Irrelevant Events and Voting Behavior: Replications Using Principles from Open Science

RUNNING HEADER:

Irrelevant Events and Voting Behavior

Matthew H. Graham* Gregory A. Huber[†] Neil Malhotra[‡]
Cecilia Hyunjung Mo[§]

June 11, 2020

*Matthew H. Graham is a Ph.D. Candidate in the Department of Political Science, Yale University, PO Box 208301, New Haven, CT 06520-8301, matthew.graham@yale.edu.

[†]Gregory A. Huber is the Forst Family Professor of Political Science, Yale University, PO Box 208209, New Haven, CT 06520-82092, gregory.huber@yale.edu.

[‡]Neil Malhotra is the Edith M. Cornell Professor of Political Economy, Stanford Graduate School of Business, 655 Knight Way, Stanford, CA 94305-7298, neilm@stanford.edu.

[§]Cecilia Hyunjung Mo is an Assistant Professor of Political Science, University of California, Berkeley, 210 Barrows Hall, Berkeley, CA 94720-1950, cecilia.h.mo@berkeley.edu.

Abstract

How well do voters hold politicians accountable? Although a longstanding research tradition claims that elections are effective tools for the sanctioning and selection of leaders, a more-recent literature argues that voters often reward and punish incumbents for “irrelevant events.” The empirical literature on this topic is characterized by conflicting findings. Drawing upon ideas from the open science movement, and showing how they can advance the transparency of observational research, we replicated three prominent studies on “irrelevant” events and voting behavior: (1) Achen and Bartels’s (2016) study of droughts and floods; (2) Healy, Malhotra, and Mo’s (2010) study of college football; and (3) Healy and Malhotra’s (2010) study of tornadoes. Each study replicates well in some areas and poorly in others. Had we sought to debunk any of the three with *ex post* specification search, we could have done so. However, our approach required us to see the full, complicated picture.

KEYWORDS: irrelevant events, observational studies, open science, replications, voting behavior

Supplementary material for this article is available in the appendix in the online edition. Replication files are available in the JOP Data Archive on Dataverse (<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/J3UBZ6>). We pre-registered our study design in the Evidence in Governance and Politics (EGAP) repository (<http://www.egap.org/registration/4715>). Support for this research was provided by the Institution for Social and Policy Studies at Yale University.

A core question in political science concerns democratic accountability, or the extent to which elections are effective tools for enhancing social welfare via the electoral sanctioning and selection of leaders. One research tradition holds that voters effectively hold politicians accountable by, for example, rewarding (punishing) them for good (poor) economic performance (Ashworth 2012). More recent work challenges these canonical models of retrospective voting by arguing that voters often reward and punish incumbents for shifts in affect caused by events beyond their control (Healy and Malhotra 2013). This phenomenon would decrease the likelihood that voters consistently select competent leaders, and, if known to politicians, could also reduce incentives for good governance. Achen and Bartels (2016) introduced this idea, pointing to several instances in which voters apparently punished politicians for “acts of God” such as droughts, floods, and shark attacks. Healy, Malhotra, and Mo (2010) originated the term “irrelevant events” to characterize the category of events for which incumbents are not responsible, and to which voters should not respond.¹ This research agenda has been highly influential and widely cited (see Online Appendix A for more details).

It is fair to characterize the observational² empirical literature on irrelevant events as a “mess.” Several studies suggest that irrelevant events like natural disasters (Achen and Bartels 2016; Heersink, Peterson, and Jenkins 2017), sporting events (Healy, Malhotra, and Mo 2010; Busby, Druckman, and Fredendall 2017; Miller 2013), lottery results (Bagues and Esteve-Volart 2016), and irrelevant tax referenda (Sances 2017) affect election results. Other studies are skeptical of these claims (Fowler and Montagnes 2015; Fowler and Hall 2018) or produce mixed results (Busby and Druckman 2018). Finally, some acknowledge these events’

¹As Healy, Malhotra, and Mo (2010) point out and Ashworth, Bueno de Mesquita, and Friedenberg (2018) formalize, caution is warranted when labelling events “irrelevant,” as incumbents’ responses to these events could signal quality. Further, there has been theoretical discussion about the implications of these results for democratic accountability (Ashworth and Bueno De Mesquita 2014). Our focus is on the strength of the empirical evidence that these events influence elections in the first place. In all disciplines, theories are constructed around empirical regularities. When empirical foundations are misleading, theory can be led astray.

