

# Overt Consequences of Covert Actions\*

## Success, Failure, and Voters' Preferences for Legislative Oversight

Caroline Robbins	Bryanna Major	Omer Turkomer
Alessandro Brunelli	María Alemán Martínez	Luzmi Valenzuela
José Castro	Yadianis Lara Ojeda	Valeria Veras
Ainsley Coty	Larissa Pontes	Carlisle Rainey <sup>†</sup>
Andrew Louis	Elke Schumacher	

April 19, 2024

### Abstract

A growing literature in political science examines the tension between secrecy and accountability as democratic leaders conduct foreign affairs. However, this literature presents a puzzle: while observational work shows that leaders make efforts to avoid getting caught, experimental work demonstrates that the public does not have an aversion to leaders acting secretly. What, then, are the electoral costs of acting covertly? To demonstrate the electoral costs of covert operations, we study voters' reactions to *legislative oversight* of a covert operation, not their reactions to the operation itself. We suggest that the electoral costs of interest might occur indirectly through legislative oversight. Using a preregistered and well-powered survey experiment with a  $2 \times 2$  factorial design, we show that the legislators' audience (their in-party voters) overwhelmingly prefers public investigations into covert operations when the operation fails. When the operation succeeds, though, their preferences are much weaker. This dynamic has important implications for foreign policy—the president has an incentive to act too conservatively, shunning necessary risks to protect the national interest and to avoid investigations that harm the country's international reputation.

---

\*Our preregistration plan is available on OSF at <https://doi.org/10.17605/OSF.IO/GJSFT>. All data and code to reproduce our results are available on OSF at <https://doi.org/10.17605/OSF.IO/2TF5J>. A preprint of this paper is available at <https://doi.org/10.31235/osf.io/9p5h8>.

<sup>†</sup>Carlisle Rainey is Associate Professor of Political Science, Florida State University, 540 Bellamy, Tallahassee, FL, 32306. ([crainey@fsu.edu](mailto:crainey@fsu.edu)). Other co-authors are or were undergraduate students in the Research Intensive Bachelor's Certificate Program at Florida State University.

# 1 Introduction

A large, classic literature in political science focuses on how public opinion constrains leaders in international relations, mostly to their benefit (Fearon 1994; Tomz 2007). However, a growing body of work focuses on complications that arise when the leader can act secretly (Carnegie 2021). Secrecy can create a tactical advantage (Slantchev 2010; O’Rourke 2018), so it serves an essential role in national security policy. But leaders can exploit secrecy to avoid electoral accountability for unpopular actions (Colaresi 2014), and thus undermine the core democratic principle that leaders should be accountable to citizens for their actions. To speak to this tension, we focus on the incentives for legislative oversight of covert operations.

Observational work shows that presidents avoid covert action when they might get caught (Joseph and Poznansky 2018; Smith 2019), but experimental work shows that voters do not inherently mind secrecy in the conduct of foreign affairs (Myrick 2020; Carnegie, Kertzer, and Yarhi-Milo 2023). This raises the important puzzle of “when and why constituents care about particular covert actions, and how they punish interveners who get caught” (Joseph and Poznansky 2018, 333).

In this paper, we illuminate the political costs of covert actions by focusing on how voters evaluate *legislative oversight* following an exposed covert action. We use a well-powered, preregistered survey experiment with a  $2 \times 2$  factorial design to show that an opposition party’s audience—their in-party voters—do not have a preference for restrained, measured oversight. Instead, the opposition party strongly prefers to persistently antagonize the leader following a failure.

## 2 A Strategic Motivation for the Experiment

Potter and Baum (2014) ask a critical question: “Given the established low baseline of public attention, what prompts the mass public to sometimes pay enough attention to foreign policy

matters to impose political costs?” Their answer—which we echo—is that the *opposition party in the legislature can sound fire alarms* and prompt the public to pay attention. Observational work shows that presidents worry about political costs if covert operations are exposed (Joseph and Poznansky 2018; Smith 2019), which suggests an electoral cost. Experimental work, though, undermines the assumption that the public has a distaste for secrecy. Myrick (2020) and Carnegie, Kertzer, and Yarhi-Milo (2023) show that citizens do not object to leaders acting covertly, and sometimes prefer secrecy. This presents an important question, if voters do not mind secrecy itself, what is the electoral cost of secrecy, if any?

We argue that the institutional structure for oversight of covert operations—in which the legislature acts (or appears to act) *after* observing the policy outcomes—generates asymmetric information between the president and the legislature. This asymmetric information allows the opposition party in the legislature to adopt a critical posture *after the operation* and *if the operation fails* to exploit idiosyncratic failures to boost their own political standing. This informational asymmetry makes the president politically vulnerable when the operation is uncovered.

In the aftermath of a covert operation, the opposition party in the legislature must choose whether to amplify any whistleblower or not. We use a survey experiment to compare (1) *the improvement in approval of the opposition party from amplifying the whistleblower when the operation fails* to (2) *the improvement when the operation succeeds*. We have the following hypothesis:

**Conceptual Hypothesis:** When a covert operation fails, the opposition party’s supporters strongly prefer an investigation into the failed operation. When the operation succeeds, they have less appetite for investigation. Thus, the electoral stakes are high under failure and low under success.

An opposition party’s voters do not want restrained, measured oversight of the president,

but want legislators to punish the president politically when the operation fails. This creates a pathological incentive. The failed covert operation—regardless of the *ex ante* process or wisdom of the action—is like blood in the water. The electoral incentive to antagonize the leader is too strong for the legislature to conduct reasonable, restrained oversight. This incentive has meaningful consequences. First, public investigations unnecessarily expose operational secrets and damage the international reputation of the country. Second, the potential for legislative exploitation of a failed operation prevents the leader from taking potentially necessary risks. We do not test this logic—we only study the public’s taste for oversight—but this logic does *motivate* our experiment. Section 2 in the Supporting Information (S.I. 2) explains this logic in greater detail.

## 3 Design

### 3.1 Preregistration, Data, and Code

All hypotheses, measures, statistical tests, and computer code are preregistered on OSF (see anonymized PDF appended). We present all planned analyses as preregistered and a single additional analysis noted clearly below. S.I. 8 provides the data and S.I. 9 provides the code and output for the main results.

### 3.2 Participants

We fielded the experiment in March 2023 to 1,500 US respondents through CloudResearch Connect (CR). We paid 1,500 respondents \$1 to complete the survey. While CR respondents are not designed to be representative of the United States population, an ongoing literature consistently demonstrates that “results from convenience samples provide estimates of causal effects comparable to those found on population-based samples” (Mullinix et al. 2015, 122; see

S.I. 4.1 for more references). Top journals publish survey experiments using similar samples (e.g., Graham 2023 and Hansen 2023 in the *APSR*; see S.I. 4.2 for more).<sup>1</sup> S.I. 4 provides additional justification.

As specified in the preregistration, we exclude respondents who failed to answer a mock vignette question correctly, which is only 3% of the respondents (Kane, Velez, and Barabas 2023). S.I. 6 provides details and alternatives. This brings our sample to 1,462 respondents, which gives us excellent statistical power.<sup>2</sup>

### 3.3 Pretreatment Measures

Along with several other demographic characteristics, we first measure the respondent’s party identification on a seven-point scale using the traditional ANES approach.<sup>3</sup> We use this measure to assign the partisanship of the president and Congress in the vignette (the president is always the respondent’s out-party; Congress is always the in-party) and create a measure of partisan strength to use as a control variable.

### 3.4 Treatment

We use a 2 x 2 factorial vignette design, randomly assigning each respondent to read one of four vignettes.<sup>4</sup> The vignette describes a hypothetical covert operation that ends in either success or failure. Then, the vignette describes a whistleblower coming forward and the legislature chooses to either amplify or ignore the whistleblower. In our conceptual

---

<sup>1</sup>We highlight one particular concern. Prior work shows that respondents on Amazon’s MTurk are more liberal than the general population; we find this in our CR respondents as well. As a robustness check (not preregistered), we re-weighted the data to match partisanship in the 2022 Cooperative Election Study (CES). The results from the weighted analysis are not meaningfully different from the unweighted, preregistered analysis and small relative to the standard error estimates. See S.I. 4.5 for details. Below, we present only the unweighted, preregistered analysis.

<sup>2</sup>In brief, we have 84% power for a “substantively small” value of our smallest effect of interest, described in H2 below. For the details, see Section 14 of the preregistration.

<sup>3</sup>Sections 16 and 17 of the appended preregistration plan describe the full survey instrument in detail.

<sup>4</sup>Our vignette follows the style and structure of the vignette used by Carnegie, Kertzer, and Yarhi-Milo (2023).

framework, we imagine (1) that the legislature and the president are in opposition and (2) that the legislators are responding to their in-party voters. Building this into the vignette, we describe Congress as being controlled by the respondent’s party and the president as being in the respondent’s out-party. For example, for a Republican respondent, the vignette describes the legislature as “Republicans in Congress” and the president as the “Democratic president.” After a mock vignette check (Kane, Velez, and Barabas 2023), the vignette opens by describing the (always out-party) president’s decision.

*Country B posed a significant threat to US security and economic interests. It was a dictatorship without a history of respecting the rights of all of its citizens. Country B had a moderate-sized military and was neither a major trading partner nor an ally of the US.*

*After deliberation with his civilian, military, and intelligence advisors, the Republican [Democratic] president decided to authorize the use of a small military force to overthrow the government of Country B. To protect national security secrets and avoid damaging the United States’ international reputation, the president decided that the operation should be carried out secretly and should remain secret after its completion.*

Then the vignette describes the outcome of the operation. Respondents received either the success or failure condition below:

**[Success]** *Fortunately, the operation succeeded. The dictator was removed from power, there were no known civilian casualties, and all American operatives returned home.*

**[Failure]** *Unfortunately, the operation failed. The dictator remained in power, several civilians died, and four American operatives were killed in action.*

Then the vignette describes how a whistleblower came forward and exposed the operation.

*In the weeks after the covert operation, a whistleblower told a national newspaper about the operation. The newspaper broke the story with a front-page headline.*

Next, the vignette describes how (the always in-party) Congress responds to the whistleblower. Respondents received either the amplify or ignore condition below:

**[Amplify]** *Democrats [Republicans] in Congress publicly criticized the president's handling of the situation. They conducted an investigation, held public hearings, and issued a comprehensive public report that described the operation in detail. Democrats' criticisms of the president were commonly heard on the campaign trail.*

**[Ignore]** *Democrats [Republicans] in Congress remained largely silent on the president's handling of the situation. They held no public hearings and did not investigate the operation. Democrats' [Republicans'] criticisms of the president were rarely heard on the campaign trail.*

Finally, the vignette describes the (always out-party )president's response.

*The president declined to comment publicly or testify in a closed-door hearing. His press secretary emphasized the importance of maintaining secrecy in the interest of national security.*

### 3.5 Outcomes and Hypotheses

We follow up with several questions to measure respondents' evaluations of the actions that were taken in the vignette (and could have been taken). First, we ask respondents to evaluate the actions of Congress: "Overall, do you approve or disapprove of how Democrats [Republicans] in Congress handled the situation?" We present respondents with a seven-point

Likert response scale that ranges from strongly approve (+3) to strongly disapprove (-3). We have three hypotheses for this outcome. First, when a covert operation fails, we expect the opposition party’s supporters to strongly prefer an investigation into the failed operation.

**H1:** When the operation fails, Congress amplifying the whistleblower increases respondents’ overall approval of Congress’ actions (compared to Congress ignoring the whistleblower).

Second, even when the operation succeeds, we expect the opposition party’s supporters to have a mild preference for an investigation—acting otherwise would defer power to an out-party president. Even though we expect less appetite for investigating a successful operation, we still expect some.

**H2:** When the operation succeeds, Congress amplifying the whistleblower increases respondents’ overall approval of Congress’ actions (compared to Congress ignoring the whistleblower).

Third, we expect the opposition party’s supporters to have a stronger preference for investigating a failed operation than a successful operation.

**H3: [Interaction]** The effect of amplifying the whistleblower on respondents’ overall approval of Congress’ actions is larger when the operation fails than when the operation succeeds.

As a collection, these three hypotheses capture our core theoretical claim: the opposition party’s supporters strongly prefer an investigation into a failed operation, but have less appetite for an investigation into a successful operation.

Next, we ask respondents whether they think Congress *should* have taken various actions (regardless of their actions as described in the vignette): investigated the operation, held a



public hearing, released a comprehensive report, and criticized the president on the campaign trail. We ask respondents to “Please indicate your agreement or disagreement with the following statements about how Democrats [Republicans] in Congress handled the situations” and present four statements.

1. Democrats in Congress should have investigated the operation.
2. Democrats in Congress should have held a public hearing about the operation.
3. Democrats in Congress should have released a comprehensive report about the operation.
4. Democrats on the campaign trail should have criticized the president’s handling of the situation.

