**Can Elites Escape Blame by Explaining Themselves?**
**Suspicion and the Limits of Elite Explanations**

Joshua Robison
Assistant Professor
Institute of Political Science
Leiden University
j.a.robison@fsw.leidenuniv.nl

**Abstract:**

Holding elected officials accountable for their behavior in office is a foundational task facing citizens. Elected officials attempt to influence this accountability process by explaining their behavior with an eye toward mitigating the blame they might receive for taking controversial actions. We address a critical limitation in the literature on elite explanation giving and accountability: the absence of attention to conflicting information regarding the official's behavior. We show across three pre-registered survey experiments that the effectiveness of explanations are negated when other speakers offer counter-explanations that focus attention on the potential ulterior motives of the official. We further show that this occurs even when the counter-explanation comes from a partisan source with low credibility. Our studies imply that elected officials experience less leeway for their actions than existing work allows and highlights important tensions concerning the interrelationship of elite behavior and accountability processes.

Elected officials generally wish to remain in office (Mayhew 1974). One way they can do so is by taking popular policy actions (e.g., Ansolabehere and Jones 2010). However, elected officials must sometimes take unpopular actions whether out of personal conviction or for some other reason. In such cases politicians try to limit the degree of blame accorded them by constituents (Weaver 1986). A crucial tactic on this front are *explanations*, wherein officials describe why they acted (Fenno 1978; Kingdon 1989; McGraw 2001). Existing work shows that explanations can often be an effective tool for avoiding blame (Broockman and Butler 2017; Grose, Malhotra, and Van Houweling 2015; Levendusky and Horowitz 2012; McGraw 2001; Robison 2017). While this may be good news for politicians wishing to maximize their re-election chances while acting as they see fit, it suggests a tension with accounts of democratic legitimacy grounded in the responsiveness of officials to public preferences (Dahl 1971).

Existing work on explanation giving has compared the important counterfactual of receiving an explanation versus receiving no explanation (e.g., Grose, Malhotra, and Van Houweling 2015; McGraw, Best, and Timpone 1995). However, explanation giving frequently occurs in contexts where "an 'instigator' (e.g. a challenger, the media, a disgruntled organized group) is eager to make public a representative's legislative activities" (McGraw 1991, 1154). These 'instigators' do more than publicize the behavior. Rather, they also introduce *counter-explanations* concerning why the official acted, ones that often highlight ulterior and self-serving goals potentially motivating the politician. As an example, during a 2016 US Senate race in North Carolina Senator Richard Burr explained his support for a change to Medicare by highlighting the potential economic benefits for affected seniors. Burr's opponent, meanwhile, charged that "Burr's

proposal…would change Medicare in a way that would benefit insurance companies who have donated to Burr's campaigns" (Douglas 2016). Voters in this race were thus able to consider two contrasting explanations for Burr's proposal: that he took the position to benefit his constituents or that he acted to reward private interests as a form of quid pro quo. Existing work leaves unclear how Burr's constituents will respond in this type of competitive environment. This is a crucial gap as it limits our understanding of the impact of explanations on the public as well as the incentives faced by politicians to take unpopular, but potentially necessary, policy actions.

We integrate theories on framing and the psychology of blame and suspicion to argue that elite explanations will be less effective in communication environments where rival accounts focus attention on the potential ulterior motives of the politician. These rival accounts provide a reason to be suspicious about the politician's true motives, which is important because suspicion prompts individuals to halt impression formation processes to consider the relative explanatory power of the rival interpretations (Fein, Hilton, and Miller 1990; Mayo 2015). Importantly, many people hold negative stereotypes about politicians wherein they are considered quite willing to sacrifice the public good for their own benefit (Clarke et al. 2018; Hibbing and Theiss-Morse 2002). Thus, when people compare the politician's explanation and an account suggesting the presence of ulterior motivations, we expect that many will find the latter more persuasive and thereby reject the focal politician's explanation.

We find support for the foregoing argument across three pre-registered survey experiments. While the politicians in our experiments were often if not always evaluated more positively when they explained their controversial actions, this effect disappeared

when a counter-explanation was present. We consider two scope conditions for this effect. First, we show that people discriminate between counter-explanations offered by more and less credible sources with the counter-explanation more influential in the former situation. Intriguingly, though, the negation of the focal politician's explanation *also* occurs when the counter-explanation is offered by a *low* credibility source. Second, we consider the role of partisanship and particularly the behavior of co-partisans to the focal politician. Here, we find that partisans were often willing to reject a co-partisan's explanation even when a low credibility source rebutted it. Thus, the effect of the explanation was attenuated even in two 'least likely' instances.

Our study contributes to the broader attempt to understand public reactions to competing messages from elites by focusing on a type of message salient in political discourse but largely ignored in prior work: messages that focus less on policy consequences and more on the procedural antecedents of elite positioning. In so doing it also adds important knowledge concerning the ability of elites to manage blame by highlighting a previously unexamined constraint on their ability to mold the accountability process.

### Explanations, Justifications, and the Psychology of Blame

Explanations figure into both the pre- and post- decision-making behavior of politicians (McGraw 2001). Our focus is on the role played by explanations after a politician takes an action such as casting a roll call vote. The goal of explanation giving in these contexts is to mollify constituents. A variety of explanations may be offered in these circumstances (Grose, Malhotra, and Van Houweling 2015; McGraw 1990). Our focus is on *justifications*, which tend to be more effective than alternative types of

explanation such as denials or excuses (McGraw 1990, 1991; but see, Peterson and Simonovits 2017). Justifications entail an acceptance of responsibility for the action coinciding with an attempt to recast it in a more positive light by highlighting the benefits that may follow from the action. Existing work shows that justifications are powerful blame management tools provided that the justification is considered 'satisfactory', i.e. is accepted by audience members (Broockman and Butler 2017; Grose, Malhotra, and Van Houweling 2015; McGraw 1991; Robison 2017).

The psychology of blame attribution highlights why a justification might placate constituents. Perceivers blame an actor, and thereby evaluate them negatively, when they believe the actor has *intentionally* taken a *wrongful* action (Malle, Guglielmo, and Monroe 2014). Justifications may influence either consideration thereby mitigating blame. On the one hand, a justification may persuade constituents that the action was correct through the provision of information concerning the action's consequences (Broockman and Butler 2017; McGraw 1991). Here the justification removes the initial basis for blame. On the other hand, the extent of blame attributed to an actor also hinges on a consideration of the agent's reasons for acting (Malle 2011). Perceivers blame agents less when they believe that 'positive' intentions rather than 'negative' ones motivated the action. In the political domain this means attributing to officials the intention of trying to help constituents and make good policy rather than trying to help oneself or 'special interests' (Bøggild 2016; Doherty 2015). As one politician told Bianco (1994): "You need to defend your position. If you give a rational and reasonable answer, they'll say, 'I disagree with your vote, but I understand why you did it and I don't hold it against you.'"

**H1**: Providing a justification will lead to more positive evaluations compared to when no justification is offered, all else equal

### Instigators, Suspicion, and the Limits of Elite Explanations

Many studies suggest that politicians can strategically use explanations to ameliorate the evaluative costs of their controversial actions (but see Peterson and Simonovits 2017 for a notable exception). However, this work shares a common limitation in that it does not incorporate the political contestation that occurs surrounding elite behavior. For instance, Broockman and Butler (2017) and Grose, Malhotra, and Van Houweling (2015) both examine one-sided communication environments such as when a legislator responds to a constituent communique. McGraw, meanwhile, notes that 'instigators' help publicize elite actions, but these instigators do not provide competing information in her studies (McGraw 1991; McGraw, Timpone, and Bruck 1993; McGraw, Best, and Timpone 1995). However, situating explanation giving in a more competitive environment is necessary to fully understand the efficacy of explanation giving and the resulting incentives faced by politicians.