²We do not consider experimental studies on the subject, several of which have found evidence in support of the irrelevant events hypothesis (e.g., Study 2 in Healy, Malhotra, and Mo 2010; Huber, Hill, and Lenz 2012; Healy and Lenz 2014).

influence but interpret them as instances of democratic accountability working effectively (Healy and Malhotra 2010).

An open science approach to improving the transparency of observational research can help clarify these muddy waters. Too frequently, the legitimate exercise of re-examining previously published findings gets bogged down by the possibility that both sides engaged in selective reporting, leaving consumers of research in the unenviable position of trying to adjudicate between competing claims without much insight into the process that produced them. Our approach to avoiding these problems is applicable not only to the literature on irrelevant events, but also to other areas with mixed findings.

We holistically replicated three prominent studies in the irrelevant events literature: (1) Achen and Bartels’s (2016) finding that incumbents are electorally harmed by abnormal precipitation (droughts and floods), which we hereafter refer to as AB; (2) Healy, Malhotra, and Mo’s (2010) finding that the results of college football games immediately preceding Election Day influence election results (hereafter HMM); and (3) Healy and Malhotra’s (2010) finding that incumbents are harmed by tornado damage but not by tornado deaths (hereafter HM). HM argue that their findings go *against* the irrelevant events hypothesis: voters respond to tornado damage, reflecting economic voting, but do not emotionally respond to tornado deaths. AB and HMM claim that their results suggest that voters are not able to distinguish between relevant and irrelevant sources of subjective well-being, and as such, the pain or euphoria from irrelevant events is politically consequential.

The open science approach we developed for this replication effort has five key features (for more details on our guiding principles, see Online Appendix B). First, we conducted *multiple simultaneous replications* of separate studies, using their common theoretical basis to guide our approach to model re-specification. Second, as one would do with experimental open science, we *pre-registered* our analysis. Third, we leveraged the passage of time to *extend the time series* without changing the original studies’ units of analysis. Fourth, we *collected all data independently* instead of simply using replication data from the original studies. Fifth, this study was a *collaboration between authors with mixed incentives*. Collaborations in which some authors have a stake in the original results and others have no such stake (and potentially a motivation to overturn the original findings) can make replication efforts more credible.

Using this approach, we simultaneously conducted the following analyses for each of the

three studies: (1) a replication of the original analysis with the original specification and time range, but with independently collected data; (2) a replication of the original analysis with an extended time series; (3) an assessment of theoretically relevant heterogeneous effects; (4) an evaluation of the robustness of the original findings with pre-registered alternative specifications; (5) a series of placebo tests; and (6) a consideration of multiple election types. We also explored issues idiosyncratic to each study. We present the results of the replication efforts below. Further details, including regression specifications, data quality issues, and a more extended discussion of the implications of our findings, can be found in the Online Appendix.

Study 1: Droughts and Floods

Using state-level data on presidential elections from 1900 to 2000, AB find that voters punish incumbent presidents for droughts and floods. The replication produces mixed evidence about the original findings. With an independently collected dataset, we generate slightly stronger findings than AB did originally. When we extend the time series, the estimates shrink only modestly and generally retain statistical significance. Concerns emerge when we respecify the original model. Most importantly, the original results depend on the exclusion of state fixed effects. We also find little evidence of stronger effects in more rural states. The findings are largely similar when we include gubernatorial and Senate elections for the years they are available. When additional election types are considered, the results are relatively less sensitive to the inclusion of state fixed effects; however, they are sensitive to the extension of the time series.

