Supporting Information

Contents

A Data Sources

We list the sources of our data (which we merged with the Gallup individual-level data) in Table S1 below. All sources are public with the exception of the Gallup data itself, which we obtained via our institution with geolocated respondents.

Variable	**Source**
Unemployment rate	BLS
Labor force participation rate	BLS
Total employment	QCEW
Wages	QCEW
Death rates	CDC
Life expectancy	IHME*
Crime / arrests	UCR (FBI)

*Note: *Institute for Health Metrics and Evaluation*

Table S1: Contextual variables used as (potential) predictors of answers to questions about life, emotions, and the economy.

B Variable Importance

B.1 Random forest

Our approach to estimating the importance of different predictors uses a well-developed measure of variable importance based on permutation tests. The underlying method is to:

- First, predict an outcome with the full data as accurately as possible.
- Second, break the empirical relationship between the outcome and one of the predictors by randomly reshuffling its values ("permuting" it), and then re-estimate the relationship with the same model.
- Third, record the difference between the prediction error from the model run on the true data and the prediction error obtained from the data with the permuted column.

Permutation analyses can be run with any model used in the first step. In our application, we rely on a random forest (Wright and Ziegler, 2017) approach to stay agnostic about the functional form by which our combination of individual-level and geography-level predictors may influence the outcomes of interest. Of course, a natural question is how good the "best" model is performing. Here, we visualize variation in this answer across two dimensions. First, we demonstrate that some questions are harder to model than others, at least with the covariates that we have at our disposal. Second, we demonstrate that the ability to predict the public's attitudes on a given question vary over time.

Figure S1 illustrates the average prediction error over 100 bootstrapped draws of the data, using the random forest on the full data. The left facet displays the raw mean absolute error, while the right facet normalizes this value by the empirical standard deviation of the outcome. As illustrated, it is relatively easy to predict presidential approval while it is relatively difficult to predict an individual's placement of themselves on a "life ladder", or whether they worry about the amount of money they have. These patterns are consistent with a notion of "well-behaved" attitudes, by which we refer to the organization of attitudes along easily-observable covariates. In the case of presidential approval, attitudes are exceptionally well-behaved along partisanship, allowing models to obtained greater predictive accuracy compared to less well-behaved attitudes such as subjective placement on a life ladder, for which it is less clear which combinations of covariates might best improve predictive accuracy.

Figure S2 displays the variation in prediction error for an individual's assessment of whether the national economic conditions are improving. Here we find striking evidence of the increasing difficulty of predicting attitudes on the economy as we move further away from the Global Financial Crisis. A likely explanation for these trends is that predicting attitudes on the economy is easier during periods when economic conditions are especially salient, but grows more challenging as this salience recedes.

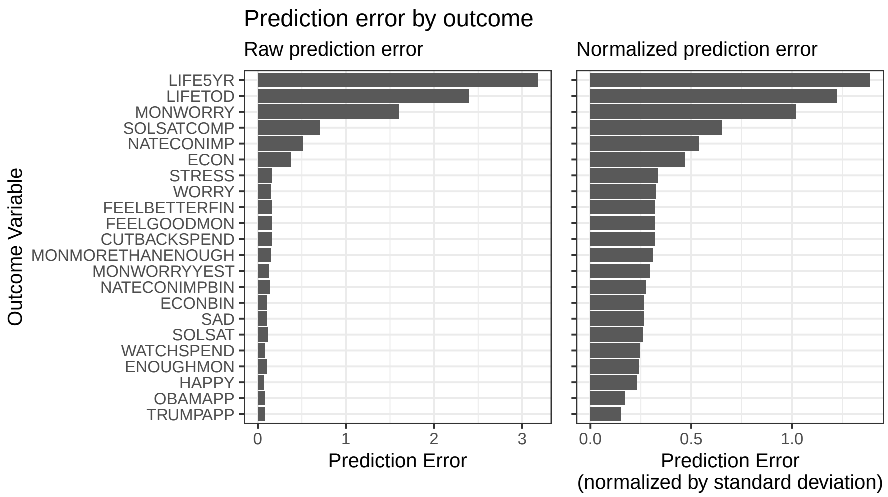

Figure S1: Raw (left) and normalized (right) prediction errors by outcome variable.

Note, however, that the ability to estimate variable importance is not sensitive to these overtime or across opinion variations. This is because the approach treats each outcome in isolation, and each period as its own dataset. Within a given outcome and month, we then apply the permutation test described above, holding constant these differences across outcomes and periods.