Instigators do not just publicize a politician's actions. Rather, they also provide conflicting interpretations of this behavior. This conflict can occur along two dimensions as implied by the foregoing discussion on blame. First, instigators may provide policy arguments to dispute the focal politician's claims that their actions will generate positive consequences. Existing work shows that message competition can, in some instances, mitigate the influence of an otherwise persuasive message (Boudreau and MacKenzie 2014; Chong and Druckman 2007a). However, this work focuses on evaluations of *policy*, not evaluations of politicians. However, it is unclear whether conflict over the merits of an action would mitigate the blame avoidant properties of justifications.

Justifications may also influence beliefs about the motives of politicians, which is important beliefs about the motives of agents appear to be central to impression formation processes (McGraw 2003).

Our focus lies on a second dimension of interpretative conflict, one which has not been investigated in the broader literature on competition and opinion formation. Justifications provide an account for why the politician acted, i.e. what (typically policy-oriented) goals they were trying to realize. Instigators, however, may have incentives to change the focus of the debate to an alternative evaluative dimension (Jerit 2009). Instigators, for instance, may provide counter-explanations that highlight ulterior and self-serving reasons motivating the politician and hence potential defects in their character. As Adams, Ezrow, and Somer-Topcu (2011, 372) note, for instance, "parties' new policy pronouncements often provoke rival party elites to publicly deride the focal party's new initiatives as being 'opportunistic,' 'pandering', [and] 'insincere'". Consider an example we return to below: US Senator Kirsten Gillibrand's policy evolution from being anti- to pro-gun control after she became a Senator. While the Senator explained this evolution as emanating from learning new information about the consequences of gun violence, her political opponents highlighted the Senator's desire to win support from party elites to advance her career (Nilsen 2019). Existing work leaves unclear how Gillibrand's constituents, or potential new voters, will respond when faced with these conflicting messages.

We expect that counter-explanations focused on the ulterior motives of the politician will lead to worse evaluations compared to a justification only counter-factual.

In making our argument we integrate two separate literatures.[1] First, individuals tend to shift to a more cognitively active form of information processing when they have reason to suspect that ulterior goals motivate the actions of an agent (Fein, Hilton, and Miller 1990; Mayo 2015; Priester and Petty 1995). In so doing suspicious perceivers consider the relative plausibility of rival interpretations for the agent's behavior to avoid being misled.

Suspicion entails considering the possibility that ulterior motives guided the politician's behavior. We further argue that many will find this type of counter-explanation persuasive and hence find the focal politician's explanation as unsatisfactory. First, news accounts often frame politicians as guided by strategic and self-interested motives (Aalberg, Strömbäck, and de Vreese 2011). Thus, many have an available stereotype of politicians as untrustworthy and corrupt (Clarke et al. 2018). Second, a variety of literatures show that information about the motives of other agents is important in impression formation processes. This type of information is more likely to be selected by participants in media choice experiments and better remembered over time (Bøggild, Aarøe, and Petersen, n.d.; Iyengar, Norpoth, and Hahn 2004). Likewise, beliefs about the motives of decision makers influences evaluations both toward the decision maker and the decision itself (Bøggild 2016; Hibbing and Alford 2004). And, information about the intentions of political candidates, individuals, and social groups is the preeminent influence on impression formation reflecting the importance of these considerations in making credit and blame attributions (Bittner 2011; Cuddy, Fiske, and Glick 2008;

---

[1] Our argument is also consistent with the literature on negative campaigning as negative advertisements tend to modestly undermine affect toward the targeted candidate (Lau, Sigelman, and Rovner 2007).

Laustsen and Bor 2017). Thus, counter-explanations focused on the ulterior ends of a

politician make accessible an available and *applicable* consideration for individuals to

ponder. This is important because considerations that are available, accessible, and

applicable tend to be highly consequential in attitude formation processes (Chong and

Druckman 2007b). Thus, we expect that many will find these counter-explanations to be

a better description for a politician's behavior than their own explanation, blunting the

latter's impact.

> **H2**: Counter-explanations will lead to worse evaluations compared to
> when only the justification is present, all else equal

### Credibility, Partisanship, and Scope Conditions

We consider two scope conditions for the foregoing argument. First, counter-

explanations may be issued by sources that vary in perceived credibility. Political rivals,

for instance, may be perceived as having a political motivation to impugn the focal

politician and thus deemed less credible, whereas non-partisan actors may be perceived

as more trustworthy.[2] We expect that counter-explanations will be more effective when

from more credible sources (Chong and Druckman 2007a; Weitz-Shapiro and Winters

---

[2] It is plausible that political rivals are more likely to offer such accounts. However, non-partisan actors are

also likely to voice these messages as well. For instance, a Washington Post article concerning lobbyist

donations to key Democrats overseeing the development of the Affordable Care Act included the following

quote: "But Jerry Flanagan, a health-care analyst with Consumer Watchdog, a California-based advocacy

group, said the tide of campaign contributions amounts to 'a huge down payment' by companies that expect

favorable policies in return. 'That is the cold reality of big-money politics'" (Eggen 2009).

2016). However, we do not make any specific prediction about whether counter-explanations are *only* effective when offered by a high credibility source.

Second, politicians explaining themselves are particularly concerned with maintaining support among their supporters and hence their co-partisans given the overlap between these two categories (Fenno 1978, 168). How should co-partisans react? While these audience members may possess a partisan motivation to discount the counter-explanation (Lodge and Taber 2006), this does not mean that they are immune to counter-argumentation as strong arguments can overcome partisan interpretations in some contexts (Boudreau and MacKenzie 2014; Bullock 2011). We argue that co-partisans should be particularly attuned to source credibility when evaluating counter-explanations. Partisans from the opposite party as the focal politician may be motivated to accept derogatory information regardless of the source of the message. Co-partisans, on the other hand, may be more resistant and only buckle when the incoming information is highly credible and thus more difficult to counter-argue.

> **H3:** Counter-explanations offered by credible sources will harm evaluations more than those offered by less credible sources, all else equal

> **H4:** Credibility will matter more for co-partisans to a politician than for opposing partisans, all else equal

### Study Design

**Table 1**: Overview of Experiments

| | **Experiment 1** | **Experiment 2** | **Experiment 3a** | **Experiment 3b** |
|---|---|---|---|---|
| Sample | MTurk (n=1816) | MTurk (n=1611) | Lucid Fulcrum (n=1214) | |
| Politician | "Representative A" | "Dennis Williams" | Kirsten Gillibrand | Bob Corker |
| Action | Vote to cut education funding | Vote to raise electricity costs | Flip-flop on gun control | Flip-flop on Trump Tax Cuts |
| Partisanship Manipulated? | Yes | Yes | No | No |
| Information Conditions | • No Vote Information<br>• Vote w/o Justification<br>• Vote w/ Justification<br>• Justification + Counter from Teachers (HC)<br>• Justification + Counter from Non-Partisan Expert (HC)<br>• Justification + Counter form Out-Party Opponent (LC) | • Vote w/o Justification<br>• Vote w/Justification<br>• Justification + Counter from Non-Partisan Expert (HC)<br>• Justification + Counter from In-Party Opponent (LC) | • Flip-Flop w/o Justification<br>• w/Justification<br>• Justification + Counter from Right-Wing Source<br>• Justification + Counter from Left-Wing Source | |
| Justification | Fairness | Long Term Interests of Constituents | New Information | Benefits to Constituents |
| Counter-Explanation | Campaign Donor Quid Pro Quo | Campaign Donor Quid Pro Quo | Career Opportunism | Financial Self-Interest |
| Total Number of Conditions | 12 | 8 | 4 | 4 |

**Notes**: LC = Low Credibility; HC = High Credibility.

We test our hypotheses across three pre-registered survey experiments. This enables us to replicate and build on initial results with new samples and stimuli thereby increasing the external validity of our argument (McDermott 2011). Across the

experiments we move from fictional politicians acting on non-salient issues to real politicians acting on salient issues where the (counter)-explanations on offer reflect actual political rhetoric. Table 1 provides an overview of some of the key elements of these studies. We use this section to describe all three experiments and the logic behind our design choices.