Replication with Original Sample Period and Specifications: The AB study passes the pure reproducibility test. In the “original” results in the top row of Figure 1, the “Overall” lines correspond to the results reported by AB in columns 1-3 of their Table 5.2. We replicated all three results with our original data collection effort for the same sample period. We find that a one-unit increase in the drought index is associated with diminished incumbent presidential vote share (first column in Figure 1). We estimate a 0.070 percentage point decrease in incumbent vote share ($se = 0.031$, $p = 0.02$), which is a stronger effect than what AB report ($\beta = -0.060$, $se = 0.031$). As shown in the second and third columns in Figure 1, a specification that included three lags of the dependent variable, and another that weighted these lags into a single “time-weighted” measure, also yield slightly stronger negative effects

Figure 1: Effect of droughts and floods: presidential elections

Note: This figure displays results from the replication of AB. The authors' theory predicts that the overall effect (squares) will be negative and that the effect will be more negative in rural areas (triangles to the left of circles). Red markers indicate that an equivalent estimate appeared in the original paper. For corresponding regression tables, see Online Appendix F.

than AB report.³

Extending the Time Series: Extending the sample largely confirms the original results, but the reported effects would have been somewhat smaller had the original study been conducted today. The second row of Figure 1 reports the effect of adding an additional 16 years of data (representing 13.5 percent of the unweighted data and 25.8 percent of the weighted data) for all three specifications. The estimates for the election year drought index with and without the lagged index shrinks by 11 percent and 14 percent, respectively, but remains significant at $p < 0.05$. The time-weighted index coefficient shrinks by 35 percent and approaches statistical significance at the 5 percent level ($p = 0.074$).

Heterogeneous Effects: In Figure 1, the plotted estimates for “Rural = 1” and “Rural = 0” are the predicted incumbent vote share loss when the rural variable is dichotomously set to 1 or 0. In Online Appendix F, we present results from models with interaction terms that formally test whether the effect of droughts is different in rural areas. While the interaction of the drought index and proportion rural is consistently negative, it is never statistically significant for presidential elections for the original time series or the full time series. The p-values for interaction coefficients that correspond to Figure 1 range from 0.18 to 0.66. In the respecified model, the interactions take the wrong sign. These results suggest that the

³For the index plus lags specification, we find $\beta = -0.065$ (se = 0.033, $p = 0.049$), whereas AB report $\beta = -0.052$ (se = 0.034). For the time-weighted specification, we find $\beta = -0.123$ (se = 0.047, $p < 0.01$), whereas AB report $\beta = -0.104$ (se = 0.045).

drought effect is not larger in more-rural states.

Robustness to Pre-registered Changes in Model Specification: We also examined the effect of two pre-registered alterations to AB’s original specification. We find that the original results reported in AB are sensitive to both modeling choices.

First, the “no weights” results in Figure 1 show the effect of weighting all observations equally rather than by the number of voters in each state-year. In every other respect, these models exactly match the original model. As AB note, the inclusion of voter population weights give more influence to more populous states and more recent years. When we compared the unweighted models to the “original” weighted models, we find that the index and index plus lags specifications have fairly similar coefficients but larger standard errors. The time-weighted index results are more sensitive to the exclusion of weights, dropping in magnitude by about 30 percent. The rural interaction terms remain statistically insignificant.

Second, the “respecified” results of Figure 1 show outcomes from our alternative specification. The point estimates are fairly similar to the original model, but fall short of statistical significance due to larger standard errors. Despite the appearance that the re-specification merely added noise around stable point estimate, we view this result as important for two reasons. First, fixed effects are theoretically necessary to correct for mean differences between states, and do in fact improve model fit (see Online Appendix Tables F.19 to F.22). Second, state fixed effects alone are sufficient to alter the original results—in fact, state fixed effects have a larger impact on the original results than is evident in our preferred specification. To assess whether the respecified model only fails to confirm AB’s results because it omits variance-absorbing control variables, we ran additional regressions (not pre-registered) that added state fixed effects alone to the original specification. When the only change to AB’s original model is the addition of state fixed effects, the “original” coefficients shrink by a larger amount than in the respecified model (by 30-50 percent) and lose statistical significance (see Online Appendix Tables F.7 and F.8).