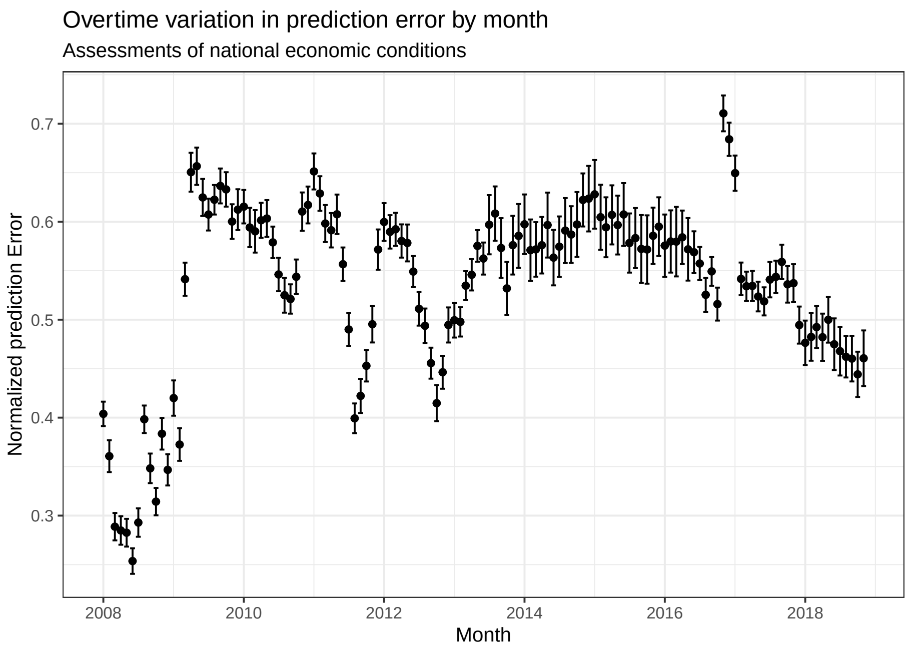

Figure S2: Errors for predicting assessments of whether the national economy is improving normalized by standard deviation (y-axis) estimated by month (x-axis). Bars indicate two standard deviations of the normalized measure.

B.2 Alternative Measures of Variable Importance

Our main results rely on permutation tests of variable importance using a random forest. Below, we confirm our findings are robust to alternative approaches to evaluating which measures are most prognostic, including LASSO regressions and expected percentage reduction in error (ePRE, Herron 1999), the latter of which is very similar in spirit to the permutation tests of our main results.

Starting with the LASSO approach, we visualize two questions from 2016 in Figure S3. In the left panel, we demonstrate that the most prognostic covariates of an individual's satisfaction with their standard of living include their income, marital status, educational attainment, and age. In the right panel, we find that their views of the national economy are far more strongly associated with their partisanship.

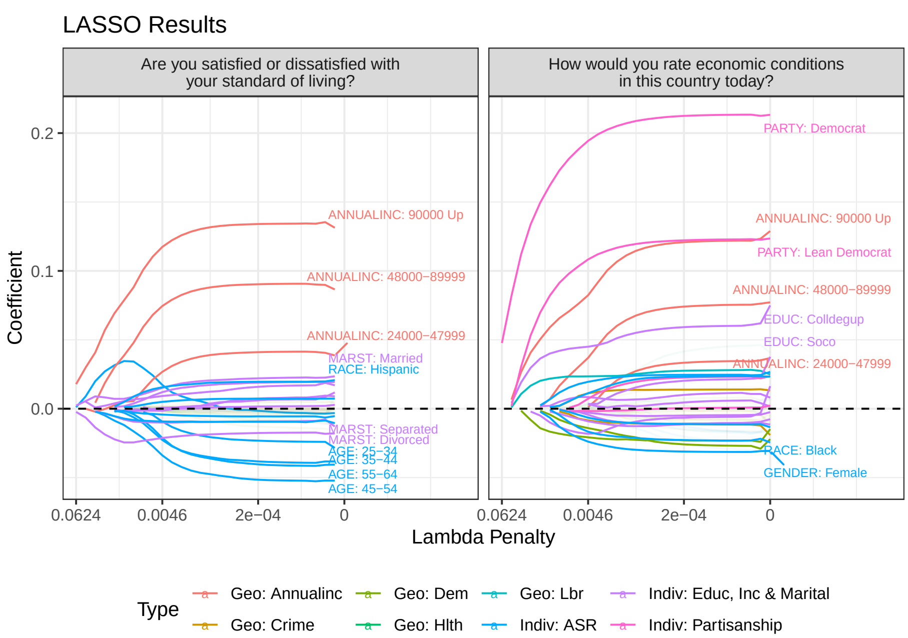

Figure S3: Estimated coefficients (y-axes) across different levels of L1 norm penalties (x-axes) for views on the economy (right panel) and self-reported satisfaction with one's standard of living (left panel). Labeled opaque values are consistently included in optimal model across 100 bootstrapped simulations. Data: 2016 Gallup responses augmented with contextual data (see Table S1).

How do these results generalize across all outcomes and all periods? Figure S4 summarizes the LASSO results by indicating the λ penalty at which each predictor is included, and highlighting the highest predictor with solid black borders. As illustrated, partisanship is the most important predictor for questions about the trend of the national economy, the current state of the national economy, and then approval for presidents Obama and Trump, as well as Hillary Clinton's favorability. As we move down the y-axis toward more egotropically-framed outcomes, we find weaker predictive power of partisanship, replaced by age, income, and marital status.