**Choice of Sample**

Experiments 1 and 2 were conducted using samples recruited from Amazon's Mechanical Turk. Experiment 3, meanwhile, used a sample recruited using Lucid Fulcrum Exchange, which is a marketplace of research firms and their survey panels. Neither data source uses a probability-based design, although recruitment using Lucid utilized quotas on education, race, and age. Nevertheless, both sources yield samples that better approximate the demographic profile of the US adult population than other convenience sampling methods; see Table OA1 for an overview of sample characteristics. Importantly, validation studies show that treatment effects generated by either Turk or Lucid resemble those generating using probability-based sampling designs (Coppock and McClellan 2019; Mullinix et al. 2015).

**Choice of Politician**

We progressively build in greater external realism via the choice of politician. In Experiment 1, subjects read about "Representative A", a politician whom they possess no prior information about (as in: Butler and Powell 2014). This is common in experimental studies of elite actors but does raise questions about whether the results apply to contexts where people possess prior attitudes toward the politician. We address this issue in Experiment 2 by first providing respondents with information about a politician and

measuring their initial impression; see Online Appendix C for these treatments. Subjects are then asked to re-evaluate this politician later in the survey after receiving the main treatment in the study. While the experimental inducement of a prior attitude moves us closer to approximating real-world contexts, Experiment 2 nevertheless focuses on a politician about whom respondents have only learned about once and whom they will never hear about again. Thus, in Experiment 3 we randomly assign respondents to read information about two real-world politicians, Senator Kirsten Gillibrand (D-NY) and former Senator Bob Corker (R-TN).

**Choice of Policy Action**

In each experiment the politician took a potentially controversial policy action. In Experiment 1 the politician, a state legislator, voted in favor of a budget amendment resulting in cuts to education spending for the legislator's district. In Experiment 2, the politician, again a state legislator, voted in favor of a change in regulations to a publicly regulated electricity company that would lead to higher costs for consumers. We chose these actions because they are ones that a legislator may need or choose to take for a variety of reasons, but ones likely unpopular because they impose 'costs' on constituents.[3] At the same time, the politicians are taking action on issues about which respondents are unlikely to have a prior attitude. Thus, in Experiment 3 we focus on real policy actions by Senators Gillibrand and Corker regarding salient political issues. The Gillibrand experiment focuses on the Senator's change from being a supporter of gun rights to a supporter of stronger gun control after her ascension to the US Senate. The Corker experiment, meanwhile, focuses on his change from opposing to supporting the

---

[3] The spending cut manipulation, moreover, was based on a similar scenario used in McGraw et al. (1993).

Republican tax-cuts passed in late 2018. Politicians may have a particularly difficult time explaining away disagreements with constituents on these types of salient political issues (Peterson and Simonovits 2017). Moreover, they involve 'flip-flopping,' a behavior that may generate negative impressions by itself (Doherty, Dowling, and Miller 2016).

**Experimental Procedure**

In each experiment we randomly assigned respondents to conditions varying in the presence of a justification and a counter-explanation.[4] In Experiment 1, respondents were assigned to one of six information conditions. Subjects in the first condition only read background information about the politician, while respondents in the remaining five conditions also learned that the politician voted in favor of the spending cut described earlier. In the No Justification condition, the politician did not provide a rationale for doing so. In the Justification condition, meanwhile, they ground this decision in a desire to "make the distribution of school funding across the state fairer for all citizens than under the present budget".[5] In the final three conditions the justification was present alongside a counter-explanation wherein the source insinuates that the representative voted for the amendment because his/her campaign donors would financially benefit from the bill. The potential source for the counter-explanation was either local teachers,

---

[4] In Experiments 1 and 2 we also randomly assigned the partisanship of the politician (Democratic vs Republican). In Experiment 3, we could not randomly assign the partisanship of the politicians but did randomly assign the order in which the respondent read about them. All experimental treatments can be found in the Online Appendices (B-D).

[5] We pre-tested this justification on a separate sample of respondents, alongside three others, and found it to be the strongest one available; see Online Appendix B.

non-partisan experts, or partisans from the opposite party as Representative A. Pre-test results reported in Online Appendix B show that local teachers and non-partisan experts are more credible than the more explicitly partisan source. However, while this manipulation should provide variation in credibility on average, it does have a drawback. Specifically, respondents that belong to a different party as Representative A and assigned to the "partisans from the opposite party" counter-explanation condition will receive a counter-explanation from *their co-partisans* due to the nature of the design. We designed Experiment 2 to help address this issue.

The procedure in Experiment 2 was different. Respondents began the survey by reporting their partisan identification and ideology. They were then told that they would be provided with a short description of an elected official provided by a non-partisan group after which they would be asked for their "overall opinion regarding the legislator". We adopted this language to stimulate respondents to form an online, i.e. "strong", attitude regarding the legislator (Chong and Druckman 2010). Respondents read information about either a Republican or Democratic version of Dennis Williams before answering a series of buffer items about their needs for cognition and affect.[6]

Respondents were assigned to one of four conditions after this buffer. In all four conditions respondents read a news article wherein they learned that Williams had cast the tie-breaking vote in favor of a budget amendment expected to lead to higher prices for customers of the publicly regulated Burlington Electric company. The four versions of this article varied according to the presence of the justification and counter-explanation.

---

[6] Online Appendix C provides the treatment wordings for the two politician profiles as well as details on the distribution of these initial impressions.

The politician does not offer an explanation in the first condition, while in the second he justifies the vote by arguing that the change was in the best long-term interests of the community because "if we do not make an investment now then prices will increase even more dramatically over time" and "sometimes you have to make tough choices that you believe are in the best long-term interests of the community".[7]

Respondents assigned to the final two conditions read this justification alongside a counter-explanation sourced to "Gary Allison". In the High Credibility condition this speaker was described as an economist at the non-partisan Center on Budget and Policy Priorities. In the Low Credibility version, meanwhile, he was described as a city councilor from the same party as the representative who was currently "trailing Williams in the polls in their upcoming primary election". In the latter case the two speakers are thus from the same party. Respondents from a different party as the representative should thus also be somewhat distrusting of this other speaker given this difference in partisanship. On the other hand, co-partisans to Representative Williams may also feel that this speaker is untrustworthy given that he has electoral incentives to impugn the motives of the representative.[8] Regardless of background, the counter-explanation always had the speaker suggesting that "lobbyists for Burlington Electric Department had

---

[7] See Online Appendix C for pre-test results concerning this account, which was found to be effective in mitigating evaluative damage from taking the action.

[8] Co-partisans to Representative Williams broadly entered the second experiment with favorable impressions of the legislator, which may bolster their propensity to be distrustful of impression incongruent information (Lodge and Taber 2006). Online Appendix C provides pre-test analyses pertaining to source credibility that also suggest that this speaker should have reduced credibility.

extensively lobbied committee members to vote for the rate increase," and that "the rate means a big cash inflow for Burlington electric, and I'm sure they were holding out future campaign donations as a carrot to vote for the hike."

The procedure in Experiment 3 was similar to Experiment 2. Respondents began the survey by answering questions about their political attitudes, including their partisan identity, attitudes regarding the two issues used in the experiment, and evaluations of Senators Gillibrand and Corker. A battery of items tapping need for cognition was used as a short buffer before respondents completed the first politician experiment, after which respondents answered a short battery of political knowledge questions and completed the second experiment.

Respondents were assigned to four experimental conditions: a No Justification condition, a Justification condition, and two conditions where a counter-explanation was also provided.[9] The justifications used in the experiment were the actual rationales publicly offered by the two politicians for their behavior. Senator Gillibrand's justification focused on how becoming a Senator had presented her with opportunities to receive new information concerning the consequences of gun violence after "meeting with the families of gun violence victims". Senator Corker, meanwhile, rooted his decision in conversations with constituents and business leaders and the possibility that the tax cuts could help "drive additional foreign direct investment in Tennessee".