Placebo Tests: We also estimated a series of placebo test regressions to examine the effect of future droughts and floods on contemporaneous election outcomes. Across the 18 non-interactive models summarized in Figure 1, as well as the 18 equivalents for all elections, the placebo coefficient is never statistically significant (columns 2, 5, and 8 of Online Appendix Tables F.1-6 and F.9-14).

Incorporating Additional Election Types: In the Online Appendix, Figure F.2 presents

the same estimates as Figure 1, but with gubernatorial and Senate elections added to the sample.⁴ The pattern of effects is broadly similar. As with the analysis of only presidential elections, there is a negative relationship between droughts and floods and incumbent vote share when considering all three election types during the original sample time period for all three specifications. Relative to the results noted above, which only examine presidential elections, the all-elections models are less sensitive to model respecification—in each of the three versions of AB’s model, the coefficient on the drought index attains or comes close to statistical significance in the original sample (see the original sample results in Online Appendix Figure F.2). However, these estimates are more sensitive to extending the time series. In the full 1900-2016 sample, no regression model with state fixed effects produces a statistically significant coefficient on the drought index.

Study 2: Football Games

The evidence presented here is mostly supportive of HMM’s conclusion that college football games affect presidential, gubernatorial, and Senate election outcomes. In the full sample, a win within 13 days of Election Day causes the incumbent to receive an additional 1.1 percentage points of the vote in Senate, gubernatorial, and presidential elections ($p < 0.01$). This is somewhat smaller than those reported by HMM ($\beta = 1.13$ as opposed to 1.61). Of the three replicated studies, HMM featured the largest extension of the time series (42%) and the fewest changes to the original model; reassuringly, neither the new data nor model respecification had much effect on the estimates. We view these results as a confirmation of the originally reported effect, albeit at a smaller magnitude.

However, two findings suggest revisions to the original conclusions. First, we could not replicate the original claim that counties with “powerhouse” college teams are more-affected. The revised result is not consistent with the irrelevant events hypothesis. Second, extending the time series suggests that games the week before the election matter more than games two weeks before the election, the opposite of the pattern in the original HMM study. This revision does not challenge the predictions of the irrelevant events hypothesis.

Replication with Original Sample Period and Specifications: As shown in the first row in the rightmost column of Figure 2, the original findings appear to have depended, in part, on

⁴Our gubernatorial election data go back to 1900. With a few exceptions, prior to 1914, Senators were elected by state legislatures rather than voters.

Figure 2: Effect of college football games: presidential, Senate, and gubernatorial elections

Note: This figure displays results from the replication of HMM. The authors' theory predicts that the overall effect (squares) will be positive and that the effect will be more positive for teams with high attendance or past championships (triangles to the right of circles). Red markers indicate that an equivalent estimate appeared in the original paper. For corresponding regression tables, see Online Appendix L.

data errors in the original study (described in Online Appendix K). HMM report that a win within two weeks of an election increases incumbent vote share by 1.6 percentage points (se = 0.58, $p < 0.01$). In our replication using the original sample (1985-2006), the coefficient is nearly one-third smaller ($\beta = 1.1$, se = 0.62, $p = 0.07$). The left and middle columns split the results according to whether the game occurred two weeks or one week before the election. Similar to the original study, the estimate for two weeks before is larger than the estimate for one week before.

Extending the Time Series: We expanded the election years of interest to 1980 through 2016 (see the bottom row of Figure 2). The additional years, 1980-1984 and 2007-2016, constitute 42 percent of the total data. Comparing the two rows of Figure 2's rightmost column, the addition of out-of-sample data has little effect on the point estimate but shrinks the standard error considerably. The estimate suggests that a win within two weeks of the election increases incumbent vote share by 1.1 percentage points (se = 0.38, $p < 0.01$).

The additional years of data change one pattern from the original study: the estimate for games one week before the election is larger ($\beta = 1.63$, se = 0.55, $p < 0.01$) than the estimate for games two weeks before ($\beta = 0.47$, se = 0.54, $p = 0.38$). HMM found the opposite. This discrepancy, however, does not weaken HMM's original claims regarding the role of irrelevant events.