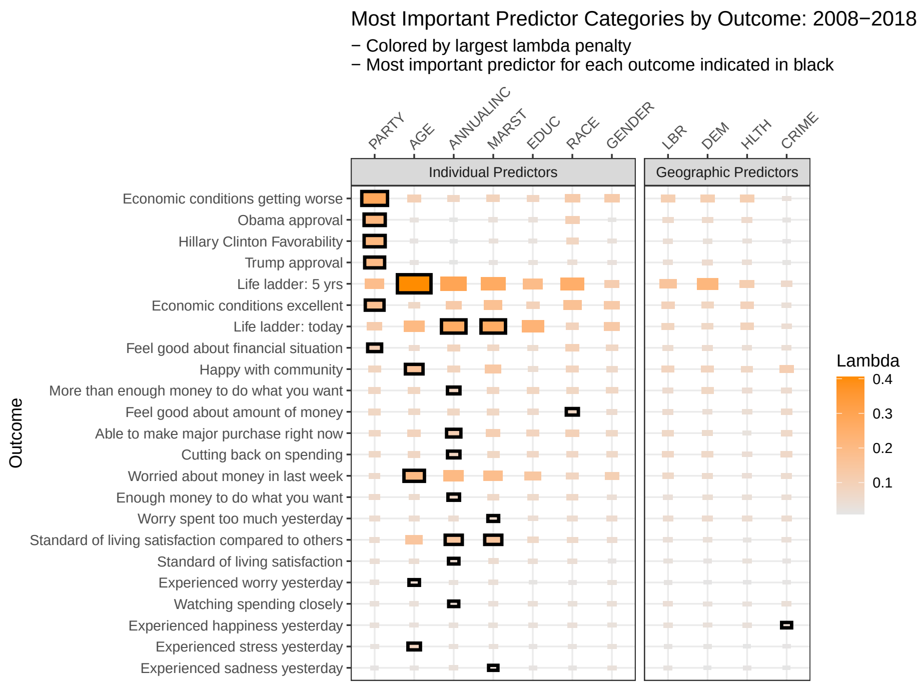

Figure S4: λ penalties at which different predictors (x-axis, top) are included in the LASSO regression for different outcomes (y-axis) across all years of the data. The largest λ penalty for each outcome is outlined in black (note that the coarseness of the grid for possible λ penalties means that some predictors are tied for importance for some outcomes).

An alternative approach to characterizing the prognostic power of a variable is to evaluate how much better we are at predicting an outcome when we add partisanship to a regression model. The simplest version of this is to compare a naive model that simply predicts the modal outcome category to a logistic regression of the outcome on an indicator for whether the respondent is a co-partisan of the president. We summarize the improvement in predictive accuracy using the expected percentage reduction in error (ePRE, Herron 1999) and visualize the results in Figure S5, which support our substantive argument that sociotropically framed questions are more sensitive to partisan motivated responding.

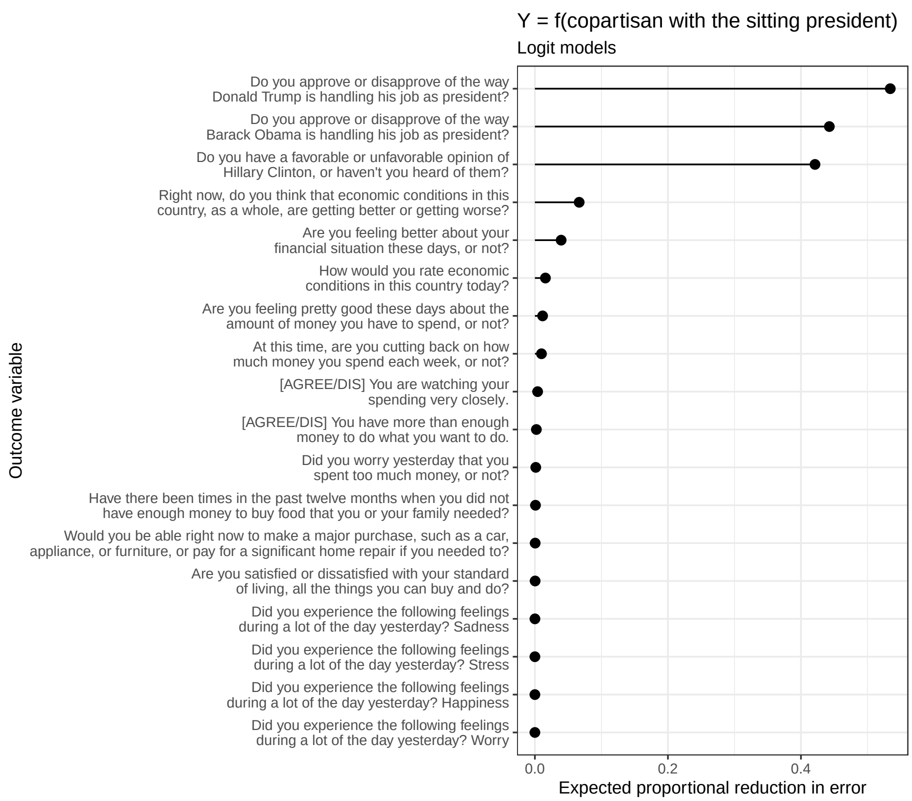

Figure S5: Expected percentage reduction in error (ePRE, Herron 1999) (x-axis) associated with the inclusion of an indicator for co-partisanship with the president, across binary survey questions (y-axis).