The counter-explanations respondents were eligible to receive also come from real discourse. The counter-explanation in the Gillibrand experiment was taken from a Republican National Committee press release suggesting that the Senator's change of

---

[9] Respondents were freely assigned across the two experiments.

heart was motivated by "political opportunism". The counter-explanation in the Corker experiment, meanwhile, focused on insinuations that Corker had changed his vote due to the inclusion of a provision in the bill that benefited him financially (e.g., Kim 2017). Respondents could receive these messages from either a left-wing or a right-wing source.[10] Variation in credibility is thus dependent on the audience member's political affiliations, e.g. a Republican should judge a right-wing source as more credible than a left-wing source and vice versa for Democrats.

**Dependent Variable and Models**

Respondents in each experiment were asked how they "would rate [the politician]" on a 0-10 scale where "0 means that you think very poorly of [the politician] and 10 means that you think very highly of [the politician]". Analyses reported below use a version of this item rescaled to range from 0-1. We regress this variable on indicators of the experimental treatment conditions to which the respondent was assigned. We use respondents assigned to the Justification condition as the baseline in these analyses given this condition's central role in all hypotheses.[11]

The analyses of Experiments 2 and 3 also incorporate pre-test covariates into these models. Including pre-test covariates that are significantly related to the dependent variable can increase the power of experimental analyses (Gerber and Green 2012 Ch. 4;

---

[10] In the Gillibrand experiment, the source was either a "conservative" or a "liberal columnist". In the Corker experiment, the counter-explanation was attributed to either "liberal" or "conservative critics" with a quote from "Brett Stevens, a Senior Fellow at the [liberal/conservative] think tank the Tax Policy Center".

[11] We thank an anonymous Reviewer for suggesting this way of presenting the analyses.

Clifford, Sheagley, and Piston 2020). In Experiment 2 we control for the respondent's pre-test evaluation of the politician, which is substantially correlated with the post-test measure (r=0.74). In Experiment 3, we include the respondent's pre-test evaluation of the politician, their co-partisanship status with the politician, issue agreement between the politician and respondent, and an indicator for the order in which they read the politician treatments. The former three items are all significantly related to post-test evaluations as we would expect them to be given existing work on the role of prior attitudes, partisanship, and policy agreement in shaping evaluations (e.g., Ansolabehere and Jones 2010; Lodge and Taber 2006).  We provide analyses omitting these covariates in Online Appendix A. In Online Appendix B we provide results for Experiment 2 focused on a difference score as well as a three-category indicator for whether evaluations were worse, did not change, or became better on the post-test. Our substantive conclusions remain the same as those in-text in both cases.[12]

## Results

### Do explanations work less well when counter-explanations are present?

We begin with Figure 1 and an examination of Hypotheses 1 and 2. Figure 1 provides OLS coefficients from four regression models. For these initial analyses we

---

[12] Online Appendices B-D also contain analyses of additional post-test measures, specifically indicators of the respondent's post-test agreement with the policy in question (Experiments 1 & 2) and motive attributions (all experiments). We provide pre-analysis plans at the end of the Online Appendices. We deviate from this plan in one aspect: while we pre-registered analyses wherein we combined the different counter-explanation treatments together for Experiments 2 and 3 (e.g. Figure 1), we did not do so for Experiment 1.

**Figure 1:** Explanations Only Work When They are Unopposed
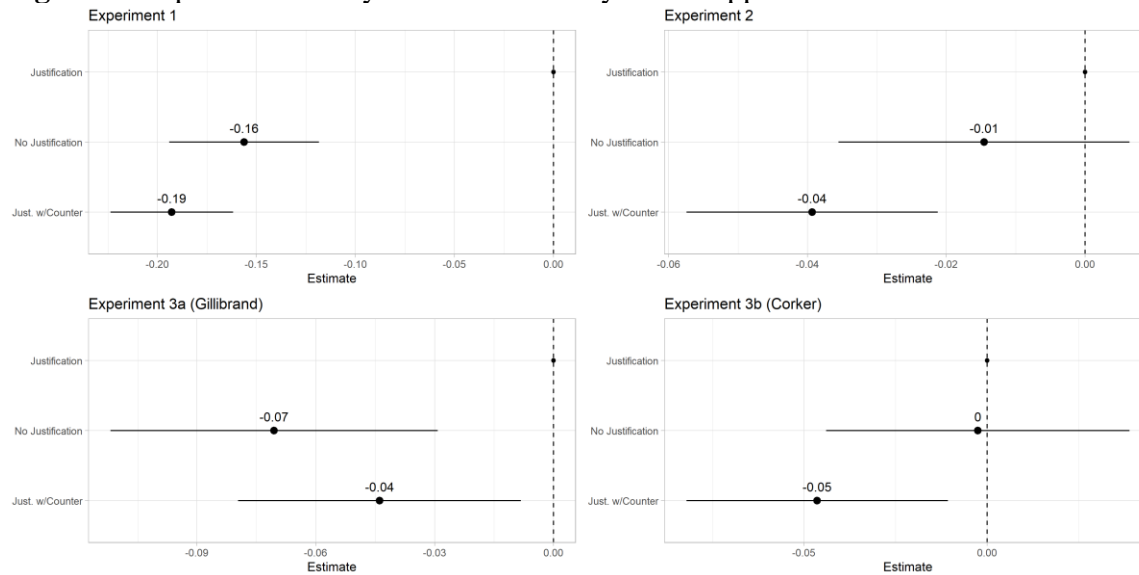


**Notes:** Markers provide the average difference in post-test evaluations (with 95% confidence intervals) relative to respondents assigned to the Justification condition (represented by a marker at 0). See Tables OA1, OA2 & OA4 for model results.

merge all respondents receiving a counter-explanation regardless of its source. The baseline for these analyses is the condition wherein the respondent received a Justification by itself. Thus, *negative* coefficients are evidence in favor of both H1 and H2.

Figure 1 communicates two key points. First, providing a justification is no panacea. In Experiments 1 and 3a (Gillibrand) respondents report significantly more positive evaluations of the politician on the post-test when they received a justification than when they did not. However, in neither of the other two cases does offering a justification improve evaluations. As we discuss later, one reason why the Corker experiment does not show evidence of a significant treatment effect is because of the offsetting reactions of partisans. The lack of a clearer effect in Experiment 2, meanwhile, may have something to do with the specific action of the politician, although our

reasoning here is speculative. In Experiment 2 the politician casts a tie-breaking vote for a policy that will impose financial costs on constituents. This was not a behavior looked kindly on by those in the no explanation baseline, with approximately 42% of respondents in this condition reporting more negative evaluations on the post-test than on the pre-test. It may be the case that this action was a particularly tough one to evade blame for as the legislator was described as making the tie-breaking vote in its favor and thus was clearly the actor culpable for imposing financial costs on constituents.

Second, respondents assigned to a counter-explanation condition reported significantly worse evaluations than those in the justification condition in all cases as expected in Hypothesis 2. Moreover, the impact of the counter-explanation did not just take the form of a balancing out of the influence of the justification. Instead, evaluations of the politician were significantly worse in the counter-explanation condition relative to the *no justification* condition in three of the four cases.[13] In other words, the counter-explanation overwhelmed the influence of the politician's justification. Ultimately, Figure 1 provides some supportive evidence in favor of H1 and much more solid evidence in favor of H2.

**Does the source of the counter-explanation matter?**

---

[13] We tested the equivalence of the two coefficients (No Justification and Justification w/Counter) with Wald tests to ascertain this. In three cases we can reject the null that the two coefficients are equivalent: Experiment 1: $F = 5.40$, $p < 0.05$; Experiment 2: $F = 7.13$, $p < 0.01$; Experiment 3a: $F=2.13$, $p =0.15$; Experiment 3b: $F = 5.64$, $p < 0.05$. Using the No Justification condition as the baseline in the regression leads to the same conclusions.