Heterogeneous Treatment Effects: We find only weak evidence of larger effects for high-attendance or championship teams. Though the predicted values and differences between

the groups have the correct sign, the interaction term misses statistical significance for the attendance measure in the 1985-2006 sample. Moreover, extending the time series more than halves the interaction terms for both championship teams ($\beta = 2.23$ drops to $\beta = 0.51$) and high attendance teams ($\beta = 1.89$ drops to $\beta = 0.87$; compare columns 11 and 12 of Online Appendix Tables L.1 and L.2). We interpret these results as a failure to replicate the original claim that football games matter more for high-attendance and championship teams.

Robustness to Pre-registered Changes in Model Specification: These effects change very little in the respecified model, which differs from the original model only in its exclusion of control variables. To see this in Figure 2, observe that in both rows, the results labeled “Original” are very similar to those labeled “Respecified.”

Placebo Tests: Our placebo tests replicated HMM’s placebo tests assessing the effects of wins one week after or two weeks *after* the election results. In placebo tests that correspond to each panel of Figure 2, we find no statistically detectable relationship between election outcomes and post-election games (see columns 2, 6, and 10 of Online Appendix Tables L.1 to L.4).

Study 3: Tornadoes

Finally, we examined HM’s finding that incumbents in presidential elections are punished for tornado damage but not for tornado deaths, and are only punished for damage when no federal disaster declaration is issued. The authors cast these results as more consistent with economic voting than blind retrospection: voters punish incumbents for not dealing with the disaster, not because severe weather happened to kill someone. We find that the main results are not robust. Although the main result using the original specification is generally confirmed out-of-sample, the result is not statistically significant when: (1) examining election types besides presidential elections; (2) using reasonable alternative specifications; and (3) using the damage measure originally reported in the article. This last point refers to our discovery that HM’s analysis transformed per capita tornado damage using a different formula than reported; consequently, we considered both the version HM reported and the version HM actually used. We view this failure to replicate as the most concerning, as results should be fairly insensitive to assumptions about functional form not clearly grounded in theory. Despite the differences in empirical results, our exercise sustains HM’s substantive conclusion that voters do not respond purely emotionally to natural disasters outside of

Figure 3: Effect of tornadoes: presidential elections

Note: This figure displays results from the replication of HM. The authors' theory predicts that the overall effect of damage (squares) will be negative and that the effect will be less negative when a disaster was declared (triangles to the right of circles). Red markers indicate that an equivalent estimate appeared in the original paper. "(used)" indicates analyses using the damage variable that HM actually used and "(reported)" indicates analyses using the damage variable that HM reported that they used. For corresponding regression tables, see Online Appendix I.

government control.

Replication with Original Sample Period and Specifications: As shown in the top left of Figure 3, we generally replicate the original result. A 10 percent increase in the tornado damage measure is associated with a 1.2 percentage point reduction in incumbent vote share ($\beta = -1.17$, $se = 0.48$, $p = 0.015$). This coefficient is smaller than the one reported by the original authors, but with a substantially smaller standard error ($\beta = -1.42$, $se = 0.83$, $p < 0.1$).⁵ The second column of Figure 3 generally confirms HM's result that tornado deaths do not negatively affect incumbent vote share.

Extending the Time Series: After including the out-of-sample observations, the result is attenuated but remains statistically significant ($\beta = -0.92$, $se = 0.47$, $p = 0.0496$). The out-of-sample data constitute 17 percent of the full dataset, and adding these data reduces the estimated effect size by about 21 percent. For the deaths measure, adding additional data slightly increases the effect size and pushes it into statistical significance. However, the positive sign (suggesting deaths improve incumbent electoral performance) contradicts the irrelevant events hypothesis.

Heterogeneous Effects: Next, we assessed the hypothesis that the negative impact of tornado damage on incumbent vote share is conditional on whether the president issues a disaster declaration. The interaction terms (see the second and sixth columns of Tables

⁵The greater precision may be due to the fact that in our independently collected data, about 20 percent more counties had tornado damage.