Respondents are given a seven-point Likert response scale that ranges from strongly agree (+3) to strongly disagree (-3). In each case, we hypothesize that respondents will report greater agreement (i.e., prefer amplification) when the operation fails. We formally test eight hypotheses (4 statements, tested separately in the amplify and ignore conditions). These eight hypotheses support our core hypotheses and serve as additional, indirect tests of our claim.

- **H4A/H4I:** Failure increases respondents’ agreement that Congress should have investigated the operation (compared to success). We test this hypothesis separately in the amplify (H4A) and ignore (H4I) conditions.
- **H5A/H5I:** Failure increases respondents’ agreement that Congress should have held a public hearing about the operation (compared to success).
- **H6A/H6I:** Failure increases respondents’ agreement that Congress should have released a comprehensive report about the operation (compared to success).
- **H7A/H7I:** Failure increases respondents’ agreement that Congress should have criticized the president on the campaign trail (compared to success).

Finally, we ask respondents to evaluate the actions of the president: “Overall, do you approve or disapprove of how the president handled the situation?” We present respondents

with a seven-point Likert response scale that ranges from strongly approve (+3) to strongly disapprove (-3). We use this outcome as a conceptual replication of Myrick (2020), who finds that failure has a large negative effect on approval of the *operation* (about one point on her seven-point Likert scale).

**H8A/H8I:** Failure decreases respondents’ overall approval of the president’s actions (compared to success). We test this hypothesis separately in the amplify (H8A) and ignore (H8I) conditions.

### 3.6 The Statistical Model

We estimate the average treatment effect using least squares regression and the standard error of the estimate with the HC3 Huber-White sandwich estimator (MacKinnon and White 1985). We include partisan strength as an interactive control (Lin 2013) and use 90% Wald confidence intervals to evaluate our directional hypotheses (McCaskey and Rainey 2015). Section 19 in the preregistration describes the details. S.I. 5 shows that the conclusions remain unchanged when we use no controls or include controls additively.

## 4 Results

We have three sets of hypotheses: evaluations of Congress’ actions as described in the vignette, evaluations of actions Congress could have taken (regardless of the vignette), and evaluations of the president’s actions. Figure 1 shows the estimated treatment effect and 90% confidence for the effects described by each hypothesis. S.I. 7 provides the full details for all regression estimates. We discuss each set of hypotheses in turn.

## 4.1 Evaluations of Congress' Actions as Described in the Vignette

We hypothesize that respondents will express greater overall approval of their party in Congress when the party amplifies the whistleblower, but this effect will be much larger when the operation fails than when the operation succeeds. The top panel of Figure 1 shows that the experiment strongly supports this conclusion.

[H1] The effect of amplification when the operation fails is about two points on the seven-point scale ( $Est. = 2.04$ ;  $90\% CI = [1.84, 2.24]$ ;  $Wald z\text{-statistic} = 16.82$ ). This effect is quite large—about two points on the seven-point scale. For comparison, (Myrick 2020) estimates that failure lowers approval of a covert operation by about one point on her seven-point scale.<sup>5</sup> This leads to a stark and surprising result: failing to amplify the whistleblower when the operation fails lowers approval of Congress by about *twice* as much as the operation itself failing lowers approval of the president (see the discussion of H8A/I for more on these relative effects.)

[H2] The effect of amplification when the operation succeeds is about one-half of a point on the seven-point scale ( $0.58$ ;  $[0.36, 0.79]$ ;  $z = 4.36$ ). As we hypothesize, amplification increases approval even when the operation succeeds. However, notice that this effect is *much* smaller.

[H3] Indeed, the effect of amplification when the operation fails is about 1.5 points larger on the seven-point scale than the effect of amplification when the operation succeeds ( $1.46$ ;  $[1.17, 1.76]$ ;  $z = 8.16$ ). Thus, amplification under failure increases approval by about *three times* more than amplification under success. These data offer strong support for our hypothesis that the electoral stakes are much higher for the legislature when the operation fails compared to when it succeeds.

---

<sup>5</sup>See the “Success” rows in her Tables 4, 5, and 6 and Figure 3.

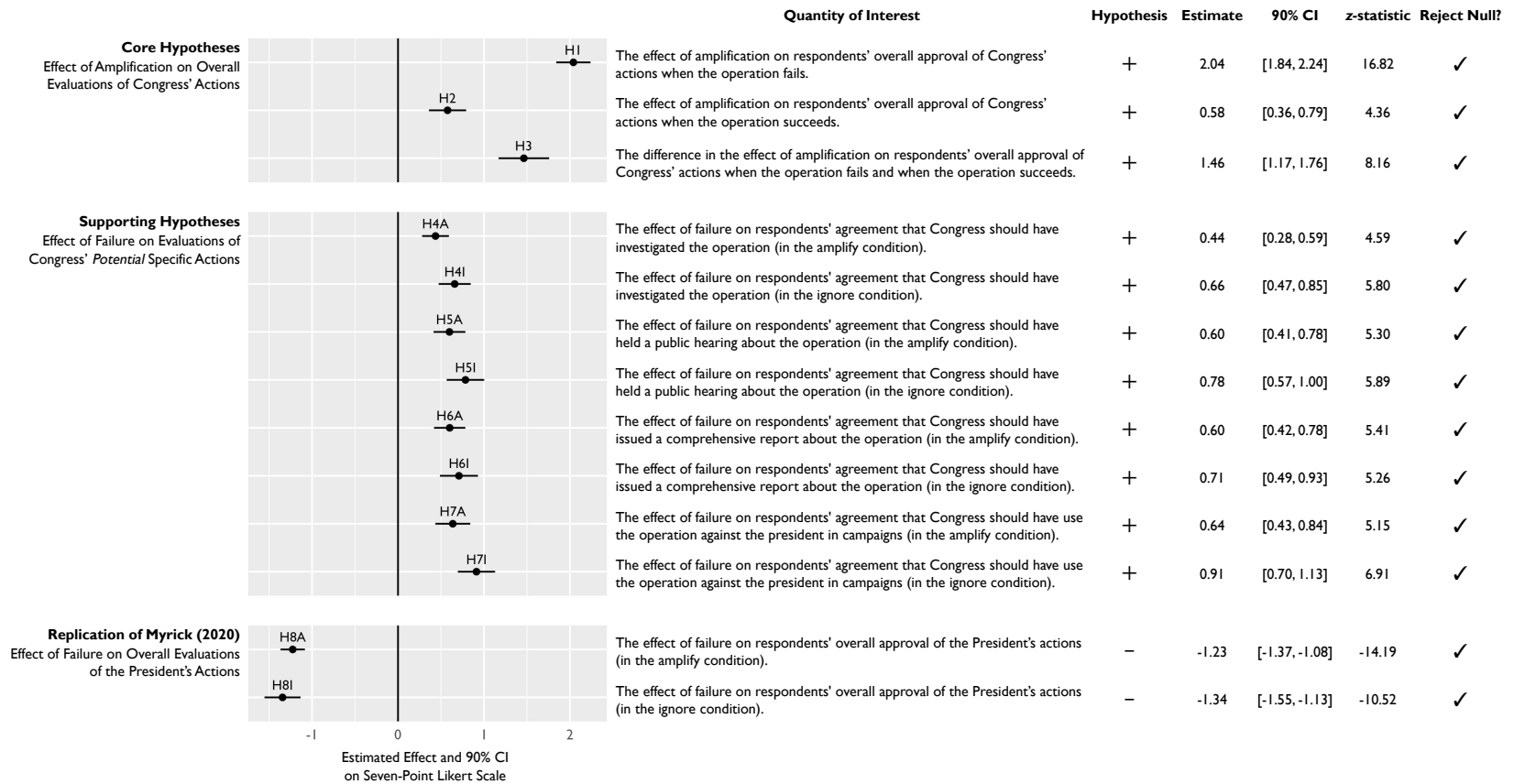


Figure 1: This figure summarizes the results from our preregistered analysis. We convincingly reject each null hypothesis. We conceptually replicate important experimental results from Myrick (2020) in H8A and H8I using evaluations of the president's actions in the vignette. But more importantly, we find even larger effects on evaluations of Congress' actions in the vignette when the operation fails. When the operation fails, respondents strongly prefer amplifying the whistleblower (H1). When the operation succeeds, preferences are much weaker (H2 and H3). Further, we demonstrate that these results are robust—regardless of the strategy Congress employs in the vignette, respondents prefer that Congress amplify the whistleblower using a range of tools (H4A through H7I). S.I. 7 provides the full details for all regression estimates.

## 4.2 Evaluations of Congress’ Potential Actions (Regardless of the Vignette)

[H4A through H7I] Next, we hypothesize that respondents will express greater agreement that Congress should take specific actions that amplify the whistleblower when the operation fails compared to when the operation succeeds: investigate the operation, hold a public hearing, issue a comprehensive report, and use the operation to campaign against the president. The vignette describes Congress as either taking action or not, but we ask specifically whether Congress *should have* done these things (regardless of the vignette). We specify a pair of hypotheses for each action (one for the amplify condition and another for the ignore condition) for eight total hypotheses. Figure 1 provides detailed statistics and shows that, in each case, failure increases support for the amplifying action compared to success. Differences across the actions and the conditions are small. The results suggest that failure increases respondents’ agreement that Congress should take a particular action by about three-quarters of a point on the seven-point scale.

## 4.3 Evaluations of the President’s Actions as Described in the Vignette

Lastly, we hypothesize that respondents will express lower overall approval of the president when the operation fails compared to when the operation succeeds. Again, we have separate hypotheses for the amplify and ignore conditions. These hypotheses serve two purposes. First, they allow us to conceptually replicate important results from (Myrick 2020). She writes that “the public primarily evaluates foreign action abroad on the basis of *policy outcomes*” (841, italics added) and she estimates a large effect of a little more than one point on the seven-point scale. Second, these results provide a useful benchmark to assess the effects of amplification. Myrick identifies a large effect of failure on approval of the operation. How

does approval of Congress amplifying the whistleblower compare?

[**H8A and H8I**] Although our experiment mimics Myrick’s only conceptually, Figure 1 shows that we successfully and convincingly replicate her results. When the vignette describes a Congress that amplifies the whistleblower, failure decreases respondents’ approval of the president by about 1.25 points on the seven-point scale ( $-1.23$ ;  $[-1.37, -1.08]$ ;  $z = -14.19$ ). When the vignette describes a Congress that ignores the whistleblower, the estimate is similar ( $-1.34$ ;  $[-1.55, -1.13]$ ;  $z = -10.52$ ). Thus, the experiment convincingly replicates Myrick’s (2020)—not only the direction of the effect but a surprisingly similar magnitude of about a 1.25-point decrease.<sup>6</sup>

[**Exploratory Analysis**] These estimates also serve as a useful baseline to highlight the effects of amplification on approval of Congress. We did not preregister this comparison, but cannot help but notice that the negative effect of failure on approval of the president is *smaller* than the negative effect of ignoring the whistleblower on approval of Congress when the operation fails. This gives important context for the results. The experiment suggests (though, again, this is not preregistered) that *Congress can have more at stake electorally than the president in the sequence of covert action and accountability*. We estimate that ignoring the whistleblower when the operation fails hurts Congress about 60% more than the president is hurt by failing in the first place ( $59\%$ ;  $90\%$  bootstrap  $CI = [41\%, 80\%]$ ). And while this hypothesis is not preregistered, the bootstrapped  $p$ -value for the null hypothesis that these effects are the same is effectively zero.

## 5 Conclusion

We address an important gap between theoretical and experimental work on electoral accountability in the domain of covert operations. The theoretical work specifies a precise mechanism

---

<sup>6</sup>Compare our estimates of  $-1.23$  and  $-1.37$  to her estimates in Tables 4, 5, and 6 and Figure 3.

to expose the covert operation to an unknowing public, such as a whistleblower (Spaniel and Poznansky 2018; Joseph and Poznansky 2018; Joseph, Poznansky, and Spaniel 2022). However, experimental work abstracts away from exposure and asks respondents to evaluate a series of choices by the leader (Myrick 2020; Carnegie, Kertzer, and Yarhi-Milo 2023). While it is helpful to understand how citizens evaluate actions taken secretly, citizens' evaluations remain inconsequential if actions remain secret. We help connect these two literatures by experimentally studying voters' evaluations of *legislative oversight*.

Our study of an explicit exposure mechanism is more than mundane realism. We emphasize the theoretical importance of the distinct information environments in secret foreign affairs of presidents—who choose policies using the information available at the time—and only legislators—who observe the outcomes before responding publicly. In short, we provide experimental oomph to Smith's (2019) evidence that leaders avoid covert operations when the opposition party is strong. Both the president and legislators would like to respond to voters. However, in the domain of covert action, voters care mostly about outcomes (Myrick 2020), and legislators observe the outcome before acting as “effective overseers” (Johnson 2018). We hypothesize and show that voters' preferences for legislative oversight depend strongly on the outcome. When the operation succeeds, voters have only a weak preference for oversight. In these situations, perhaps legislators pursue other political goals. However, when the operation fails, voters have a strong preference for oversight. The legislators face overwhelming political pressure to investigate and politically harm the president. As Rep. Trey Gowdy noted, the objective of such investigations is “not to find out what happened—it's to win” (Cogan 2015).