Figure 1 showed that politicians hoping to avoid negative evaluations by justifying themselves had better hope that another actor does not insinuate that they possess ulterior motives. In Hypothesis 3 we argued that counter-explanations should be more effective when voiced by more credible sources. We test this claim in analyses reported in Figure 2, which provides plots for each experiment. The top facet in each experiment plot provides the average difference in evaluations between those in the assigned condition and those in the Justification counterfactual with the counter-explanation conditions now separated by source. The bottom facet, meanwhile, provides the difference in coefficients between the various counter-explanation conditions. Evidence consistent with Hypothesis 3 would take the form of a larger negative coefficient for the high credibility source in the top facet and a negative difference in the bottom half. Our main attention will be on Experiments 1 and 2 as they provide the clearest differentiation in source credibility at this level of respondent aggregation.

We see two key results in Figure 2. First, credibility increased the influence of the counter-explanation in the first two experiments. The coefficient for the more credible source is always numerically greater with the difference between the high and low credibility effects in the bottom facet negative and statistically significant as well. Interestingly, Figure 2 shows that it was ultimately the higher credibility sources driving the additional negative evaluative consequences of the counter-explanation seen in Figure 1 and discussed earlier as the coefficient for the low credibility source is in line with, and not statistically distinct form, the coefficient for those assigned to the No Justification condition.

**Figure 2:** Both Partisan and Non-Partisan Sources Harm Evaluations
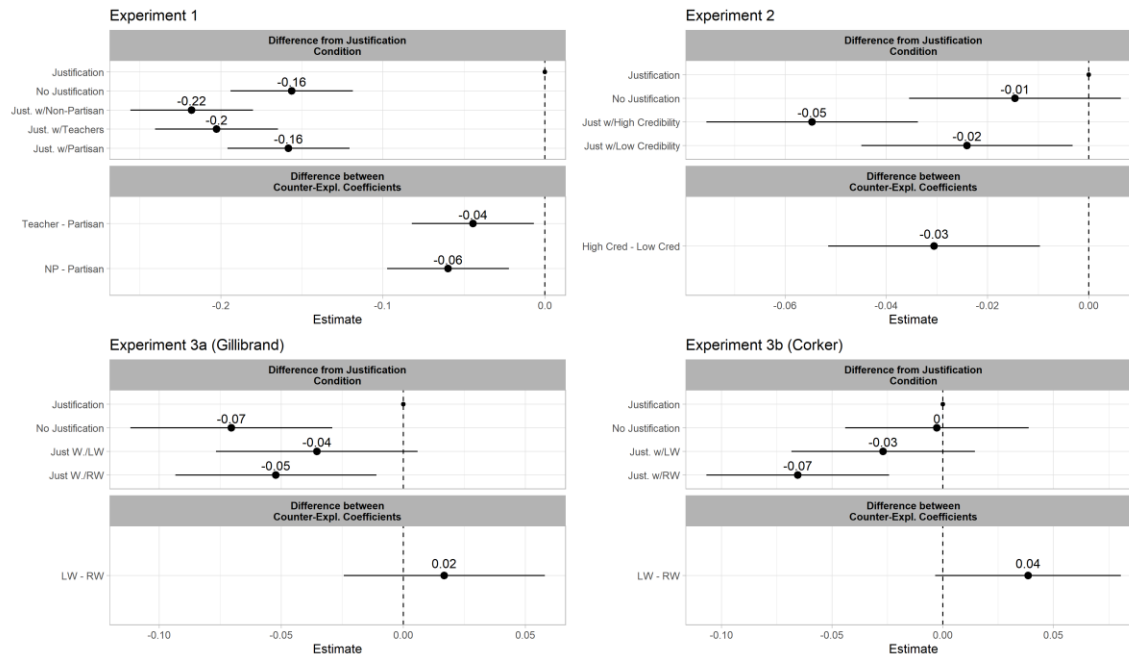


**Notes:** The top facet within each plot provides the average difference in evaluations relative to respondents in the Justification condition. The bottom facet provides the difference between the counter-explanation condition coefficients (High Credibility – Low Credibility for Experiments 1 & 2; Left Wing – Right Wing for Experiments 3a & 3b). Negative coefficients indicate that the former source had a larger negative impact on evaluations than the latter source. See Tables OA1, OA2, and OA5 for model results.

Second, Figure 2 shows that even a source with low credibility could undermine the effectiveness of the politician's explanation in these first two experiments even if not to the same extent as a more credible source. Evaluations were significantly worse in the low-credibility counter-explanation condition relative to the justification condition in both experiments. At first glance this result resembles findings from the competitive framing literature wherein a weak argument does not influence opinions and two equally strong arguments tend to balance out in their influence on resulting policy opinions (Chong and Druckman 2007a; Druckman, Fein, and Leeper 2012). However, this result is novel as the low credibility source manipulation should have undermined the perceived

strength of the counter-explanation, as it does in Chong and Druckman (2007), thereby enabling the justification to still influence evaluations (particularly in Experiment 1). In other words, that the counter-explanation from the low credibility source continued to fully balance out the effect of the otherwise substantially effective explanation signifies something about the inherent strength of the underlying message.

The distinction between more and less credible sources is not as clear cut in Experiment 3. As a result, we do not expect to see much aggregate level differences in these experiments insofar as those on the left-and right may react in off-setting ways. Nonetheless, Figure 3 does suggest some slight evidence of differential effectiveness based on source characteristics. In both cases the right-wing source appears to have a greater impact than the left-wing source, although in neither case is the effect statistically significant (see bottom facet). We return to this point in the following section. For now, though, the clear message for politicians is to hope that media coverage of their behavior focuses attention on their partisan rivals rather than critical comments from non-partisan actors as the former situation enables them to at least tread water.

**Do co-partisans pay more careful attention to source credibility?**

We now turn our attention to the role of partisanship. We are especially interested in the reactions of co-partisans to the politician offering a justification. We argued in Hypothesis 4 that these co-partisans should be especially attentive to the credibility of counter-explanation providers given partisan motivations that may make them both an 'easy' audience for the focal politician and a 'hard' one for information that conflicts with their partisan loyalties. There are two key tests here. First, is the effect of a high credibility counter-explanation significantly different from that of a low credibility message among co-partisans? Second, are the differences observed within partisan group

also significant from one another, i.e. is the difference between high and low credibility source effects greater among co-partisans than among opposing partisans? Full support for Hypothesis 4 would manifest in both patterns being the case.

We begin with Figure 3 which focuses on Experiments 1 & 2. The results in Figure 3 are based on a model where evaluations are regressed on information condition, co-partisanship status, and their interaction (along with pre-test evaluations for Experiment 2); see Tables OA1, OA2, and OA6. The top facet provides the average marginal effect of being assigned to each condition relative to the Justification counterfactual separately for co-partisans (circles) and opposing partisans (triangles). The middle panel compares the effect of credibility within each partisan group (e.g. Teachers – Partisan for Co-Partisans), while the bottom panel then provides the difference in difference estimate (e.g. [Co-Partisan: Teachers – Partisan] – [Opposing Partisan: Teachers – Partisan]) Significant negative estimates in these two latter panels would support Hypothesis 4.

A first thing to note in Figure 3 is the effect of receiving a counter-explanation among co- and opposing partisans as shown in the top facet. Evaluations in both experiments were significantly worse among both type of partisan when the counter-explanation stemmed from a high credibility source. Co-partisans in both experiments, meanwhile, also reported significantly worse evaluations when a *low credibility* source offered it, which is specifically notable in Experiment 1 where the low credibility source for co-partisans are out-partisans. Co-partisans were thus not immune to argumentation impugning a politician from their side.