I.1-I.4 in the Online Appendix) fail to back this claim: effects are close to zero in the pure replication, though the estimates in the respecified model come close to statistical significance when using the damage variable used by HM ($p = 0.15$ in the original sample and 0.07 in the full sample; first column of Figure 3). Evidence is weaker still using the damage formula originally reported by HM (see the third column of Figure 3). Figure I.2 and Tables I.5-I.8 in the Online Appendix show a similar lack of evidence for heterogeneous effects when pooling across presidential, gubernatorial, and Senate elections.

Robustness to Pre-registered Changes in Model Specification: The result is weaker when using the respecified model. The estimated coefficient for the tornado damage variable shrinks by about half with our pre-registered preferred specification (see first column of bottom panel of Figure 3; with the original sample, $\beta = -0.52$, $se = 0.30$, $p = 0.09$; in the full sample, $\beta = -0.46$, $se = 0.29$, $p = 0.11$).

Placebo Tests: The models also fail the pre-registered placebo tests. As shown in the Online Appendix, future tornado damage is positively associated with past incumbent vote share, even when conditioning on expectations (see Online Appendix A for details on an additional, non-preregistered placebo test that HM pass).

Incorporating Additional Election Types: The negative relationship between tornado damage and election outcomes seems to be confined to presidential elections. Across the different models, including Senate and gubernatorial elections halves the effect sizes and renders them statistically insignificant (see Figure I.2 and Tables I.5-I.8 in the Online Appendix). This seems inconsistent with a story about purely emotional responses to natural disasters, as tornado damage should also adversely affect other officeholders if voters are misattributing their dissatisfaction with the status quo to incumbents more generally.

Additional Issues: Our main data quality discovery is that HM’s findings are very sensitive to functional form assumptions. As described in Online Appendix H, HM’s analysis used a different damage variable than reported. The results generally do not replicate using the reported damage measure.⁶ Given that it is unclear which functional form is more theoretically appropriate, it is reasonable to expect that the result to hold using both measures.

⁶Using the reported damage formula, only one coefficient comes close to statistical significance. When considering all elections, employing the original specification, and including out-of-sample observations, we estimated $\beta = -0.058$, $se = 0.036$, $p = 0.11$.

Discussion

Political scientists increasingly agree about the importance of replication efforts and the use of open science tools to achieve greater research transparency. Although most of the discussion on replicability and the creation of open science procedures has taken place within the context of experiments, observational studies still constitute the bulk of political science research. We devised our approach with the hope that it can serve as a guide for conducting thorough but fair replications of observational studies. Just as published observational studies underpin much of what is known about the workings of politics, the quality of replication practices will determine whether the right evidence stands or falls.

Our results provide mixed evidence for the irrelevant events hypothesis. Even as the original findings successfully replicated in many instances, the strength of the evidence for the electoral effects of irrelevant events is weaker than was originally reported. In the case of HM, the original result was not robust to a seemingly innocuous difference between the reported and actually-used formula for transforming the independent variable. Though we were able to replicate the main HMM and AB results, the former’s estimates were inflated by data errors, while the latter’s depended on the exclusion of state fixed effects. In all three replications, we failed to find solid evidence of the originally reported heterogeneous effects.

Though our results did not fully confirm or overturn the original findings of any study, we view this as a feature, not a bug. Absent the disciplining features of our approach, we could have easily “null hacked” (Protzko 2018) away all three of these studies by selectively reporting the failed tests and failing to report the successful ones. Publication bias may even have rewarded such a strategy. From this, we infer that publication biases (or researchers’ anticipation of them) may make published findings—both original studies and replications of these studies—appear more dispositive than they are.

Acknowledgements

We would like to thank Christopher H. Achen, Larry Bartels, Jamie Druckman, and Anthony Fowler for their helpful comments and advice. We also want to thank our anonymous reviewers, as well as participants of the 2019 Annual Meeting of the Midwest Political Science Association, the 2019 Berkeley Initiative for Transparency in the Social Sciences (BITSS) Annual Meeting, the 2019 BITSS workshop titled “Unlocking the File Drawer” at UC Berkeley, the Institute of Social Science Special Lecture Series at the University of Tokyo, the

Research Workshop in American Politics (RWAP) at UC Berkeley, and the USC Marshall School of Business Workshop for helpful feedback.