Scholars in international relations are increasingly careful to test the microfoundations of their theories (Tomz 2007; Kertzer 2017). However, public opinion influences policy through complex institutional arrangements. We encourage scholars of public opinion and foreign policy to think carefully about these institutions. In our case, explicitly considering the institutional mechanism through which voters can hold a president accountable (i.e.,

legislative oversight after the outcome is known) provides a richer understanding of the tension between democratic accountability and national secrets. Though beyond the scope of this paper, we suggest that the link between domestic audiences and national security is more complex than our theoretical and empirical models. Exploring these complexities will improve our understanding of international interactions.

## References

- Carnegie, Allison. 2021. “Secrecy in International Relations and Foreign Policy.” *Annual Review of Political Science* 24: 213–233.
- Carnegie, Allison, Joshua D. Kertzer, and Keren Yarhi-Milo. 2023. “Democratic Peace and Covert Military Force: An Experimental Test.” *Journal of Conflict Resolution* 67: 235–265.
- Cogan, Marin. 2015. “Trey Gowdy on Hillary Clinton, the PR Disasters of the Benghazi Committee, and Why He’s Had It With D.C.” *New York Magazine* .
- Colaresi, Michael P. 2014. *Democracy Declassified: The Secrecy Dilemma in National Security*. USA: Oxford University Press.
- Fearon, James D. 1994. “Domestic Political Audiences and the Escalation of International Disputes.” *American Political Science Review* 88 (3): 577–592.
- Graham, Matthew H. 2023. “Measuring Misperceptions?” *American Political Science Review* 117 (1): 80–102.
- Hansen, Kristina Jessen. 2023. “Greed, Envy, and Admiration: The Distinct Nature of Public Opinion About Redistribution From the Rich.” *American Political Science Review* 117 (1): 217–234.
- Johnson, Loch K. 2018. *Spy Watching*. New York, NY: Oxford University Press.



- Joseph, Michael F., and Michael Poznansky. 2018. "Media Technology, Covert Action, and the Politics of Exposure." *Journal of Peace Research* 55 (3): 320–335.
- Joseph, Michael F., Michael Poznansky, and William Spaniel. 2022. "Shooting the Messenger: The Challenge of National Security Whistleblowing." *Journal of Politics* 84 (2): 846–860.
- Kane, John V., Yamil R. Velez, and Jason Barabas. 2023. "Analyze the Attentive and Bypass Bias: Mock Vignette Checks in Survey Experiments." *Political Science Research and Methods* 11 (2): 293–310.
- Kertzer, Joshua D. 2017. "Microfoundations in International Relations." *Conflict Management and Peace Science* 34 (1): 81–97.
- Lin, Winston. 2013. "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique." *Annals of Applied Statistics* 7 (1): 295–318.
- MacKinnon, James G., and Halbert White. 1985. "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics* 29 (3): 305–325.
- McCaskey, Kelly, and Carlisle Rainey. 2015. "Substantive Importance and the Veil of Statistical Significance." *Statistics, Politics, and Policy* 6: 77–96.
- Mullinix, Kevin J., Thomas J. Leeper, James N. Druckman, and Jeremy Freese. 2015. "The Generalizability of Survey Experiments." *Journal of Experimental Political Science* 2 (2): 109–138.
- Myrick, Rachel. 2020. "Why So Secretive? Unpacking Public Attitudes toward Secrecy and Success in US Foreign Policy." *Journal of Politics* 82 (3): 828–843.
- O'Rourke, Lindsey A. 2018. *Covert Regime Change: America's Secret Cold War*. Ithaca, NY: Cornell University Press.

- Potter, Philip B. K., and Matthew A. Baum. 2014. "Looking For Audience Costs in All the Wrong Places: Electoral Institutions, Media Access, and Democratic Constraint." *Journal of Politics* 76 (1): 167–181.
- Slantchev, Branislav L. 2010. "Feigning Weakness." *International Organization* 64 (3): 357–388.
- Smith, Gregory L. 2019. "Secret but Constrained: The Impact of Elite Opposition on Covert Operations." *International Organization* 73 (3): 685–707.
- Spaniel, William, and Michael Poznansky. 2018. "Credible Commitment in Covert Affairs." *American Journal of Political Science* 62 (3): 668–681.
- Tomz, Michael. 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61 (4): 821–840.

# Supplementary Information (S.I.) for: Overt Consequences of Covert Actions

## Table of contents

<b>1</b>	<b>Discussion of Study Ethics</b>	<b>3</b>
<b>2</b>	<b>A Game to Motivate the Experiment</b>	<b>4</b>
2.1	Policy Payoffs . . . . .	4
2.2	Electoral Payoffs . . . . .	5
2.3	Payoffs and Equilibrium <u>When Electoral Payoffs Are Absent</u> . . . . .	5
2.4	Payoffs and Equilibrium <u>When Electoral Payoffs Are Present</u> . . . . .	5
2.5	Implications . . . . .	6
<b>3</b>	<b>Survey Instrument</b>	<b>8</b>
<b>4</b>	<b>Justification of the CloudResearch Sample</b>	<b>9</b>
4.1	Convenience Samples Work Well for Survey Experiments . . . . .	9
4.2	Similar Samples Appear in Top Journals . . . . .	9
4.3	CloudResearch Connect Respondents Are Attentive . . . . .	10
4.4	CloudResearch Connect Respondents Are Diverse . . . . .	10
4.5	Reweightings to Match the 2022 CES Produces Similar Estimates . . . . .	11
<b>5</b>	<b>Estimates Using Alternative Specifications</b>	<b>12</b>
<b>6</b>	<b>Estimates Using Alternative MVCs</b>	<b>13</b>
<b>7</b>	<b>Regression Tables</b>	<b>17</b>
7.1	H1, H2, and H3: Overall Approval of Congress . . . . .	18
7.2	H4: “Should Congress have investigated?” . . . . .	19
7.3	H5: “Should Congress have held a hearing?” . . . . .	20
7.4	H6: “Should Congress have issued a report?” . . . . .	21
7.5	H7: “Should Congress have campaigned on the operation?” . . . . .	22
7.6	H8: Overall Approval of the President . . . . .	23

<b>8</b>	<b>A Link to the Data</b>	<b>24</b>
8.1	Cleaned Data . . . . .	24
8.2	Raw Qualtrics Data . . . . .	25
<b>9</b>	<b>A Reproduction of Our Main Results</b>	<b>26</b>
	<b>References</b>	<b>31</b>

# 1 Discussion of Study Ethics

Our study was approved by our universities IRB and adheres to APSA's [Principles and Guidance for Human Subjects Research](#). Participants understood they were part of a research study and we did not use deception.

We obtained voluntary consent through a survey screen, which is shown in Figure 1.

**Consent**

---

You are being asked to voluntarily participate in a research study. We are doing this study to better understand how citizens evaluate elected officials. If you choose to participate, you will be asked to answer a few questions about yourself, read a description of a series of hypothetical events, and evaluate the elected officials described in those events. This will take about ten minutes.

If you agree to take part in this study, you will receive \$1.

We will not record your name or any information that shows your identity. You will not be signing this form.

If you have any questions, please contact [REDACTED] at [REDACTED], or [REDACTED]  
[REDACTED]

If you have any questions or concerns about your rights as a research participant, or questions or concerns regarding the study and would like to talk to someone other than the researcher(s), you are encouraged to contact the [REDACTED]  
[REDACTED] at [REDACTED], by email at [REDACTED], or by mail at [REDACTED]  
[REDACTED].

Figure 1: A screenshot of the consent form on the survey.

We paid respondents \$1 to complete the survey. The median completion time was about seven minutes; a pay rate of about \$8.50/hour. In our IRB documentation, we estimated a pay rate of about \$9.00/hour (above minimum wage for most of the US), and widely considered a generous rate among respondents.

## 2 A Game to Motivate the Experiment

Potter and Baum (2014) ask a critical question: “Given the established low baseline of public attention, what prompts the mass public to sometimes pay enough attention to foreign policy matters to impose political costs?” Their answer—which we echo—is that the opposition party in the legislature prompts the public to pay attention. So we build on their work and ask: “*When does the opposition party have an incentive to prompt the public to pay attention?*” Experimental work suggests that the policy process does not matter much; voters do not mind secrecy in the conduct of foreign affairs. Instead, voters care deeply about outcomes (Myrick 2020; Carnegie, Kertzer, and Yarhi-Milo 2022). In our experiment, we extend this prior work to voters’ preferences about *legislative oversight*. We connect theoretical and observational work that posits an explicit mechanism through which voters learn about covert actions (Michael F. Joseph, Poznansky, and Spaniel 2022; Michael F. Joseph and Poznansky 2017; Potter and Baum 2014) with experimental work that examines approval of covert actions in the abstract (Myrick 2020; Carnegie, Kertzer, and Yarhi-Milo 2022).

We use a simple extensive form game to motivate the experiment. Importantly, our experiment is not a formal test of the game; instead, we use the game to explain the critical importance of the voters’ responses to legislative oversight. The game summarizes the strategic interaction between a president and the opposition party in the legislature and shows why voters’ reactions to oversight might be important. Our experiment teaches us something important about politics—the game highlights the importance of our experimental results.

### ! Purpose of Game

Our experiment is not a formal test of the game; instead, we use the game to explain the critical importance of the voters’ responses to legislative oversight. Rather than testing implications of the game with our experiment, we use the game to highlight the implications of our experimental results.

The game occurs in three stages. In the first stage, the president decides whether to take a risky covert action or not. If the president does not take action, then all payoffs are normalized to zero. If the president does take action, then the action succeeds with probability  $p$  and fails with probability  $1 - p$ . If the president takes action, then we imagine that a whistleblower comes forward and the opposition party must choose whether to ignore the whistleblower or amplify the whistleblower.

### 2.1 Policy Payoffs

Along a policy dimension, both the president and opposition party have common preferences. They prefer a successful operation to no action to a failed operation. (This is absent the electoral incentives, which we discuss below.) We denote the policy payoffs for a successful

and failed operation as  $v_S$  and  $v_F$  such that  $v_S > 0 > v_F$ . The two actors also share a common preference to maintain operational secrets and avoid embarrassing the country by exposing the operation. We denote this reputational cost as  $R < 0$ .

## 2.2 Electoral Payoffs

Electoral, the four paths following the president taking action produce four different payoffs for both the opposition party and the president. We denote the electoral payoffs as  $E$ , use a superscript ( $P$  or  $O$ ) to denote the actor and a subscript ( $SA$ ,  $SI$ ,  $FA$ , or  $FI$ ) to denote the path. (Note that these path correspond the the 2 x 2 design of our experiment.) For example,  $E_{FA}^O$  represents Electoral payoff to the Opposition party when the operation Fails and they Amplify the whistleblower.

In our experiment, we are primarily interested in comparing (1) the marginal benefit of amplifying the whistleblower when the operation fails to (2) the marginal benefit when the operation succeeds. *These relative benefits have important implications for the equilibrium of the game.*

## 2.3 Payoffs and Equilibrium When Electoral Payoffs Are Absent

Solving by backward induction, the equilibrium depends on the size of the electoral payoffs. It is helpful to first solve the game assuming that the electoral payoffs are zero. In this simplified game, the president and opposition party share payoffs. Ignoring the whistleblower is always preferable for the opposition party, so the president will always act when the expected policy benefits are greater than zero. When the operation fails, the opposition always ignores the whistleblower, so the country never suffers the reputational damage and strategic cost of exposing the details of the operation. In this game, the payoff to the president and the opposition party is  $pv_S + (1 - p)v_F$  if this value is greater than zero. If this value is less than zero, then the president does not take action and the payoff is zero for both actors.

## 2.4 Payoffs and Equilibrium When Electoral Payoffs Are Present

When the electoral payoffs enter the game, the equilibrium becomes more interesting. Many equilibria are possible, depending on the size of the electoral payoffs relative to the reputational cost and policy payoffs. Solving by backward induction, the equilibrium depends on whether the marginal benefit to the opposition party of amplification is greater than the reputational and strategic cost. Our experiment focuses on these payoffs.

We have the following hypothesis:

When a covert operation fails, the opposition party’s supporters strongly prefer an investigation into the failed operation. When the operation succeeds, they have less appetite for investigation. In other words, the electoral stakes are high under failure and low under success.

In the analysis below, we assume that the marginal benefit of amplifying the whistleblower is greater than the reputational and strategic cost of exposing the operation only when the operation fails. Formally, we assume that  $E_{FA}^O - E_{FI}^O + R > 0$  but  $E_{SA}^O - E_{SI}^O + R < 0$ . That is, the opposition party amplifies the whistleblower when the operation fails and ignores the whistleblower when the operation succeeds. If this assumption holds—and our experiment supports this assumption—what are the equilibrium strategies and payoffs in the game?