C Interrupted Time Series Analysis

Our main results present the evidence for flip-flopping with the president descriptively, simply plotting the attitudes on the economy among Democrats and Republicans by day in Figure 3. This visualization trades off methodological robustness for descriptive clarity, and in so doing assumes that there are not other things occurring around the time of the presidential transitions that might make the clear evidence of partisan flip-flopping on the economy spurious. Here, we estimate a rich interrupted time series specification in which we predict attitudes on the economy as a function of partisanship interacted with a binary variable *post* that takes on the value 1 if the outcomes are measured following the presidential inauguration or 0 otherwise. In addition, we control for all other individual-level covariates using a similar interacted specification, allowing us to both hold constant any alternative explanations that might be correlated with partisanship and economic attitudes, as well as examine the interacted coefficients on these non-partisan predictors. Finally, we include linear and cubic time trends measured before and after the inauguration, and also interact these with all individual-level covariates. The full specification is written below:

$$\begin{aligned} ECONBIN_{it} = \alpha_s &+ \beta_1 post_t + \beta_2 PARTY_{it} + \beta_3 PARTY_{it} * post_t \\ &+ \beta_4 preTrend_t + \beta_5 preTrend_t^2 + \beta_6 postTrend_t + \beta_7 postTrend_t^2 \\ &+ \beta_8 PARTY_{it} * preTrend_t + \beta_9 PARTY_{it} * preTrend_t^2 \\ &+ \beta_{10} PARTY_{it} * postTrend_t + \beta_{11} PARTY_{it} * postTrend_t^2 \\ &+ \gamma_{\mathbf{1}} \mathbf{X}_{it} + \gamma_{\mathbf{2}} \mathbf{X}_{it} * post_t + \gamma_{\mathbf{3}} \mathbf{X}_{it} * preTrend_t + \gamma_{\mathbf{4}} \mathbf{X}_{it} * preTrend_t^2 \\ &+ \gamma_{\mathbf{5}} \mathbf{X}_{it} * postTrend_t + \gamma_{\mathbf{6}} \mathbf{X}_{it} * postTrend_t^2 + \varepsilon_{it} \end{aligned} \quad (1)$$

where the binarized evaluation of the economy by respondent i in time t is predicted as a function of their partisanship interacted with the $post_t$ indicator variable, controlling for quadratic time trends in the pre ($preTrend_t$) and post ($postTrend$) periods, and a full set of individual-level covariates represented with the matrix $\mathbf{X}_{it}$. $\gamma_{\mathbf{j}}$ are vectors of coefficients premultiplied by the covariate matrices, which are also interacted with the same $post$, $preTrend$ and $postTrend$ variables.

We set the pre/post binary variable to the inauguration date, but also explore robustness of the interacted effect with partisanship over alternative dates between the election and two months after the election. In each regression, we subset the data to 60 days prior to and 60 days following the threshold date for the $post_t$ binary variable. Our main coefficients of interest are the β_3 and vector of $\gamma_{\mathbf{2}}$ coefficients, all of which describe how much partisanship (or other individual-level covariates) changed sign following the inauguration date. In using this rich specification, we can imagine this as a type of placebo test, except the placebos are the predictors instead of the outcome. Specifically, we don't expect a respondent's income to suddenly flip its sign with the respondent's evaluation of the economy following the inauguration of a new president. However, we do expect their partisanship to exhibit this flip, under the assumption that their evaluations of the economy are interpreted as a referendum on the president.

We visualize these coefficients by both the individual-level predictors (y-axis) and choice of pre/post date (x-axes) in Figure S6, exploring the 18 weeks between election day and March 1st of the following year for the 2008, 2012, and 2016 elections. Each tile is colored by the coefficient magnitude of the interaction term linking an individual-level covariate (y-axis) with the pre/post shift in its predict views of the economy by each date (x-axes). Note that the specification can be interpreted as a special case of a difference-in-differences estimator. Taking partisanship as an example, the interaction term β_3 captures how much the gap between strong Democrats and strong Republicans (difference one) changed before and after the inauguration (difference two). As illustrated, the interaction term between all partisanship categories (category 1, meaning strong Republicans, is the omitted category) and the $post_t$ binary variable is consistently statistically significant in both the 2008 and 2016 elections, which are when the incoming president was of a different party than the outgoing president. Importantly, none of the other individual-level predictors are consistently associated over these same periods. While one or another crops up as significant at the 5% threshold in a given week, there is no evidence of a consistent pattern in the data, especially when considering the signs of the coefficient estimates that cross the null week-by-week. These results strengthen our interpretation of the clear descriptive evidence of flip-flopping with the president as not spurious byproducts of some other predictor.

One other approach, albeit one very similar to the interrupted time series described above, is to run a regression discontinuity design with time as the forcing variable. We use the `rdrobust` package for `R` (Calonico, Cattaneo and Titiunik, 2015) and confirm that the interrupted time series results obtained above are replicated with this similar approach. (Note that the interrupted time series solution is effectively a regression discontinuity with time as a forcing variable, with the weights set to uniform and a second order polynomial. An added benefit of the ITS approach is that it is trivial to include interacted terms to examine the difference in partisanship before and after the break. With the `rdrobust` package, we instead subset the data and estimate several regression discontinuity designs for different partisan groups.) When estimated among strong Democrats using election day 2016 as the break point, the RD estimate is -0.117 (-0.181,-0.065), suggesting that Democrats' assessments of the economy declined substantially after the election of Donald Trump, despite the surrounding However, the same estimate among strong Republicans is a tightly estimated zero.