References

- Achen, Christopher H., and Larry M. Bartels. 2016. *Democracy for Realists*. Princeton, NJ: Princeton University Press.
- Ashworth, Scott. 2012. "Electoral Accountability: Recent Theoretical and Empirical Work." *Annual Review of Political Science* 15: 183–201.
- Ashworth, Scott, and Ethan Bueno De Mesquita. 2014. "Is Voter Competence Good for Voters?: Information, Rationality, and Democratic Performance." *American Political Science Review* 108 (3): 565–587.
- Ashworth, Scott, Ethan Bueno de Mesquita, and Amanda Friedenberg. 2018. "Learning about Voter Rationality." *American Journal of Political Science* 62 (1): 37–54.
- Bagues, Manuel, and Berta Esteve-Volart. 2016. "Politicians' Luck of the Draw: Evidence from the Spanish Christmas Lottery." *Journal of Political Economy* 124 (5): 1269–1294.
- Busby, Ethan C., and James N. Druckman. 2018. "Football and Public Opinion: A Partial Replication and Extension." *Journal of Experimental Political Science* 5 (1): 4–10.
- Busby, Ethan C., James N. Druckman, and Alexandria Fredendall. 2017. "The Political Relevance of Irrelevant Events." *The Journal of Politics* 79 (1): 346–350.
- Fowler, Anthony, and Andrew B. Hall. 2018. "Do Shark Attacks Influence Presidential Elections? Reassessing a Prominent Finding on Voter Competence." *The Journal of Politics* 80 (4): 1423–1437.
- Fowler, Anthony, and B. Pablo Montagnes. 2015. "College Football, Elections, and False-Positive Results in Observational Research." *Proceedings of the National Academy of Sciences* 112 (45): 13800–13804.
- Healy, Andrew, and Gabriel S. Lenz. 2014. "Substituting the End for the Whole: Why Voters Respond Primarily to the Election-Year Economy." *American Journal of Political Science* 58 (1): 31–47.
- Healy, Andrew, and Neil Malhotra. 2010. "Random Events, Economic Losses, and Retrospective Voting: Implications for Democratic Competence." *Quarterly Journal of Political Science* 5 (May): 193–208.
- Healy, Andrew, and Neil Malhotra. 2013. "Retrospective Voting Reconsidered." *Annual Review of Political Science* 16: 285–306.
- Healy, Andrew, Neil Malhotra, and Cecilia Hyunjung Mo. 2010. "Irrelevant Events Affect Voters' Evaluations of Government Performance." *Proceedings of the National Academy of Sciences* 107 (29): 12804–12809.
- Heersink, Boris, Brenton D. Peterson, and Jeffery A. Jenkins. 2017. "Disasters and Elec-

- tions: Estimating the Net Effect of Damage and Relief in Historical Perspective.” *Political Analysis* 25 (2): 260–268.
- Huber, Gregory A., Seth J. Hill, and Gabriel S. Lenz. 2012. “Sources of Bias in Retrospective Decision Making: Experimental Evidence on Voters’ Limitations in Controlling Incumbents.” *American Political Science Review* 106 (4): 720–741.
- Miller, Michael K. 2013. “For the Win! The Effect of Professional Sports Records on Mayoral Elections.” *Social Science Quarterly* 94 (1): 59–78.
- Protzko, John. 2018. “Null-Hacking, a Lurking Problem in the Open Science Movement.” *PsyArXiv*. June 21. doi:10.31234/osf.io/9y3mp .
- Sances, Michael W. 2017. “Attribution Errors in Federalist Systems: When Voters Punish the President for Local Tax Increases.” *The Journal of Politics* 79 (4): 1286–1301.

Biographical Statements

Matthew H. Graham is a Ph.D. Candidate in the Department of Political Science at Yale University, New Haven, CT 06511. Gregory A. Huber is the Forst Family Professor of Political Science at Yale University, New Haven, CT 06511. Neil Malhotra is the Edith M. Cornell Professor of Political Economy in the Graduate School of Business at Stanford University, Stanford, CA 94305. Cecilia Hyunjung Mo is an Assistant Professor of Political Science at UC Berkeley, Berkeley, CA 94720.