First, the president’s payoff for acting becomes  $pv_S + E_{SI}^P + (1-p)v_F + E_{FA}^P + R$ . We can assume that the electoral benefit of a successful covert operation (that remains secret) is about zero, so this simplifies to  $pv_S + (1-p)v_F + E_{FA}^P + R$ . Notice how this compares to the president’s payoff without electoral payoffs. We can rewrite this payoff as  $pv_S + (1-p)v_F + (1-p)E_{FA}^P + R$ . The component in the first bracket is exactly the president’s payoff for acting in the game without electoral payoffs. The component in the second bracket is an additional cost. The opposition party is now willing to amplify the whistleblower and expose operational details. This will no doubt hurt the president electorally so that  $E_{FA}^P < 0$  (Kriner and Schickler 2014), but also harm the country’s reputation and strategic advantage.

## 2.5 Implications

This equilibrium, which is supported by our experiment, identifies two important implications of the hypothesis above. First, in order to take action, the benefits of a successful operation must be much larger than in the game without electoral payoffs. That is, *the president can no longer take calculated risks*. The president acts too conservatively, which reduces the payoffs for both the president and the opposition party compared to the game with no electoral payoffs. Second, the electoral payoffs cause the opposition party to amplify the whistleblower in some instances. In the game without electoral payoffs, the opposition always ignores the whistleblower. When we introduce the electoral payoffs, *the opposition party will sometimes choose to damage the international reputation of the country and expose tactically advantageous secrets*.

The game we describe in technical language above has a simple and intuitive structure. The nature of covert operations and legislative oversight requires that legislators act after observing the outcome of the operation. Previous work shows that voters care much more about failed operations than secret operations (Myrick 2020). This implies that voters expect legislators to punish the executive politically when the operation fails. The high electoral stakes when the operation fails means that the political gain from punishing the executive exceeds the benefit from preserving operational secrets or maintaining the country’s international reputation. The



relatively lower electoral stakes when the operation succeeds means that the political gains might be relatively small and warrant legislative discretion.

### **3 Survey Instrument**

The pretreatment measures, treatments, and posttreatment measures are described in detail in Sections 16 and 17 of the preregistration.

## 4 Justification of the CloudResearch Sample

### 4.1 Convenience Samples Work Well for Survey Experiments

While we do not intend for our CloudResearch Connect respondents to be representative of the United States population, an on-going literature consistently demonstrates that “results from convenience samples provide estimates of causal effects comparable to those found on population-based samples” (Mullinix et al. 2015, 122). Given our goal of estimating treatment effects from survey experiments, prior work shows that convenience samples work very well. The claim that convenience samples replicate the results of very expensive, nationally representative surveys is remarkably robust (Berinsky, Huber, and Lenz 2012; Krupnikov and Levine 2014; Weinberg, Freese, and McElhattan 2014; Thomas and Clifford 2017). Coppock, Leeper, and Mullinix (2018) explain that most survey experiments lack sufficient heterogeneity in treatment effects for different estimates to emerge due to small differences between diverse crowdsourced convenience samples and nationally representative samples.

Indeed, in a survey experiment studying responses to covert actions, Carnegie, Kertzer, and Yarhi-Milo (2022) write:

We used this platform [Amazon’s MTurk, in their case] because these surveys can be run faster and at lower expense while still being “more representative of the general population” than other convenience samples frequently employed by researchers (Berinsky, Huber, and Lenz 2012). Indeed, many researchers have shown that MTurk can successfully replicate other published studies that used nationally representative samples and that the MTurk population behaves similarly to other populations” (243).

In short, research consistently shows that survey experiments using convenience samples and nationally representative samples produce similar estimates. Among convenience samples, crowdsourced samples like MTurk and CloudResearch Connect are perhaps the most diverse.

### 4.2 Similar Samples Appear in Top Journals

Many crowdsourced survey experiments appear in recent issues of top journals. See examples in the *APSR* (Hansen 2022; Graham 2022; Graham and Singh 2023; Zárate, Quezada-Llanes, and Armenta 2023; Kreft and Agerberg 2023; Pérez, Robertson, and Vicuña 2022), *AJPS* (Kim 2022; Virgin 2021; Kobayashi et al. 2023; Powers and Altman 2022; Singh and Tir 2021; Graham 2023), and *JOP* (Andrews, Delton, and Kline 2023; Renshon, Yarhi-Milo, and Kertzer 2023; Abrajano, Elmendorf, and Quinn 2023; Chow and Han 2023; Velez, Porter, and Wood 2023; Elder and Hayes 2023).

### 4.3 CloudResearch Connect Respondents Are Attentive

To assess the attentiveness of respondents, we use a mock vignette check from Kane, Velez, and Barabas' (2023) collection of validated mock vignettes. We use their Mock Vignette 1 on same-day voter registration and ask all three follow-up questions. 97% of our respondents correctly answered the first question, compared to 81% in the NORC sample. For the second and third questions, 79% and 40% of our respondents answered the questions correctly, compared to 36% and 47% in the NORC sample. Our respondents spent about 46 seconds on the vignette screen, on average, compared to about 38 seconds in the NORC sample. Our survey concluded with optional open-ended questions and many respondents chose to provide thoughtful, detailed answers.

### 4.4 CloudResearch Connect Respondents Are Diverse

Our sample is not intended to be a nationally representative sample. Prior research shows that online convenience samples tend to be more liberal (Levy, Freese, and Druckman 2016). Indeed, we find that strong Republicans are under-represented in our data (compared to 2022 CES). Figure 2 compares the distribution of partisanship in our CloudResearch sample, the unweighted 2022 CES, and the 2022 CES weighted (using their sample weights). While we would not claim that the partisanship of our sample closely matches the CES, we would claim that our convenience sample is *diverse*. We obtain a large number of respondents across the partisan spectrum.

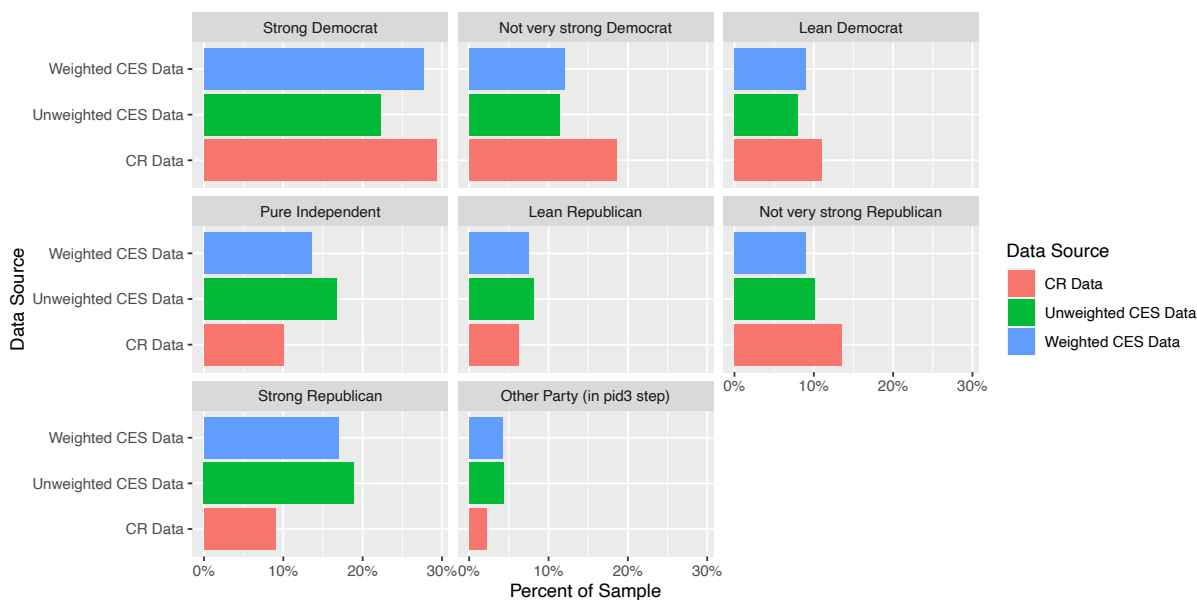


Figure 2: A figure showing the distribution of partisan identification for the CloudResearch sample, the unweighted 2022 CES, and the weighted 2022 CES.

## 4.5 Reweighting to Match the 2022 CES Produces Similar Estimates

To address the concern that a nationally representative sample might have produced different treatment effects, we can reweight our CloudResearch sample to match the 2022 CES distribution of partisanship. In practice, this means weight the responses of strong Republicans in our sample higher and not very strong Democrats lower. Figure 3 shows that this approach changes our estimates very little.

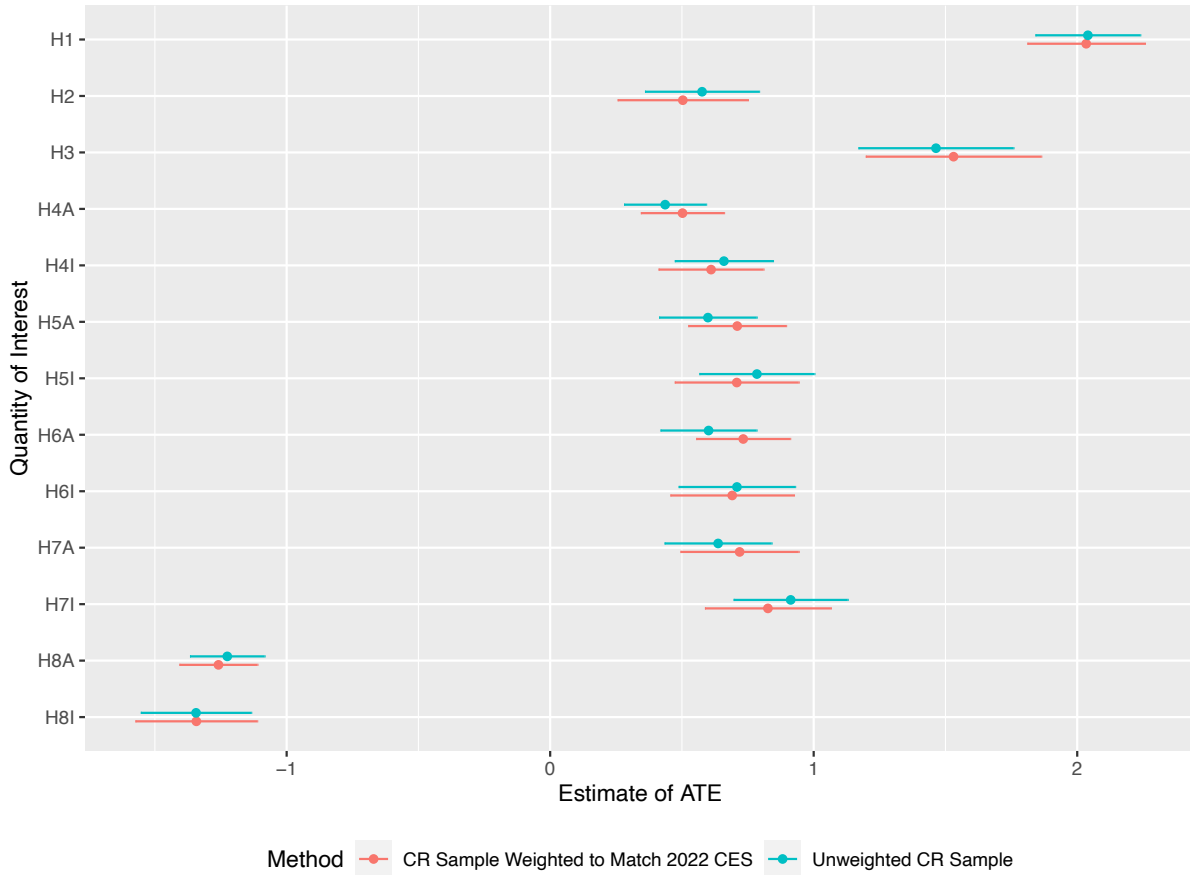


Figure 3: A figure showing the estimate from our CloudResearch data and estimates weighted those data to match the distribution of partisan identification in the 2022 CES.

## 5 Estimates Using Alternative Specifications

### **i** The Estimates Are Robust to Alternative Specifications

The estimates of interest reported in Figure 1 of the paper do not change when controlling for partisan strength additively or using no controls.

Our main specification, presented in Figure 1 in the paper, includes partisan strength as an interactive control (Lin 2013). In the figure below, we also provide the estimates using partisan strength as an additive control and using no controls. Figure 4 shows that the estimates of interest do not change depending on the specification.