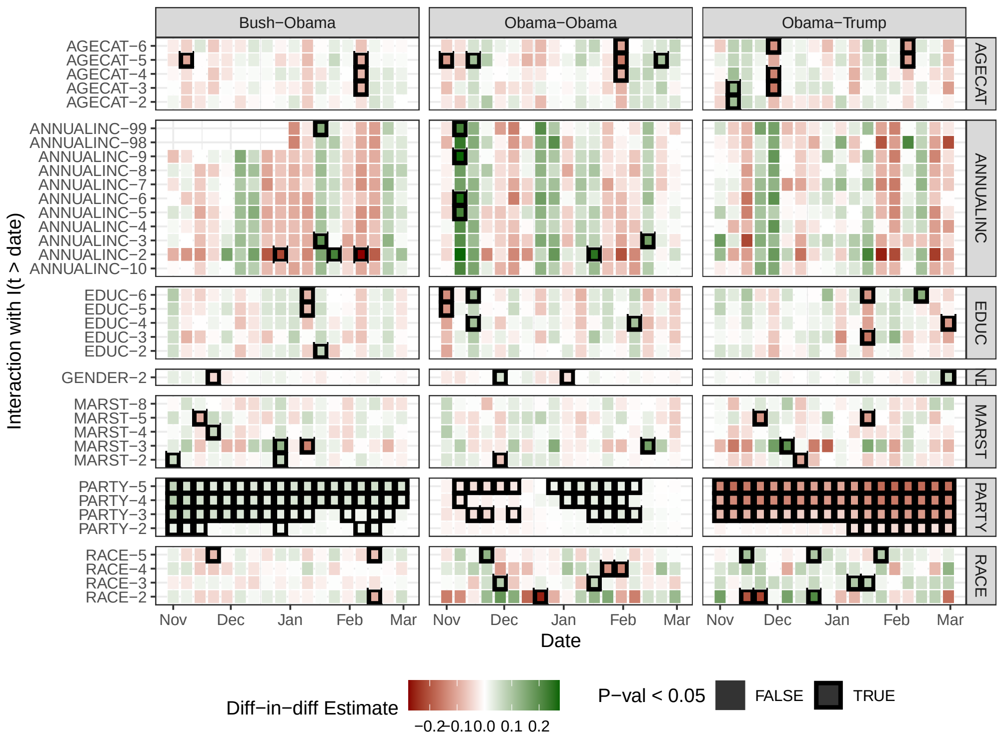

Figure S6: β_3 and $\gamma_{\mathbf{2}}$ coefficients estimated by week ranging from election day to March 1st of the following year for the 2008 (left panel), 2012 (center panel), and 2016 (right panel) presidential elections. Tiles are colored by the magnitude and sign of the coefficient. Estimates that are significant at the 95% threshold are highlighted with a black border.

D Other approaches to flip-flopping

Our main results simply displayed the clear evidence of partisan flip-flopping on the state of the economy before and after a change in the partisanship of the president. Here, we implement an agnostic alternative approach. Specifically, we run multiple change point analysis using the `mcp` package for `R`. This allows us to stipulate two potential change points, reflecting our theoretical assumption that partisan shifts on the state of the economy should reflect the election date and the inauguration date. However, the change point analytic method allows the data to determine when there are discontinuous breaks in a time series vector of observations. We expect these to occur around the election and the inauguration date, reflecting our presumption that these are two opportunities for partisan respondents to reinterpret the association between economic health and the performance of the president. However, there are two factors to consider when expecting the changes to occur around these dates.

First is the outcome variable itself. Among the sociotropically-worded questions, the one most amenable to a substitution for the respondent's support for the (incoming) president is the question phrased about whether the national economy is improving. The word "improving" should allow the respondent to easily substitute their opinion of an incoming president with the economic question, and should reveal the strongest evidence of partisan cheerleading. Conversely, the similarly sociotropic question of "How would you rate economic conditions in the country today" should be more muddled around the time of an election, reflecting the greater uncertainty about to whom economic conditions should be attributed.

Second is the election. The 2008 and 2016 elections saw a transition of partisan control of the Oval office, with Obama (a Democrat) replacing Bush (a Republican) in 2008, and Trump (a Republican) replacing Obama in 2016. Conversely, we should see less clear evidence of discontinuities in 2012, since Obama won re-election, defeating the Republican challenger Mitt Romney.

We visualize these results in Figure S7, which visualizes the structural breaks in the time series data of the difference between Republicans and Democrats in their views on the economy (y-axes). Blue histograms represent the most likely structural breaks in the data, which are also reflected in the grey lines of best fit. As illustrated, we see clear evidence of two structural breaks around the election and the inauguration in 2008 and 2016, but less clear evidence of the same in 2012. Furthermore, the evidence of these twin breaks is starker when we use the sociotropic question about "improving" the economy, compared to the more nebulous question about how the respondent would rate the economoy "today". Here we see striking evidence of two structural breaks exactly at the 2008 and 2016 elections, and then again just after the 2009 inauguration, and again exactly at the 2017 inauguration.