**Figure 3:** Co-Partisans Pay More Attention to Credibility in Experiments 1 & 2
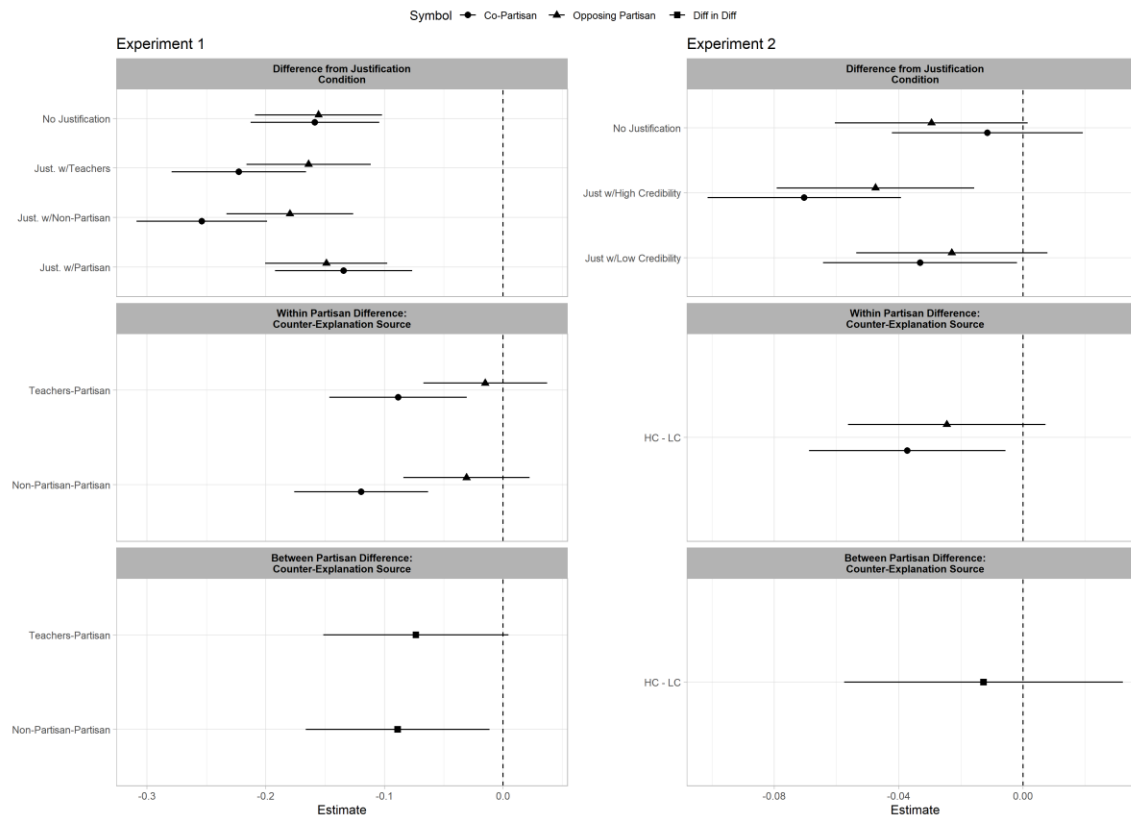


**Notes:** The top facet provides the average difference in evaluations based on treatment condition with separate markers for co-partisans (circle) and opposing partisans (triangles). The middle facet provides the difference *within* partisan group between the counter-explanation coefficients (e.g.: Co-Partisan: HC-LC). The final facet provides a difference in difference (e.g., Co-Partisan[HC-LC] – Opposing Partisan[HC-LC]).

The middle and bottom facets of Figure 3 consider whether co-partisans were more sensitive to source credibility with mixed evidence resulting. The middle panels show that high credibility sources had a significantly greater effect on evaluations for co-partisans in both experiments. Opposing partisans, meanwhile, reacted broadly the same regardless of the source, although we should be cautious in interpreting the Experiment 1 results given that opposing partisans in the Partisan counter-explanation condition read a message from someone on their partisan team. The bottom panels, however, show that the difference in difference estimate is only statistically significant in Experiment 1 when

the high credibility source was a non-partisan.[14] We thus fail to observe consistent evidence in support of Hypothesis 4.[15]

Figure 4 provides the same styled graph but for Experiments 3a and 3b. A first thing to note is the role of partisanship in shaping reactions to the politician's justification. While both types of partisan reacted positively to the Gillibrand justification, we observe more 'traditional' partisanship effects in the Corker experiment as partisans polarized in their reaction to the justification. One plausible reason why this occurred may be the explicit connection between this issue and Donald Trump, an almost inherently polarizing agent in contemporary American politics. Ultimately, this offsetting pattern helps explain the lack of an overall effect for this justification as noted earlier.

---

[14] Experiment 1, Local teachers: difference = -0.07 [-0.15, 0.01], p < 0.10; Experiment 2: -0.01 [-0.06, 0.03], p = 0.58).

[15] One possible reason for this may be a lack of power. We consider the role of power in Online Appendix A where we conclude this consideration is unlikely to be driving the lack of effect here.

**Figure 4:** Co-Partisanship Matters Less in Experiment 3
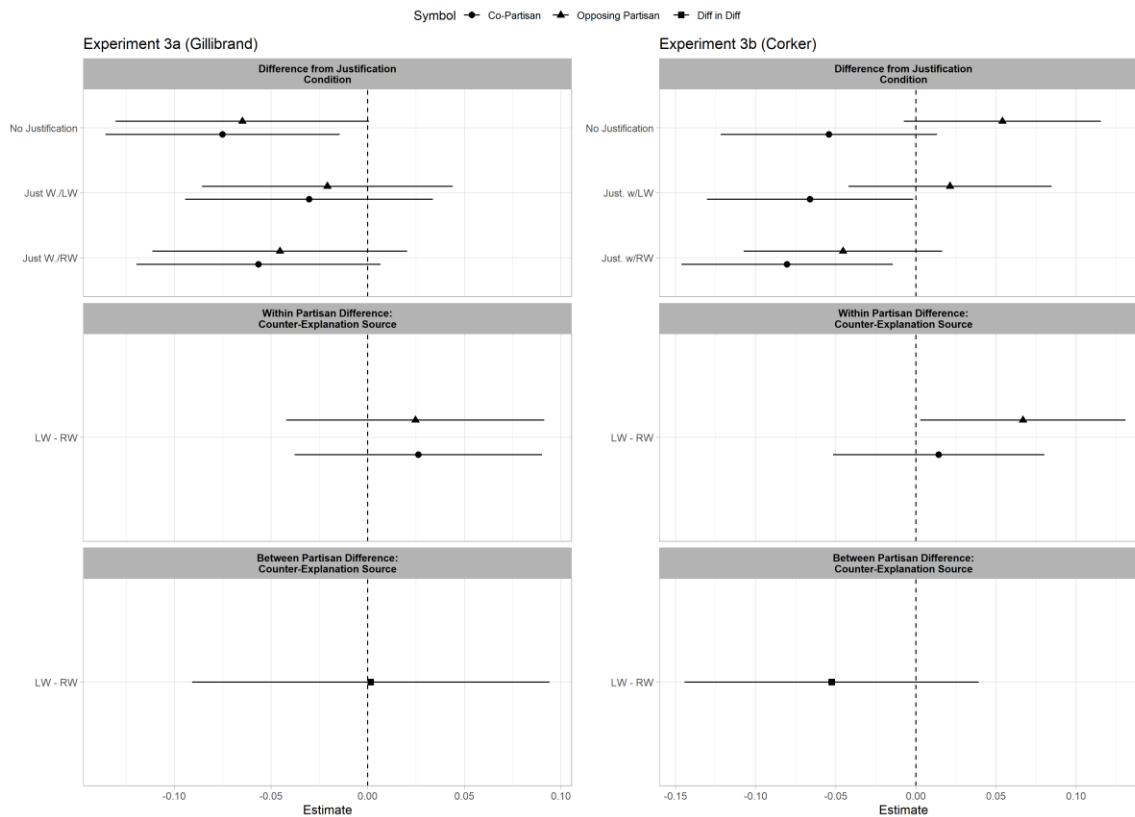


**Notes:** The top facet provides the average difference in evaluations based on treatment condition with separate markers for co-partisans (circle) and opposing partisans (triangles). The middle facet provides the difference *within* partisan group between the counter-explanation coefficients. The final facet provides a difference in difference (e.g., Co-Partisan[LW-RW]– Opposing Partisan[LW-RW]).