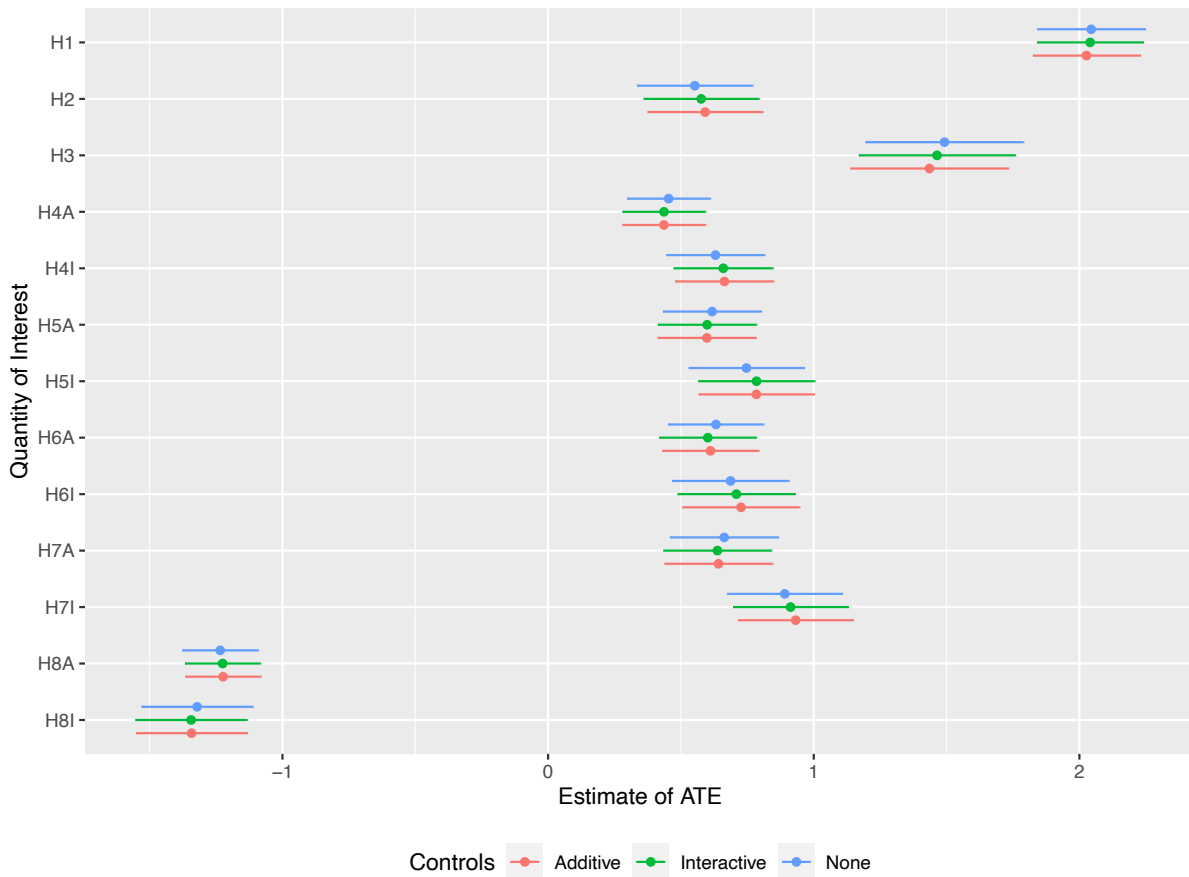


Figure 4: A figure showing the ATE estimates using our preregistered control specification (interactive, following Lin 2013), an additive specification, and a specification with no controls. Results do not change.

## 6 Estimates Using Alternative MVCs

### **i** The Estimates Are Robust to Alternative Attention Checks

With two reasonable exceptions, the estimates of interest reported in Figure 1 of the paper do not change when using alternative exclusion rules for the mock vignette check.

In this section, we consider exclude respondents using alternative thresholds from the mock vignette checks. We used the following validated mock vignette from Kane, Velez, and Barabas (2023).

Many state legislatures are currently considering enacting "same-day registration" policies, which would allow residents of the state (who are eligible to vote) to register and vote within the same day.

As with any policy, there are many factors to consider, including factors regarding potential costs of implementation. According to the National Conference of State Legislatures, "Same day registration procedures vary within states, and so costs vary as well. Some states indicate there is little to no additional cost in implementing same day registration, especially those that have had this option available for a long time. Some costs that may be associated with implementing same day registration include increased election staff or poll workers to process same day registrations. This extra administrative task can be time consuming at the same day registration site and when verifying registration information after the election. Many states report this is more a reallocation of costs and resources, though, rather than an additional cost."

After asking respondents to read the vignette, we present the following factual questions.

Which policy area was discussed in the article you just read?

- ☐ Voter identification policies
  - ☐ Voting age policies
  - ☐ Voter registration policies
  - ☐ Voting privacy policies
  - ☐ Voting location policies
- 

In the article you just read, which specific organization was quoted regarding "same-day registration" policies?

- ☐ Bureau of Legislative and Electoral Processes
  - ☐ State Board of Elections
  - ☐ National Conference of State Legislatures
  - ☐ National Governors Group
  - ☐ Council on Foreign Relations
- 

According to the article you just read, many states report that:

- ☐ There is now "significantly greater gubernatorial oversight of the voting process"
- ☐ They have "recently reversed their position on voting registration policies"
- ☐ "Increased election staff" may be one cost of implementation
- ☐ Policymakers "do not believe this policy will significantly change voter turnout"
- ☐ Same-day registration has "resulted in many more voters coming to the polls"

Based on pretesting, we expected that almost all our CloudResearch respondents would be sufficiently attentive. Based on timing data, open-ended responses, and attention checks, only about 1 in 50 respondents seemed to rush through the survey. We preregistered the decision to exclude respondents who did not answer the first question correctly, which is the easiest. In pretesting, only about 3 to 5% of respondents failed this check, and this indicator seems to correspond to respondents who rushed through the survey.

The variables `passed_mvc1`, `passed_mvc2`, and `passed_mvc3` indicate respondents who answered the three questions above correctly. 97% of respondents answered MVC 1 correctly, 40% answered MVC 2 correctly, and 79% answered MVC 3 correctly. In each case, our CloudResearch respondents out-performed the respondents in Kane, Velez, and Barabas' (2023)



high-quality NORC sample. Compare our percent-correct of 97%, 79%, and 40% to Kane, Velez, and Barabas' percent-correct of 81%, 36%, and 47%. Our respondents spent about 46 seconds on the vignette screen, on average, compared to about 38 seconds in Kane, Velez, and Barabas' NORC sample.

Figure 5 shows that the results are robust to excluding different sets of respondents, except when we exclude the 60% of respondents to did not correctly answer the very difficult MVC 2. In that extreme case, we fail to reject the null hypothesis for H4A and H5A (but reject the null in all other cases).<sup>1</sup> For the reasonable strategies of including all data, excluding respondents who missed the not-very-difficult MVC 1, and excluding respondents who missed the moderately-difficult MVC 3, the results are nearly identical.

---

<sup>1</sup>We do not consider dropping 60% of respondents a reasonable strategy to consider here. We included this item to compare our results to Kane, Velez, and Barabas (2023).

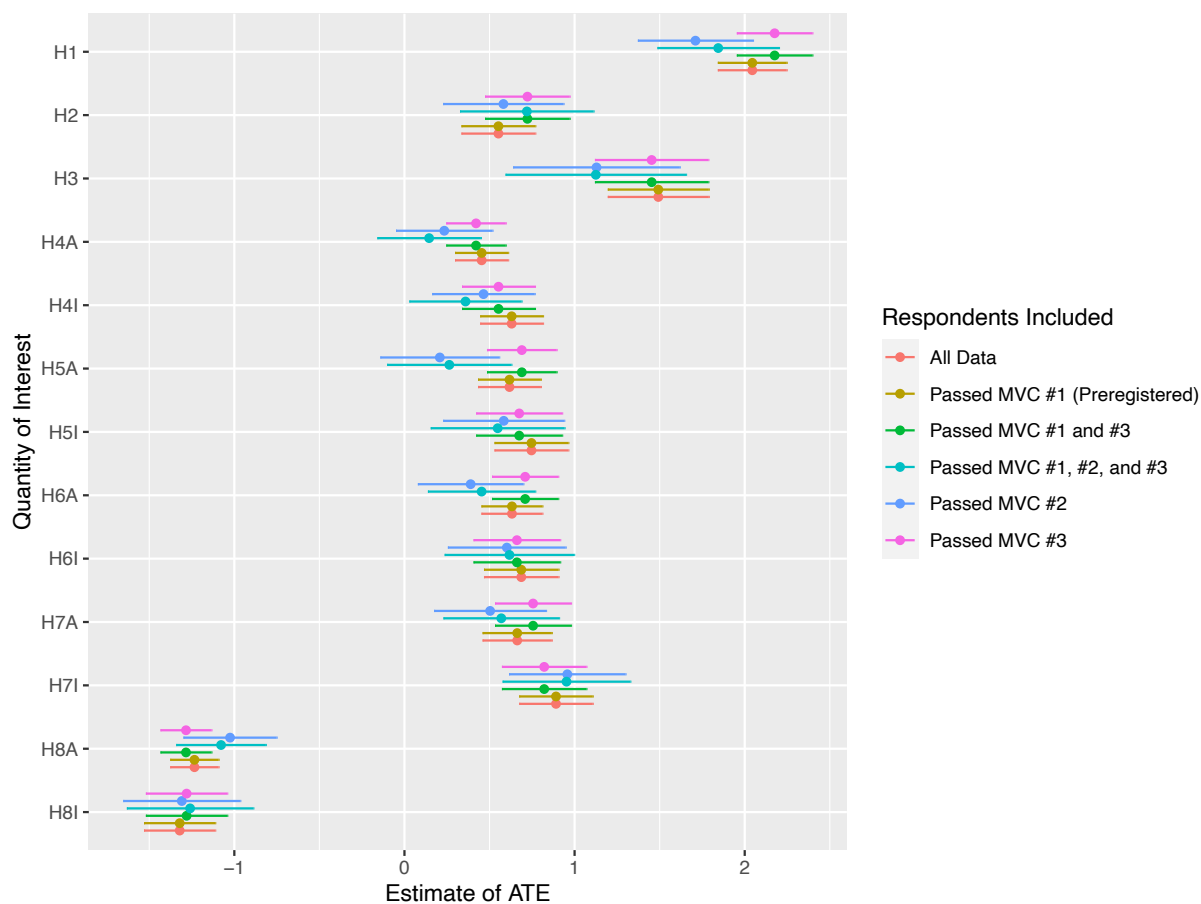


Figure 5: A figure showing the ATE estimates using our preregistered inclusion criterion (passed MVC 1) and other possible criteria. Except in the extreme case of dropping the 60% of respondents who answered the very difficult MVC 2 incorrectly, the results do not change.

## 7 Regression Tables

Below, we provide the full regression tables for the estimates summarized in Figure 1 of the paper and Figure 4 above. The figures allow easy comparison across models and the tables provide the precise numerical estimates and robust standard errors.

### Details of Statistical Models

Section 19 of the preregistration describes the statistical model in detail and how to compute quantities of interest and confidence intervals, but theoretically and in R.

In these tables and figures below, these coefficients are the key quantities of interest.

1. Intercept
2. Failure Indicator (FI)
3. Amplify Indicator (AI)
4. FI x AI

All other coefficients are nuisance parameters. Because we center our control variable(s), the intercept and coefficients for the treatment indicators have the same interpretation as a model with no controls.

### Important Point about Interpretation

Because we center our control variable(s), the intercept and coefficients for the treatment indicators have the same interpretation as a model with no controls.

## 7.1 H1, H2, and H3: Overall Approval of Congress

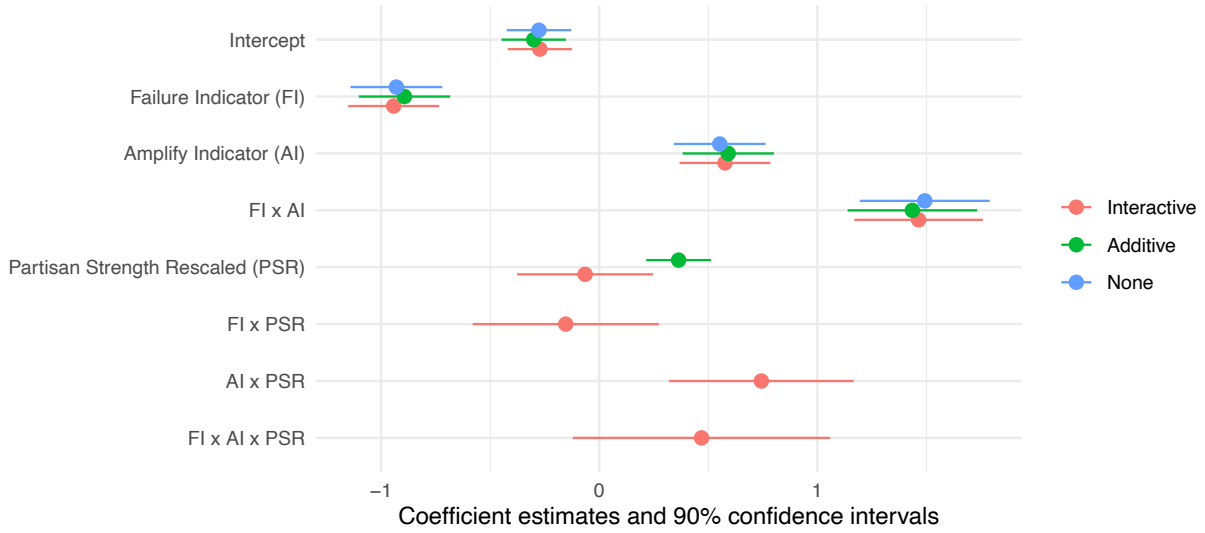


Figure 6: The regression estimates and robust 90% confidence intervals. Outcome variable is overall approval of Congress.