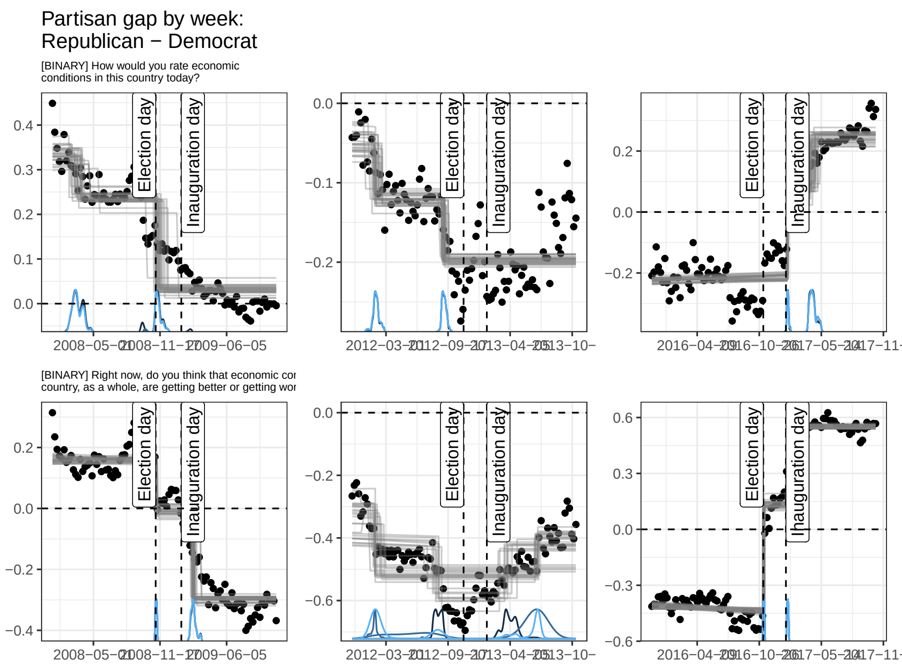

Figure S7: Bayesian change point analysis with two breaks estimated on whether the respondent thinks the economy today is good (top row of plots) or whether the national economy is playing (bottom row of plots). Blue densities reflect the most likely location of the structural breaks in the data.

E Extended Analysis of Survey Design Change

Our main analyses focused on the change in the variable importance measure of partisanship just prior to, and just following, the change in Gallup's survey design on January 1st, 2013. We present several extensions here.

First, we evaluate these results subject to finer-grained temporal units, including weeks and days. As illustrated in Figures S8 and S9, the substantive conclusions hold, albeit with increasing noise as we disaggregate to smaller temporal units.

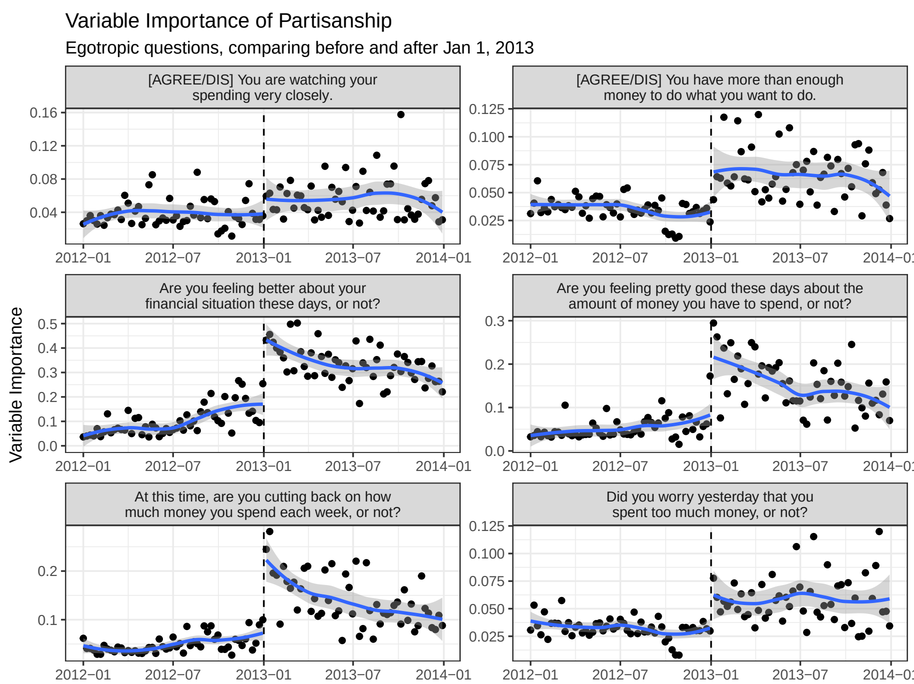

Figure S8: Variable importance of party affiliation (from models estimated on a monthly basis) on a battery of egotropically-phrased questions about the respondent's economic condition. Loess smoothers fit separately prior to, and following, January 1st 2013 when Gallup split the survey into two samples. Data aggregated to week level.

Second, we apply Bayesian Change Point Analysis (BCP) to let the data inform if and when a discontinuous break appears in these data. As illustrated in Figures S10, S11, and S12, the analysis consistently chooses the period on or around January 1st, 2013 across different outcomes, regardless of whether we aggregate to the month, week, or day.