Turning to the influence of credibility, the bottom and middle panels provide the difference in coefficients between those assigned to the left-wing and right-wing sources within partisan grouping (middle panel) and then across partisanship (bottom panel). If co-partisans pay more attention to credibility then we would expect to see significant *negative* coefficients in Experiment 3a (since co-partisans here are Democrats and thus the left-wing source should be more influential) and a significant *positive* coefficient in

Experiment 3b (since co-partisans here are Republicans and hence should react more strongly to the right wing source).

Our results here are not consistent with Hypothesis 4. Co-partisans in the Corker experiment reported significantly worse evaluations relative to the justification only condition regardless of the source of the counter-explanation with the resulting difference between the two coefficients being statistically insignificant. Opposing partisans, i.e. Democrats, were actually significantly more responsive to the *right-wing* source, although this largely reflects the lack of influence of the left-wing source as seen in the top facet. While speculative, this could reflect a reasoning process wherein the right-wing source is seen as offering a 'costly' signal because they are acting in a counter-stereotypical manner that goes against their general partisan interests (e.g., Alt, Lassen, and Marshall 2015). The resulting difference in difference estimate, meanwhile, was also insignificant. Partisans in the Gillibrand experiment, meanwhile, did not react significantly different based on the source of the counter-explanation regardless of the test being done.

Figure 2 showed that politicians wishing to explain themselves have more to fear from credible non-partisan sources than from partisan actors. However, the evidence that this is because their co-partisan allies in the mass public are especially attentive to such sources appears limited. Co-partisans in Experiments 1, 2, and 3b felt significantly more negative toward the politician even when a lower credibility source impugned the politician's motives. Overall, these patterns show a greater openness to attacks on co-partisan politicians than prior work on motivated reasoning might suggest.

**Conclusion**

Democracies require balance between elected officials free to take the long view yet still responsive to constituent preferences. The literature on explanation giving suggests that legislators may use rhetoric to balance themselves on this edge, although this naturally raises worries that the efficacy of explanation giving will raise the temptation to shirk mass preferences (Grose, Malhotra, and Van Houweling 2015, 741). We contribute to the literature on explanation giving, as well as work on elite accountability and democratic representation, by showing that explanations provide a less secure way to balance these competing ends than previously thought. On the one hand, justifications do not always move public attitudes (see also: Peterson and Simonovits 2017). On the other, we have shown that rival explanations that root a politician's behavior in an ulterior motive can neuter the positive influence of a justification. Such counter-messages are particularly harmful when they stem from highly credible sources, but also appear to be effective when they come from more partisan and less credible sources. These studies converge on the conclusion that explanation giving as a means for blame management is a risky strategy that may mostly be effective when the behavior of the politician fails to trigger additional scrutiny.

The results reported here suggest several avenues for further investigation. One topic that deserves greater attention is the timing of explanations and counter-explanations. Participants in our experiments received both type of rhetoric at the same time much as occurs in most studies of explanation giving and competitive elite communications (e.g., Chong and Druckman 2007a; McGraw, Timpone, and Bruck 1993). However, exposure to explanations of elite behavior may be staggered over time; some may be exposed to the explanation first and then the rival account later in time

whereas others may receive the inverse pattern. This has two potential implications that future work could explore. First, it is plausible that elite explanations could still positively influence evaluations even after exposure to a counter-explanation provided the explanation comes first and generates a "strong" attitude toward the focal politician (Druckman, Fein, and Leeper 2012). If so, then elected officials would face the incentive of 'going public' with rationales for their policy actions as quickly as possible to inoculate supporters against alternative understandings of their behavior.

Second, it may be that the absence of a positive explanation effect seen in our experiments is a best-case scenario for politicians. The existing literature on elite explanations contrasts two counterfactuals: one where no explanation is provided and one where the explanation is present. However, the relevant counterfactuals for politicians may be one where it is the *counter-explanation* that stands alone versus one where it competes against the politician's own account. Perhaps in such instances a politician's best hope is to tread water by providing the explanation. In Online Appendix F we report the results of a small experiment that sheds some light on this possibility. There, participants were randomly assigned to conditions like those in Study 1 but with an additional wrinkle: some participants received the "counter"-explanation but not the explanation itself. As in text we find that justifications mollify otherwise upset respondents and that counter-explanations attenuated this effect. Notably, we did not find that the counter-explanation undermined evaluations relative to a baseline condition, although we do not wish to put too much weight on these results given that evaluations of the politician in the baseline condition were already quite low and the sample size of the

experiment is small. Regardless, future work is required to fully untangle how people respond to rival accounts for elite behavior over time.

Across our four experiments we saw inconsistent effects for the focal politician's justification. Justifications are not panaceas and, indeed, the existing literature on explanation giving does highlight cases where providing one fails to significantly move evaluations on average (McGraw 1991 Table 1; Peterson and Simonovits 2017). Some of this variation may stem from differences in the strength of the justification as weak or 'unsatisfactory' arguments do not influence public opinion (Chong and Druckman 2007a; McGraw 1991). However, future work could also benefit from considering whether the type of action involved influences receptiveness to justifications.[16] Experiments 2 and 3b suggest some initial possibilities. In the former case, the politician was unable to explain away an action wherein they were both clearly responsible and where the material costs to voters were direct and easy to understand. Perhaps justifications are more effective on 'hard' issues where individuals possess weaker prior attitudes and have a greater difficulty ascertaining the full consequences of elite actions. On the other hand, the experiment involving Corker highlights the potential role of polarization and identity. Corker's justification polarized partisans leading to an overall null effect of the justification. His actions were also closely tied with President Trump. One possibility is that actions which conjure an identity litmus test (if you take position A you are with them, if you take position B you are with us) may also limit a politician's ability to explain themselves, although the positive effect of Gillibrand's justification on the issue

---

[16] We thank an anonymous Reviewer for calling attention to this important question.

of gun control (Experiment 3a) cuts against this to a certain extent. This area is one that future research should unpack.

Our study has important potential implications for understanding elite incentives and behavior. On the one hand, our results may help explain a potential puzzle about elite behavior. The existing literature on explanation giving suggests that many are willing to give their legislators leeway provided a satisfactory explanation is forthcoming. However, politicians seem reluctant to fully trade in on their apparent ability to do so as suggested by several other literatures. For instance, politicians strategically focus on issues where their stance is already popular with voters (Canes-Wrone 2006). Otherwise, they try to remain ambiguous about where they stand to avoid alienating voters (Tomz and Van Houweling 2009). Policy changes, meanwhile, may be structured to obfuscate what is happening and who is responsible (Pierson 1994). Our results may provide one way of addressing this tension: politicians may anticipate the reactions of various 'instigators' and alter their behavior as a result.

The effectiveness of the character-based counter-explanations explored here, meanwhile, has important implications for elite strategy and broader normative debates. Our results suggest that politicians may face incentives to focus on intention-based counter-messages when attempting to rebut the persuasive appeals of rival actors. This type of message appears to be highly effective even in the hands of less credible speakers. They are likely easier to comprehend than much policy-based rhetoric. They also tie into a broadly held worry concerning the role of special interests in driving policy outcomes (Hibbing and Theiss-Morse 2002). An optimist might suggest that these types of messages serve an important purpose of motivating individuals to think again and thus be

more careful in simply accepting the stance of their elected officials. More pessimistically, however, the counter-explanations explored here worked by changing the grounds of debate from policy to character. If one goal of elite rhetoric is to educate the public (Mansbridge 2003), then the attractiveness of character based counter-explanations may have dire political implications insofar as politicians feel incentivized to orient their attention to the contents of their character and away from the contents of their policies.

