectively Predict Election Outcomes.**

Left: Schematic of experimental design. All participants provided real-time subjective ratings while viewing the debate; a subset (denoted by an arrow) of these also underwent EEG recordings. Right: Actual versus predicted votes for each candidate. Blue diamonds illustrate performance in the Iowa caucuses and red circles depict the full primary results (normalized by each candidate's campaign length). Prediction models achieved high accuracy ( $r > .90, p < .01$ ).

- Barnett, S.B. and Cerf, M., 2017. A Ticket for Your Thoughts: Method for Predicting Content Recall and Sales Using Neural Similarity of Moviegoers. *Journal of Consumer Research*, 44(1), pp.160-181.
- Bartels, L. M., 2002. Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, Volume 24.2, pp. 117-150.
- Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D. and Craven, P.L., 2007. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, 78(5), pp.B231-B244.
- Boksem, M. A. & Smidts, A., 2015. Brain responses to movie trailers predict individual preferences for movies and their population-wide commercial success. *Journal of Marketing Research*, 52(4), pp. 482-492.
- Delorme, A. and Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1), pp.9-21.
- Delorme, A., Sejnowski, T. & Makeig, S., 2007. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage*, 34(4), pp. 1443-1449.
- Dmochowski, J. P. et al., 2015. Audience preferences are predicted by temporal reliability of neural processing. *Nature Communications*, Volume 5.
- Dmochowski, J. P., Sajda, P., Dias, J. & Parra, L. C., 2012. Correlated components of ongoing EEG point to emotionally laden attention-A possible marker of engagement?. *Frontiers in Human Neuroscience*, 6(112).
- Ely, J., Frankel, A. & Kamenica, E., 2015. Suspense and surprise. *Journal of Political Economy*, 123(1), pp. 215-260.
- Falk, E. B., Berkman, E. T. & Lieberman, M. D., 2012. From neural responses to population behavior neural focus group predicts population-level media effects. *Psychological Science*, 23(5), pp. 439-445.
- Fowler, J.H. and Dawes, C.T., 2008. Two genes predict voter turnout. *The Journal of Politics*, 70(3), pp.579-594.
- Griffin, A. & Hauser, J. R., 1993. The voice of the customer. *Marketing Science*, 12(1), pp. 1-27.
- Haidt, J., 2012. *The righteous mind: Why good people are divided by politics and religion*. s.l.:Vintage.
- Hall, L. et al., 2013. How the polls can be both spot on and dead wrong: Using choice blindness to shift political attitudes and voter intentions. *PloS One*, 8(4), p. 360554.
- Hasson, U., Furman, O., Clark, D., Dudai, Y. and Davachi, L., 2008. Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron*, 57(3), pp.452-462.
- Hasson, U. et al., 2004. Intersubject synchronization of cortical activity during natural vision. *Science*, Volume 303.5664, pp. 1634-1640.
- Healy, A. J., Malhotra, N. & Mo, C. H., 2010. Irrelevant events affect voters' evaluations of government performance. *Proceedings of the National Academy of Sciences*, 107(29), pp. 12804-12809.
- HuffPost Pollster, 2016. National Republican Primary. *HuffPost*. <https://elections.huffingtonpost.com/pollster/2016-national-gop-primary> [Accessed 23 March 2018].
- Hyvärinen, A. & Oja, E., 2000. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4), pp. 411-430.
- Jackson, N. & Sparks, G., 2017. Most Americans Don't Trust Public Opinion Polls. *HuffPost*. [https://www.huffingtonpost.com/entry/most-americans-dont-trust-public-opinion-polls\\_us\\_58de94ece4b0ba359594a708](https://www.huffingtonpost.com/entry/most-americans-dont-trust-public-opinion-polls_us_58de94ece4b0ba359594a708) [Accessed 23 March 2018].
- Jost, J. T. & Amodio, D. M., 2012. Political ideology as motivated social cognition: Behavioral and neuroscientific evidence. *Motivation and Emotion*, 36(1), pp. 55-64.
- Kamberelis, G. & Dimitriadis, G., 2005. Focus groups. *The Sage Handbook of Qualitative Research-Third Edition*, Volume 3, pp. 887-908.
- Ki, J. J., Kelly, S. P. & Parra, L. C., 2016. Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. *The Journal of Neuroscience*, 36(10), pp. 3092-3101.
- Lakoff, G., 2008. *The political mind: A cognitive scientist's guide to your brain and its politics*. Penguin.
- Levy, I., Lazzaro, S.C., Rutledge, R.B. and Glimcher, P.W., 2011. Choice from non-choice: predicting consumer preferences from blood oxygenation level-dependent signals obtained during passive viewing. *Journal of neuroscience*, 31(1), pp.118-125.
- Mason, M. et al., 2007. Wandering minds: the default network and stimulus-independent thought. *Science*, 315(5810), pp. 393-395.
- Mauser, G. A. & Kopel, D. B., 1992. 'Sorry, wrong number': Why media polls on gun control are often unreliable. *Political Communications*, 9(2), pp. 69-91.
- Nisbett, R. E. & Wilson, T. D., 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), p. 231.
- Todorov, A., Mandisodza, A. N., Goren, A. & Hall, C. C., 2005. Inferences of competence from faces predict election outcomes. *Science*, 308(5728), pp. 1623-1626.