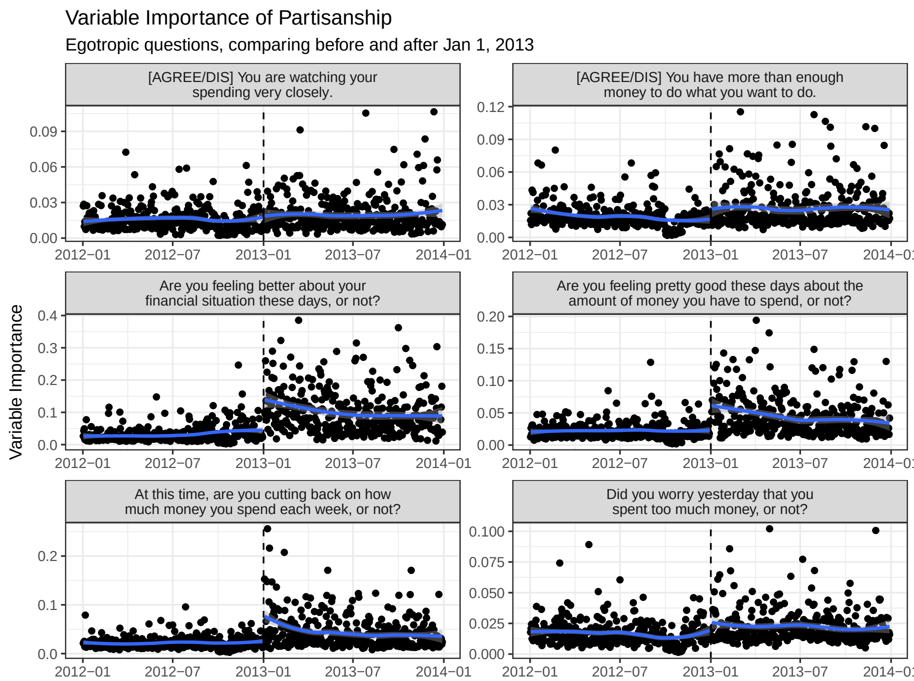

Figure S9: Variable importance of party affiliation (from models estimated on a monthly basis) on a battery of egotropically-phrased questions about the respondent's economic condition. Loess smoothers fit separately prior to, and following, January 1st 2013 when Gallup split the survey into two samples. Data aggregated to day level.

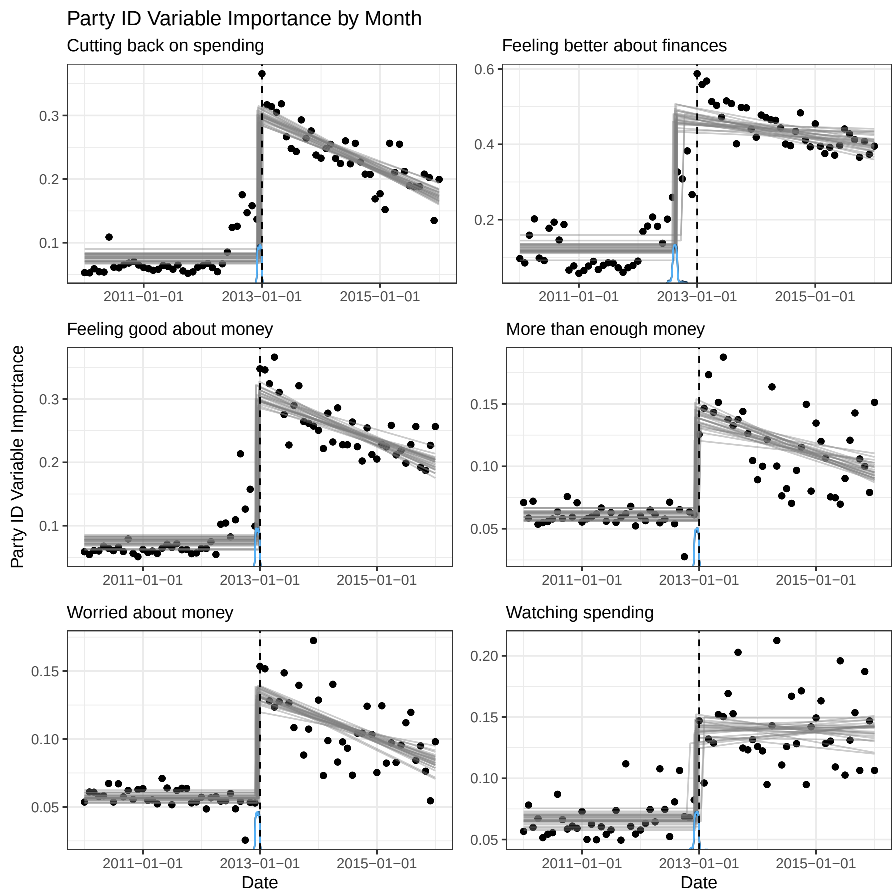

Figure S10: Bayesian change point detection, data aggregated to months.