Table 1: The regression estimates and robust standard errors. Outcome variable is overall approval of Congress.

	Interactive		Additive		None	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
<b>Coefficients of interest</b>						
Intercept	-0.27**	0.10	-0.30**	0.10	-0.28**	0.10
Failure Indicator (FI)	-0.94***	0.13	-0.89***	0.14	-0.93***	0.13
Amplify Indicator (AI)	0.58***	0.13	0.59***	0.13	0.55***	0.13
FI x AI	1.46***	0.18	1.44***	0.18	1.49***	0.18
<b>Nuisance coefficients</b>						
Partisan Strength Rescaled (PSR)	-0.06	0.22	0.36***	0.09		
FI x PSR	-0.15	0.28				
AI x PSR	0.74**	0.28				
FI x AI x PSR	0.47	0.37				
Num.Obs.	1462		1462		1462	
RMSE	1.70		1.72		1.73	
R2	0.187		0.169		0.160	
AIC	5717.1		5742.9		5757.1	

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

## 7.2 H4: “Should Congress have investigated?”

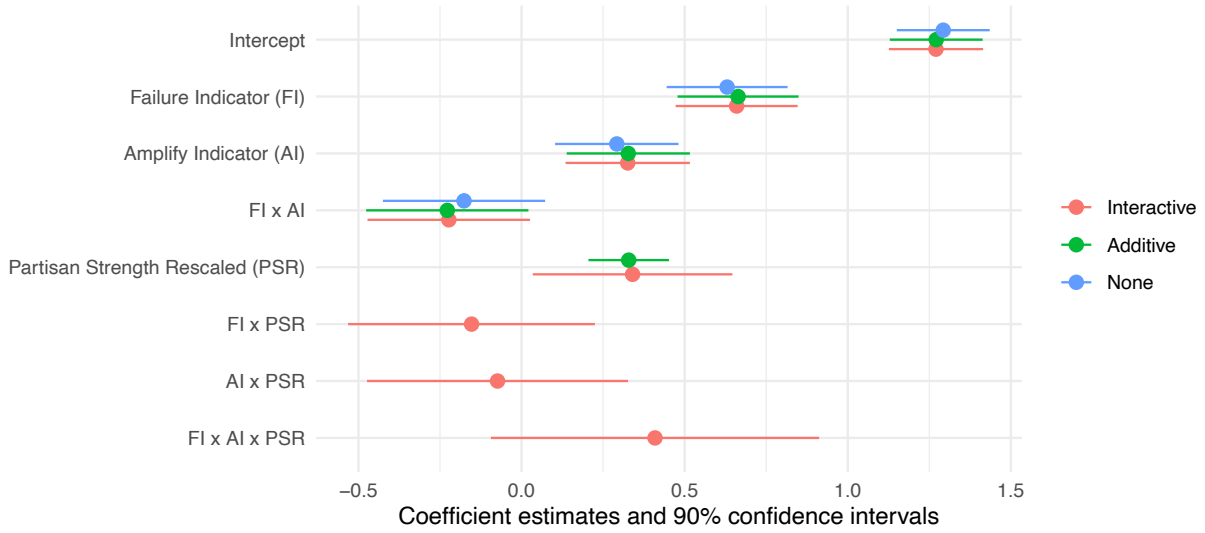


Figure 7: The regression estimates and robust 90% confidence intervals. Outcome variable is agreement that Congress should have investigated.

Table 2: The regression estimates and robust standard errors. Outcome variable is agreement that Congress should have investigated.

	Interactive		Additive		None	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
<b>Coefficients of interest</b>						
Intercept	1.27***	0.09	1.27***	0.09	1.29***	0.09
Failure Indicator (FI)	0.66***	0.11	0.66***	0.11	0.63***	0.11
Amplify Indicator (AI)	0.33**	0.12	0.33**	0.11	0.29*	0.11
FI x AI	-0.22	0.15	-0.23	0.15	-0.18	0.15
<b>Nuisance coefficients</b>						
Partisan Strength Rescaled (PSR)	0.34+	0.19	0.33***	0.07		
FI x PSR	-0.15	0.23				
AI x PSR	-0.07	0.24				
FI x AI x PSR	0.41	0.31				
Num.Obs.	1462		1462		1462	
RMSE	1.43		1.43		1.44	
R2	0.054		0.052		0.040	
AIC	5215.8		5212.6		5229.5	

+  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

### 7.3 H5: “Should Congress have held a hearing?”

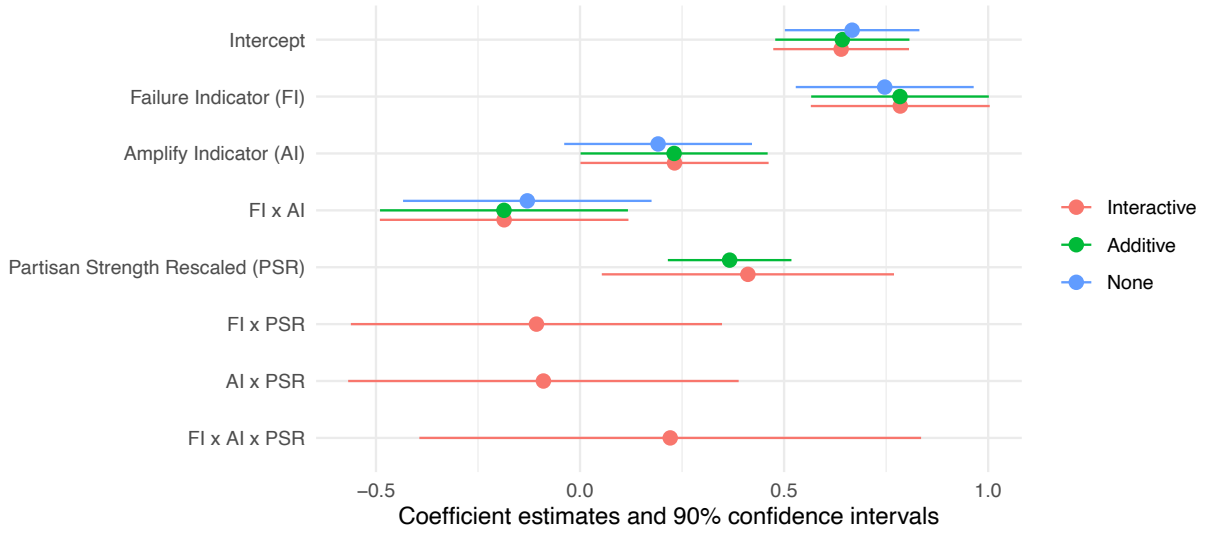


Figure 8: The regression estimates and robust 90% confidence intervals. Outcome variable is agreement that Congress should have held a hearing.

Table 3: The regression estimates and robust standard errors. Outcome variable is agreement that Congress should have held a hearing.

	Interactive		Additive		None	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
<b>Coefficients of interest</b>						
Intercept	0.64***	0.10	0.64***	0.10	0.67***	0.10
Failure Indicator (FI)	0.78***	0.13	0.78***	0.13	0.75***	0.13
Amplify Indicator (AI)	0.23+	0.14	0.23+	0.14	0.19	0.14
FI x AI	-0.19	0.19	-0.19	0.18	-0.13	0.18
<b>Nuisance coefficients</b>						
Partisan Strength Rescaled (PSR)	0.41+	0.22	0.37***	0.09		
FI x PSR	-0.11	0.28				
AI x PSR	-0.09	0.29				
FI x AI x PSR	0.22	0.37				
Num.Obs.	1462		1462		1462	
RMSE	1.76		1.76		1.76	
R2	0.048		0.048		0.038	
AIC	5811.8		5806.2		5819.9	

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

## 7.4 H6: “Should Congress have issued a report?”

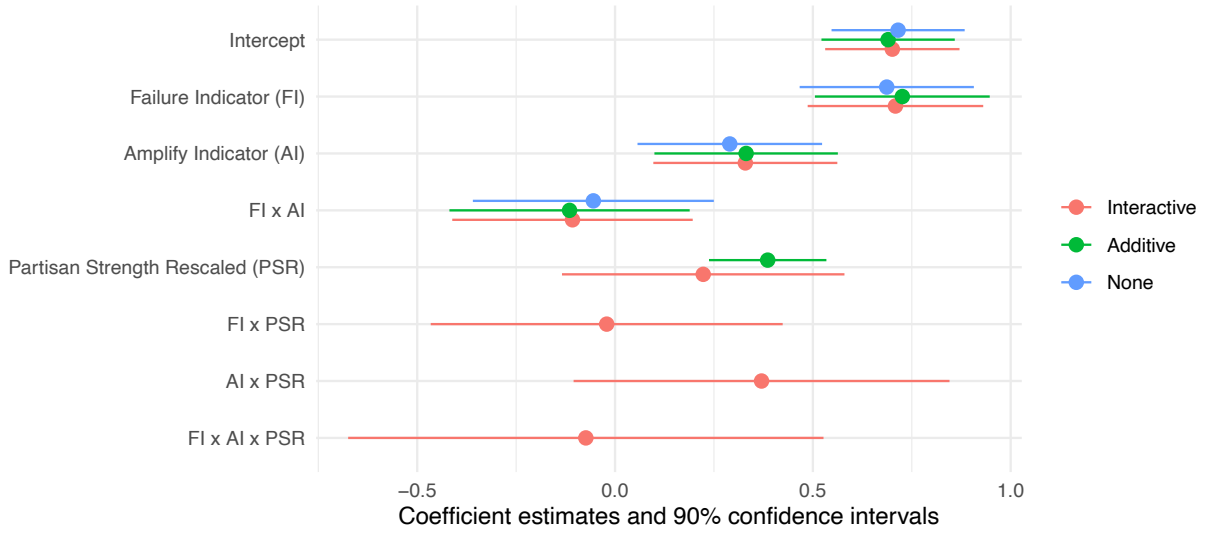


Figure 9: The regression estimates and robust 90% confidence intervals. Outcome variable is agreement that Congress should have released a report.

Table 4: The regression estimates and robust standard errors. Outcome variable is agreement that Congress should have released a report.

	Interactive		Additive		None	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
<b>Coefficients of interest</b>						
Intercept	0.70***	0.10	0.69***	0.10	0.72***	0.10
Failure Indicator (FI)	0.71***	0.13	0.73***	0.13	0.69***	0.13
Amplify Indicator (AI)	0.33*	0.14	0.33*	0.14	0.29*	0.14
FI x AI	-0.11	0.18	-0.11	0.18	-0.05	0.19
<b>Nuisance coefficients</b>						
Partisan Strength Rescaled (PSR)	0.22	0.22	0.39***	0.09		
FI x PSR	-0.02	0.27				
AI x PSR	0.37	0.29				
FI x AI x PSR	-0.07	0.36				
Num.Obs.	1462		1462		1462	
RMSE	1.75		1.76		1.77	
R2	0.053		0.050		0.039	
AIC	5808.7		5806.1		5821.6	

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

## 7.5 H7: “Should Congress have campaigned on the operation?”

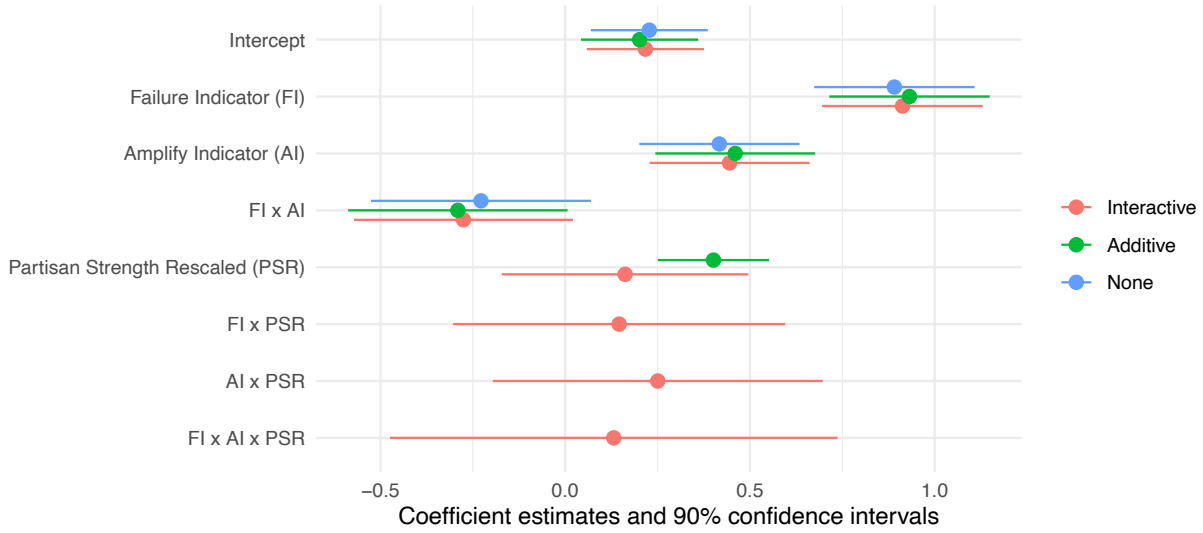


Figure 10: The regression estimates and robust 90% confidence intervals. Outcome variable is agreement that Congress should have campaigned against the president.