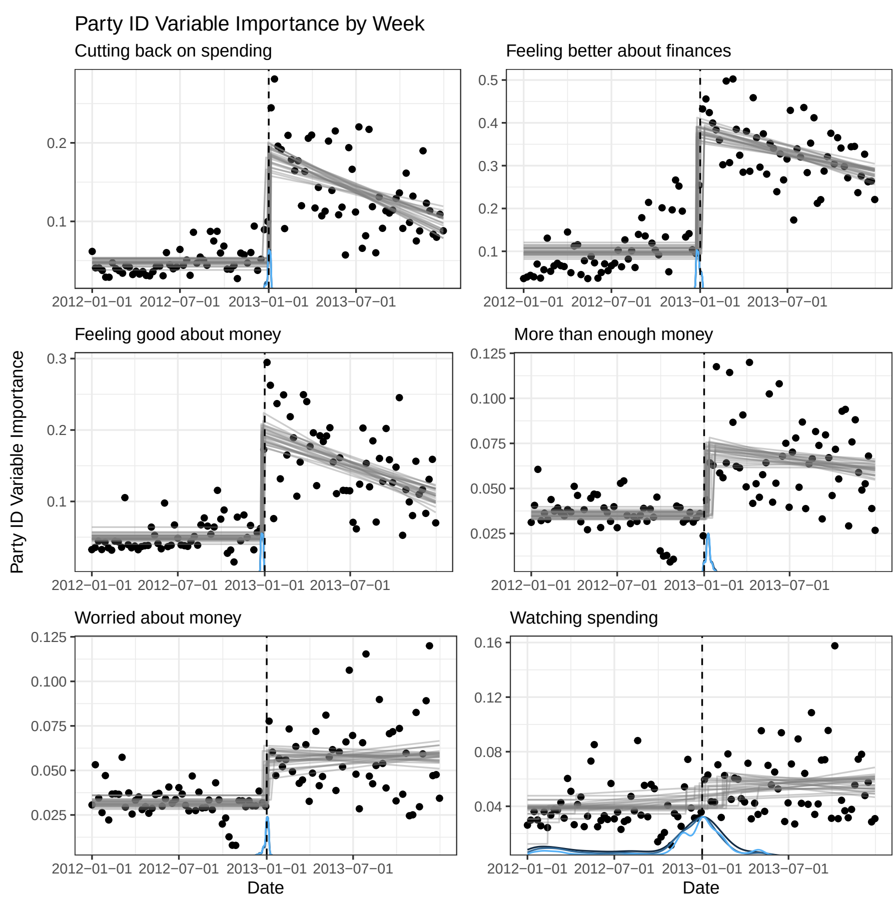

Figure S11: Bayesian change point detection, data aggregated to weeks.

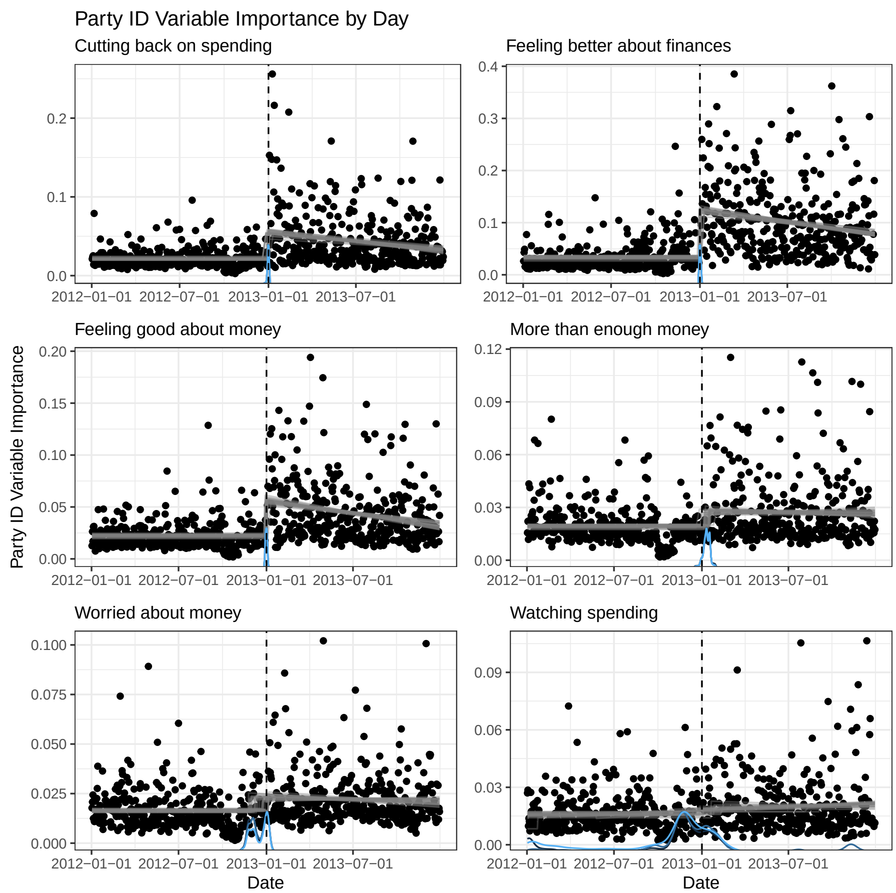

Figure S12: Bayesian change point detection, data aggregated to days.

References

Calonico, Sebastian, Matias D Cattaneo and Rocío Titiunik. 2015. "Rdrobust: An R package for robust nonparametric inference in regression-discontinuity designs." *R Journal* 7(1):38–51.

Herron, Michael C. 1999. "Postestimation Uncertainty in Limited Dependent Variable Models." *Political Analysis* 8(1):83–98.

Wright, Marvin N and Andreas Ziegler. 2017. "ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R." *Journal of Statistical Software* 77(1):1–17.