Table 5: The regression estimates and robust standard errors. Outcome variable is agreement that Congress should have campaigned against the president.

	Interactive		Additive		None	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
<b>Coefficients of interest</b>						
Intercept	0.22*	0.10	0.20*	0.10	0.23*	0.10
Failure Indicator (FI)	0.91***	0.13	0.93***	0.13	0.89***	0.13
Amplify Indicator (AI)	0.44***	0.13	0.46***	0.13	0.42**	0.13
FI x AI	-0.28	0.18	-0.29	0.18	-0.23	0.18
<b>Nuisance coefficients</b>						
Partisan Strength Rescaled (PSR)	0.16	0.20	0.40***	0.09		
FI x PSR	0.15	0.27				
AI x PSR	0.25	0.27				
FI x AI x PSR	0.13	0.37				
Num.Obs.	1462		1462		1462	
RMSE	1.71		1.71		1.72	
R2	0.072		0.069		0.056	
AIC	5736.9		5735.4		5753.1	

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001



## 7.6 H8: Overall Approval of the President

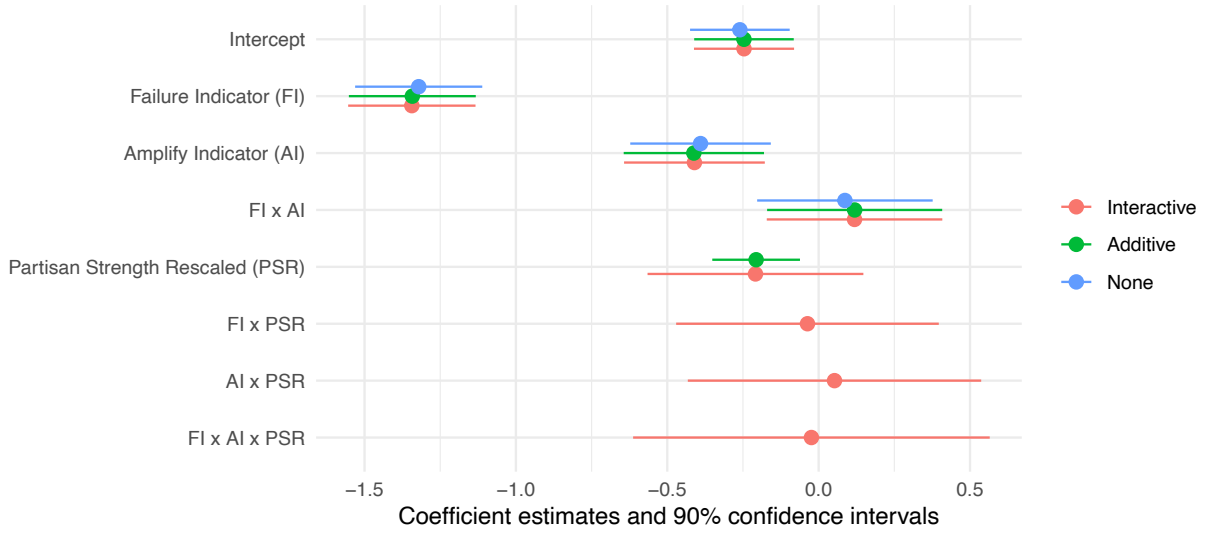


Figure 11: The regression estimates and robust 90% confidence intervals. Outcome variable is approval of the president.

Table 6: The regression estimates and robust standard errors. Outcome variable is approval of the president.

	Interactive		Additive		None	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
<b>Coefficients of interest</b>						
Intercept	-0.25*	0.10	-0.25*	0.10	-0.26**	0.10
Failure Indicator (FI)	-1.34***	0.13	-1.34***	0.13	-1.32***	0.13
Amplify Indicator (AI)	-0.41**	0.14	-0.41**	0.14	-0.39**	0.14
FI x AI	0.12	0.18	0.12	0.18	0.09	0.18
<b>Nuisance coefficients</b>						
Partisan Strength Rescaled (PSR)	-0.21	0.22	-0.21*	0.09		
FI x PSR	-0.04	0.26				
AI x PSR	0.05	0.29				
FI x AI x PSR	-0.02	0.36				
Num.Obs.	1462		1462		1462	
RMSE	1.68		1.68		1.68	
R2	0.138		0.138		0.134	
AIC	5680.9		5675.1		5678.6	

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

## 8 A Link to the Data

### 8.1 Cleaned Data

Users can obtain the cleaned data using the code below. These data are cleaned exactly as described in the preregistration.

```
# load packages
library(tidyverse)

# read cleaned data from dropbox
coversight <- read_rds("https://www.dropbox.com/s/9grn8kkb5yzwagx/data.rds?raw=1")%>%
  filter(passed_mvc1) %>% # remove respondents who failed MVC 1
  mutate(pid_strength_rs = arm::rescale(pid_strength)) # center partisan strength

# quick look
glimpse(coversight)
```

Rows: 1,462

Columns: 25

```
$ respondent_id    <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1~
$ cong_party       <chr> "Democrat", "Republican", "Democrat", "Democrat", "R~
$ pid2             <chr> "Democrat", "Republican", "Democrat", "Democrat", "R~
$ pid7             <fct> Strong Democrat, Strong Republican, Strong Democrat,~
$ pid_strength     <dbl> 3, 3, 3, 0, 2, 2, 3, 3, 2, 3, 3, 3, 2, 0, 0, 1, 2, 2~
$ amplify          <fct> Amplify, Ignore, Amplify, Ignore, Ignore, Amplify, A~
$ failure          <fct> Failure, Failure, Failure, Failure, Success, Success~
$ pres_words       <chr> "Great man ", "I feel that things could have been do~
$ pres_overall     <dbl> 0, 1, -3, 0, 2, 3, 2, -3, -3, 0, -3, -3, 1, 0, 2, -2~
$ pres_operation   <dbl> -1, -1, -3, 0, 0, 3, 2, -3, -3, 0, -3, -3, 1, 0, 2, ~
$ pres_secret      <dbl> -2, 2, -3, 0, 0, 3, -2, -3, 1, -3, -3, -3, 2, 0, 2, ~
$ pres_comment     <dbl> 0, 0, 3, 1, 1, 1, 3, 3, 3, 3, 3, 3, 2, 0, 1, 2, 2, 3~
$ pres_testify     <dbl> 3, 2, 3, 1, 1, 2, -1, 3, 2, -3, 1, 2, 2, 0, -2, -1, ~
$ cong_words       <chr> "Strong community ", "Libtards", "just", "Silence", ~
$ cong_overall     <dbl> 2, 1, 2, -1, 2, -2, -1, -3, 2, 0, -3, -2, -1, 0, -2,~
$ cong_investigate <dbl> 1, 1, 3, 1, 0, -2, 2, 3, 3, 3, 3, 3, 2, 0, 1, 1, -1,~
$ cong_hearing     <dbl> -1, 2, 3, 0, 0, -2, -1, 3, 2, 3, 3, 2, -1, 0, -1, 1,~
$ cong_report      <dbl> 0, -1, 3, 0, 0, -3, 1, 3, 2, 3, 3, 2, 2, 0, -1, 1, 0~
$ cong_campaign    <dbl> 0, 1, 3, 0, 0, -2, 0, 3, 2, 3, 3, 3, 0, 0, -1, 0, -1~
$ whistle_overall  <dbl> -2, 0, 3, 0, 2, -2, -1, 3, 2, 0, 2, 3, -1, 0, 0, 1, ~
$ newspaper_overall <dbl> 1, 0, 2, 0, -2, -3, -1, 3, 1, 0, 2, 3, 0, 0, 0, 0, 0~
$ passed_mvc1     <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE~
```

```
$ passed_mvc2      <lgl> FALSE, TRUE, FALSE, FALSE, TRUE, TRUE, FALSE, FALSE, ~
$ passed_mvc3      <lgl> FALSE, FALSE, TRUE, TRUE, FALSE, TRUE, FALSE, TRUE, ~
$ pid_strength_rs  <dbl> 0.50270200, 0.50270200, 0.50270200, -0.97310186, 0.0~
```

```
# reproduce first model (overall approval of congress)
fit <- lm(cong_overall ~ failure*amplify*pid_strength_rs, data = coversight)

# summarize fit
arm::display(fit, detail = TRUE)
```

```
lm(formula = cong_overall ~ failure * amplify * pid_strength_rs,
    data = coversight)
```

	coef.est	coef.se	t value	Pr(> t )
(Intercept)	-0.27	0.09	-3.04	0.00
failureFailure	-0.94	0.13	-7.44	0.00
amplifyAmplify	0.58	0.13	4.56	0.00
pid_strength_rs	-0.06	0.19	-0.34	0.73
failureFailure:amplifyAmplify	1.46	0.18	8.18	0.00
failureFailure:pid_strength_rs	-0.15	0.26	-0.59	0.55
amplifyAmplify:pid_strength_rs	0.74	0.26	2.89	0.00
failureFailure:amplifyAmplify:pid_strength_rs	0.47	0.36	1.31	0.19

```
---
n = 1462, k = 8
residual sd = 1.70, R-Squared = 0.19
```

## 8.2 Raw Qualtrics Data

Readers can download the raw Qualtrics data using [this link](#). The preregistered code to clean these data can be downloaded using [this link](#).

## 9 A Reproduction of Our Main Results

We preregistered the code below (see Section 25 of the preregistration). The results are shown in Figure 1 in the paper.

```
# read cleaned data from dropbox
data <- read_rds("https://www.dropbox.com/s/9grn8kbb5yzwagx/data.rds?raw=1")%>%
  filter(passed_mvc1) %>% # remove respondents who failed MVC 1
  mutate(pid_strength_rs = arm::rescale(pid_strength)) # center partisan strength

# outcome: cong_overall
# -----
fit <- lm(cong_overall ~ failure*amplify*pid_strength_rs, data = data)
v <- sandwich::vcovHC(fit)

## H1
est <- coef(fit)["amplifyAmplify"] + coef(fit)["failureFailure:amplifyAmplify"]
c1 <- v["amplifyAmplify", "amplifyAmplify"]
c2 <- v["failureFailure:amplifyAmplify", "failureFailure:amplifyAmplify"]
c3 <- v["amplifyAmplify", "failureFailure:amplifyAmplify"]
se <- sqrt(c1 + c2 + 2*c3)
ci <- est + c(-1.64, 1.64)*se
cat(paste0("Effect = ", round(est, 2), "; 90% CI = [", round(ci[1], 2), ", ", round(ci[2],
```

Effect = 2.04; 90% CI = [1.84, 2.24]

```
## H2
est <- coef(fit)["amplifyAmplify"]
se <- sqrt(v["amplifyAmplify", "amplifyAmplify"])
ci <- est + c(-1.64, 1.64)*se
cat(paste0("Effect = ", round(est, 2), "; 90% CI = [", round(ci[1], 2), ", ", round(ci[2],
```

Effect = 0.58; 90% CI = [0.36, 0.79]

```
## H3
est <- coef(fit)["failureFailure:amplifyAmplify"]
se <- sqrt(v["failureFailure:amplifyAmplify", "failureFailure:amplifyAmplify"])
ci <- est + c(-1.64, 1.64)*se
cat(paste0("Effect = ", round(est, 2), "; 90% CI = [", round(ci[1], 2), ", ", round(ci[2],
```

Effect = 1.46; 90% CI = [1.17, 1.76]

```
# outcome: cong_investigate
# -----
fit <- lm(cong_investigate ~ failure*amplify*pid_strength_rs, data = data)
v <- sandwich::vcovHC(fit)

## H4(A)
est <- coef(fit)["failureFailure"] + coef(fit)["failureFailure:amplifyAmplify"]
c1 <- v["failureFailure", "failureFailure"]
c2 <- v["failureFailure:amplifyAmplify", "failureFailure:amplifyAmplify"]
c3 <- v["amplifyAmplify", "failureFailure:amplifyAmplify"]
se <- sqrt(c1 + c2 + 2*c3)
ci <- est + c(-1.64, 1.64)*se
cat(paste0("Effect = ", round(est, 2), "; 90% CI = [", round(ci[1], 2), ", ", round(ci[2],
```

Effect = 0.44; 90% CI = [0.28, 0.59]

```
## H4(I)
est <- coef(fit)["failureFailure"]
se <- sqrt(v["failureFailure", "failureFailure"])
ci <- est + c(-1.64, 1.64)*se
cat(paste0("Effect = ", round(est, 2), "; 90% CI = [", round(ci[1], 2), ", ", round(ci[2],
```

Effect = 0.66; 90% CI = [0.47, 0.85]

```
# outcome: cong_hearing
# -----
fit <- lm(cong_hearing ~ failure*amplify*pid_strength_rs, data = data)
v <- sandwich::vcovHC(fit)

## H5(A)
est <- coef(fit)["failureFailure"] + coef(fit)["failureFailure:amplifyAmplify"]
c1 <- v["failureFailure", "failureFailure"]
c2 <- v["failureFailure:amplifyAmplify", "failureFailure:amplifyAmplify"]
c3 <- v["amplifyAmplify", "failureFailure:amplifyAmplify"]
se <- sqrt(c1 + c2 + 2*c3)
ci <- est + c(-1.64, 1.64)*se
cat(paste0("Effect = ", round(est, 2), "; 90% CI = [", round(ci[1], 2), ", ", round(ci[2],
```

Effect = 0.6; 90% CI = [0.41, 0.78]

```
## H5(I)
est <- coef(fit)["failureFailure"]
se <- sqrt(v["failureFailure", "failureFailure"])
ci <- est + c(-1.64, 1.64)*se
cat(paste0("Effect = ", round(est, 2), "; 90% CI = [", round(ci[1], 2), ", ", round(ci[2],
```

Effect = 0.78; 90% CI = [0.57, 1]

```
# outcome: cong_report
# -----
fit <- lm(cong_report ~ failure*amplify*pid_strength_rs, data = data)
v <- sandwich::vcovHC(fit)

## H6(A)
est <- coef(fit)["failureFailure"] + coef(fit)["failureFailure:amplifyAmplify"]
c1 <- v["failureFailure", "failureFailure"]
c2 <- v["failureFailure:amplifyAmplify", "failureFailure:amplifyAmplify"]
c3 <- v["amplifyAmplify", "failureFailure:amplifyAmplify"]
se <- sqrt(c1 + c2 +2*c3)
ci <- est + c(-1.64, 1.64)*se
cat(paste0("Effect = ", round(est, 2), "; 90% CI = [", round(ci[1], 2), ", ", round(ci[2],
```

Effect = 0.6; 90% CI = [0.42, 0.78]

```
## H6(I)
est <- coef(fit)["failureFailure"]
se <- sqrt(v["failureFailure", "failureFailure"])
ci <- est + c(-1.64, 1.64)*se
cat(paste0("Effect = ", round(est, 2), "; 90% CI = [", round(ci[1], 2), ", ", round(ci[2],
```

Effect = 0.71; 90% CI = [0.49, 0.93]

```
# outcome: cong_campaign
# -----
fit <- lm(cong_campaign ~ failure*amplify*pid_strength_rs, data = data)
v <- sandwich::vcovHC(fit)
```

```
## H7(A)
est <- coef(fit)["failureFailure"] + coef(fit)["failureFailure:amplifyAmplify"]
c1 <- v["failureFailure", "failureFailure"]
c2 <- v["failureFailure:amplifyAmplify", "failureFailure:amplifyAmplify"]
c3 <- v["amplifyAmplify", "failureFailure:amplifyAmplify"]
se <- sqrt(c1 + c2 + 2*c3)
ci <- est + c(-1.64, 1.64)*se
cat(paste0("Effect = ", round(est, 2), "; 90% CI = [", round(ci[1], 2), ", ", round(ci[2],
```

Effect = 0.64; 90% CI = [0.43, 0.84]

```
## H7(I)
est <- coef(fit)["failureFailure"]
se <- sqrt(v["failureFailure", "failureFailure"])
ci <- est + c(-1.64, 1.64)*se
cat(paste0("Effect = ", round(est, 2), "; 90% CI = [", round(ci[1], 2), ", ", round(ci[2],
```

Effect = 0.91; 90% CI = [0.7, 1.13]

```
# outcome: pres_overall
# -----
fit <- lm(pres_overall ~ failure*amplify*pid_strength_rs, data = data)
v <- sandwich::vcovHC(fit)

## H8(A)
est <- coef(fit)["failureFailure"] + coef(fit)["failureFailure:amplifyAmplify"]
c1 <- v["failureFailure", "failureFailure"]
c2 <- v["failureFailure:amplifyAmplify", "failureFailure:amplifyAmplify"]
c3 <- v["amplifyAmplify", "failureFailure:amplifyAmplify"]
se <- sqrt(c1 + c2 + 2*c3)
ci <- est + c(-1.64, 1.64)*se
cat(paste0("Effect = ", round(est, 2), "; 90% CI = [", round(ci[1], 2), ", ", round(ci[2],
```

Effect = -1.23; 90% CI = [-1.37, -1.08]

```
## H8(I)
est <- coef(fit)["failureFailure"]
```

```
se <- sqrt(v["failureFailure", "failureFailure"])
ci <- est + c(-1.64, 1.64)*se
cat(paste0("Effect = ", round(est, 2), "; 90% CI = [", round(ci[1], 2), ", ", round(ci[2],
```

Effect = -1.34; 90% CI = [-1.55, -1.13]



## References

- Abrajano, Marisa, Christopher S. Elmendorf, and Kevin M. Quinn. 2023. "Measuring Perceived Skin Color: Spillover Effects and Likert-Type Scales." *The Journal of Politics* 85 (1): 320–27. <https://doi.org/10.1086/720941>.
- Andrews, Talbot M., Andrew W. Delton, and Reuben Kline. 2023. "Who Do You Trust? Institutions That Constrain Leaders Help People Prevent Disaster." *The Journal of Politics* 85 (1): 64–75. <https://doi.org/10.1086/720650>.
- Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20 (3): 351–68. <https://doi.org/10.1093/pan/mpr057>.
- Carnegie, Allison, Joshua D. Kertzer, and Keren Yarhi-Milo. 2022. "Democratic Peace and Covert Military Force: An Experimental Test." *Journal of Conflict Resolution* 67 (2-3): 235–65. <https://doi.org/10.1177/00220027221116289>.
- Chow, Wilfred M., and Enze Han. 2023. "Descriptive Legitimacy and International Organizations: Evidence from United Nations High Commissioner for Refugees." *The Journal of Politics* 85 (2): 357–71. <https://doi.org/10.1086/723017>.
- Coppock, Alexander, Thomas J. Leeper, and Kevin J. Mullinix. 2018. "Generalizability of Heterogeneous Treatment Effect Estimates Across Samples." *Proceedings of the National Academy of Sciences* 115 (49): 12441–46. <https://doi.org/10.1073/pnas.1808083115>.
- Elder, Elizabeth Mitchell, and Matthew Hayes. 2023. "Signaling Race, Ethnicity, and Gender with Names: Challenges and Recommendations." *The Journal of Politics* 85 (2): 764–70. <https://doi.org/10.1086/723820>.
- Graham, Matthew H. 2022. "Measuring Misperceptions?" *American Political Science Review* 117 (1): 80–102. <https://doi.org/10.1017/s0003055422000387>.
- . 2023. "Detecting and Deterring Information Search in Online Surveys." *American Journal of Political Science*, April. <https://doi.org/10.1111/ajps.12786>.
- Graham, Matthew H., and Shikhar Singh. 2023. "An Outbreak of Selective Attribution: Partisanship and Blame in the COVID-19 Pandemic." *American Political Science Review*, March, 1–19. <https://doi.org/10.1017/s0003055423000047>.
- Hansen, Kristina Jessen. 2022. "Greed, Envy, and Admiration: The Distinct Nature of Public Opinion about Redistribution from the Rich." *American Political Science Review* 117 (1): 217–34. <https://doi.org/10.1017/s0003055422000582>.
- Joseph, Michael F., Michael Poznansky, and William Spaniel. 2022. "Shooting the Messenger: The Challenge of National Security Whistleblowing." *The Journal of Politics* 84 (2): 846–60. <https://doi.org/10.1086/715595>.
- Joseph, Michael F., and Michael Poznansky. 2017. "Media Technology, Covert Action, and the Politics of Exposure." *Journal of Peace Research* 55 (3): 320–35. <https://doi.org/10.1177/0022343317731508>.
- Kane, John V., Yamil R. Velez, and Jason Barabas. 2023. "Analyze the Attentive and Bypass Bias: Mock Vignette Checks in Survey Experiments." *Political Science Research and Methods* 11 (2): 293–310. <https://doi.org/10.1017/psrm.2023.3>.
- Kim, Eunji. 2022. "Entertaining Beliefs in Economic Mobility." *American Journal of Political*

- Science* 67 (1): 39–54. <https://doi.org/10.1111/ajps.12702>.
- Kobayashi, Yoshiharu, Menevis Cilizoglu, Tobias Heinrich, and William Christiansen. 2023. “No Entry in a Pandemic: Public Support for Border Closures.” *American Journal of Political Science*, April. <https://doi.org/10.1111/ajps.12790>.
- Kreft, Anne-Kathrin, and Mattias Agerberg. 2023. “Imperfect Victims? Civilian Men, Vulnerability, and Policy Preferences.” *American Political Science Review*, April, 1–17. <https://doi.org/10.1017/s0003055423000345>.
- Kriner, Douglas L., and Eric Schickler. 2014. “Investigating the President: Committee Probes and Presidential Approval, 1953–2006.” *The Journal of Politics* 76 (2): 521–34. <https://doi.org/10.1017/s0022381613001448>.
- Krupnikov, Yanna, and Adam Seth Levine. 2014. “Cross-Sample Comparisons and External Validity.” *Journal of Experimental Political Science* 1 (1): 59–80. <https://doi.org/10.1017/xps.2014.7>.
- Levay, Kevin E., Jeremy Freese, and James N. Druckman. 2016. “The Demographic and Political Composition of Mechanical Turk Samples.” *SAGE Open* 6 (1): 215824401663643. <https://doi.org/10.1177/2158244016636433>.
- Lin, Winston. 2013. “Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman’s Critique.” *The Annals of Applied Statistics* 7 (1). <https://doi.org/10.1214/12-aos583>.
- Mullinix, Kevin J., Thomas J. Leeper, James N. Druckman, and Jeremy Freese. 2015. “The Generalizability of Survey Experiments.” *Journal of Experimental Political Science* 2 (2): 109–38. <https://doi.org/10.1017/xps.2015.19>.
- Myrick, Rachel. 2020. “Why So Secretive? Unpacking Public Attitudes Toward Secrecy and Success in US Foreign Policy.” *The Journal of Politics* 82 (3): 828–43. <https://doi.org/10.1086/707308>.
- Pérez, Efrén, Crystal Robertson, and Bianca Vicuña. 2022. “Prejudiced When Climbing Up or When Falling Down? Why Some People of Color Express Anti-Black Racism.” *American Political Science Review* 117 (1): 168–83. <https://doi.org/10.1017/s0003055422000545>.
- Potter, Philip B. K., and Matthew A. Baum. 2014. “Looking for Audience Costs in All the Wrong Places: Electoral Institutions, Media Access, and Democratic Constraint.” *The Journal of Politics* 76 (1): 167–81. <https://doi.org/10.1017/s0022381613001230>.
- Powers, Kathleen E., and Dan Altman. 2022. “The Psychology of Coercion Failure: How Reactance Explains Resistance to Threats.” *American Journal of Political Science* 67 (1): 221–38. <https://doi.org/10.1111/ajps.12711>.
- Renshon, Jonathan, Keren Yarhi-Milo, and Joshua D. Kertzer. 2023. “Democratic Reputations in Crises and War.” *The Journal of Politics* 85 (1): 1–18. <https://doi.org/10.1086/722043>.
- Singh, Shane P., and Jaroslav Tir. 2021. “Threat-Inducing Violent Events Exacerbate Social Desirability Bias in Survey Responses.” *American Journal of Political Science* 67 (1): 154–69. <https://doi.org/10.1111/ajps.12615>.
- Thomas, Kyle A., and Scott Clifford. 2017. “Validity and Mechanical Turk: An Assessment of Exclusion Methods and Interactive Experiments.” *Computers in Human Behavior* 77 (December): 184–97. <https://doi.org/10.1016/j.chb.2017.08.038>.

- Velez, Yamil R., Ethan Porter, and Thomas J. Wood. 2023. "Latino-Targeted Misinformation and the Power of Factual Corrections." *The Journal of Politics* 85 (2): 789–94. <https://doi.org/10.1086/722345>.
- Virgin, Sheahan G. 2021. "The Effect of Core Values on Support for Electoral Reform: Evidence from Two Survey Experiments." *American Journal of Political Science* 67 (1): 239–56. <https://doi.org/10.1111/ajps.12658>.
- Weinberg, Jill, Jeremy Freese, and David McElhattan. 2014. "Comparing Data Characteristics and Results of an Online Factorial Survey Between a Population-Based and a Crowdsourced-Recruited Sample." *Sociological Science* 1: 292–310. <https://doi.org/10.15195/v1.a19>.
- Zárate, Marques G., Enrique Quezada-Llanes, and Angel D. Armenta. 2023. "Se Habla Español: Spanish-Language Appeals and Candidate Evaluations in the United States." *American Political Science Review*, March, 1–17. <https://doi.org/10.1017/s0003055423000084>.