**References**

Aalberg, Toril, Jesper Strömbäck, and Claes H. de Vreese. 2011. "The Framing of Politics as Strategy and Game: A Review of Concepts, Operationalizations and Key Findings." *Journalism* 13 (2): 162–78.

Adams, James, Lawrence Ezrow, and Zeynep Somer-Topcu. 2011. "Is Anybody Listening? Evidence That Voters Do Not Respond to European Parties' Policy Statements During Elections." *American Journal of Political Science* 55 (2): 370–82.

Alt, James E., David D. Lassen, and John Marshall. 2015. "Credible Sources and Sophisticated Voters: When Does New Information Induce Economic Voting?" *The Journal of Politics* 78 (2): 327–42.

Ansolabehere, Stephen, and PHilip Edward Jones. 2010. "Constituents' Responses to Congressional Roll-Call Voting." *American Journal of Political Science* 54 (3): 583–97.

Bianco, William T. 1994. *Trust: Representatives & Constituents*. Ann Arbor, MI: The University of Michigan Press.

Bittner, Amanda. 2011. *Platform or Personality? The Role of Party Leaders in Elections*. Oxford: Oxford University Press.

Bøggild, Troels. 2016. "How Politicians' Reelection Efforts Can Reduce Public Trust, Electoral Support, and Policy Approval." *Political Psychology* 37 (6): 901–19.

Bøggild, Troels, Lene Aarøe, and Michael Bang Petersen. n.d. "Citizens as Complicits: Distrust in Politicians and Biased Social Dissemination of Political Information." *American Political Science Review*.

Boudreau, Cheryl, and Scott A. MacKenzie. 2014. "Informing the Electorate? How Party Cues and Policy Information Affect Public Opinion About Initiatives." *American Journal of Political Science* 58 (1): 48–62.

Broockman, David E., and Daniel M. Butler. 2017. "The Causal Effects of Elite Position-Taking on Voter Attitudes: Field Experiments with Elite Communication." *American Journal of Political Science* 61 (1): 208–21.

Bullock, John G. 2011. "Elite Influence on Public Opinion in an Informed Electorate." *American Political Science Review* 105 (3): 496–515.

Butler, Daniel M., and Eleanor Neff Powell. 2014. "Understanding the Party Brand: Experimental Evidence on the Role of Valence." *The Journal of Politics* 76 (2): 492–505.

Canes-Wrone, Brandice. 2006. *Who Leads Whom? Presidents, Policy, and the Public*. Chicago: University of Chicago Press.

Chong, Dennis, and James N Druckman. 2007a. "Framing Public Opinion in Competitive Democracies." *American Political Science Review* 101 (4): 637–55.

———. 2007b. "Framing Theory." *Annual Review of Political Science* 10: 103–26. doi:10.1146/annurev.polisci.10.072805.103054.

———. 2010. "Dynamic Public Opinion: Communication Effects Over Time." *American Political Science Review* 104 (4): 663–80. doi:10.1017/S0003055410000493.

Clarke, Nick, Will Jennings, Jonathan Moss, and Gerry Stoker. 2018. *The Good Politician: Folk Theories, Political Interaction, and the Rise of Anti-Politics*. Cambridge: Cambridge University Press.

Clifford, Scott, Geoffrey Sheagley, and Spencer Piston. 2020. "Increasing Precision in Survey Experiments Without Introducing Bias." *APSA Preprints*. doi:10.33774/apsa-2020-tbd3c.

Coppock, Alexander, and Oliver A. McClellan. 2019. "Validating the Demographic, Political, Psychological, Experimental Results Obtained from a New Source of Online Survey Respondents." *Research & Politics* 6 (1): 1–14.

Cuddy, Amy J.C., Susan T. Fiske, and Peter Glick. 2008. "Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map." *Advances in Experimental Social Psychology* 40: 61–149.

Dahl, Robert A. 1971. *Polyarchy: Participation and Opposition*. New Haven: Yale University Press.

Doherty, David. 2015. "Perceived Motives in the Political Arena." *American Politics Research* 43 (3): 363–93.

Doherty, David, Conor M. Dowling, and Michael G. Miller. 2016. "When Is Changing Policy Positions Costly for Politicians? Experimental Evidence." *Political Behavior* 38 (2): 455–84.

Douglas, Anna. 2016. "NC GOP Senator Says He's Dropping Medicare Plan Attacked by

Opponent." *McClatchy DC*. http://www.mcclatchydc.com/news/politics-government/election/article100489837.html.

Druckman, James N., Jordan Fein, and Thomas J. Leeper. 2012. "A Source of Bias in Public Opinion Stability." *American Political Science Review* 106 (2): 430–54. doi:10.1017/S0003055412000123.

Eggen, Dan. 2009. "Industry Cash Flowed to Drafters of Reform; Key Senator Baucus Is a Leading Recipient." *The Washington Post*, July 21.

Fein, Steven, James L. Hilton, and Dale T. Miller. 1990. "Suspicion of Ulterior Motivation and the Correspondence Bias." *Journal of Personality and Social Psychology* 58 (5): 753–64.

Fenno, Richard F. 1978. *Home Style: House Members in Their Districts*. Scott Foresman & Co.

Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W.W. Norton & Company, Inc.

Grose, Christian, Neil Malhotra, and Robert P. Van Houweling. 2015. "Explaining Explanations: How Legislators Explain Their Policy Positions and How Citizens React." *American Journal of Political Science* 59 (3): 724–43.

Hibbing, John R., and John R. Alford. 2004. "Accepting Authoritative Decisions: Humans as Wary Cooperators." *American Journal of Political Science* 48 (1): 62–76.

Hibbing, John R, and Elizabeth Theiss-Morse. 2002. *Stealth Democracy: Americans' Beleifs about How Government Should Work*. Cambridge: Cambridge University Press.

Iyengar, Shanto, Helmut Norpoth, and Kyu S. Hahn. 2004. "Consumer Demand for Election News: The Horserace Sells." *Journal of Politics* 66 (1): 157–75.

Jerit, Jennifer. 2009. "How Predictive Appeals Affect Policy Opinions." *American Journal of Political Science* 53 (2): 411–26.

Kim, Seung Min. 2017. "Why Corker Flipped on the Tax Bill." *Politico*. https://www.politico.com/story/2017/12/18/bob-corker-tax-bill-kickback-republicans-respond-302482.

Kingdon, John W. 1989. *Congressman's Voting Decisions*. Ann Arbor, MI: University of

Michigan Press.

Lau, Richard R., Lee Sigelman, and Ivy Brown Rovner. 2007. "The Effects of Negative Political Campaigns: A Meta-Analytic Reassessment." *The Journal of Politics* 69 (4): 1176–1209.

Laustsen, Lasse, and Alexander Bor. 2017. "The Relative Weight of Character Traits in Political Candidate Evaluations: Warmth Is More Important than Competence, Leadership, and Integrity." *Electoral Studies* 49: 96–107.

Levendusky, Matthew S., and Michael C. Horowitz. 2012. "When Backing Down Is the Right Decision: Partisanship, New Information, and Audience Costs." *The Journal of Politics* 74 (2): 323–38.

Lodge, Milton, and Charles S Taber. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50 (3). Midwest Political Science Association: 755–69. http://www.jstor.org/stable/3694247.

Malle, Bertram F. 2011. "Time to Give Up the Dogmas of Attribution: An Alternative Theory of Behavior Explanation." *Advances in Experimental Social Psychology* 44: 297–352.

Malle, Bertram F., Steve Guglielmo, and Andrew E. Monroe. 2014. "A Theory of Blame." *Psychological Inquiry* 25: 147–86.

Mansbridge, Jane J. 2003. "Rethinking Representation." *American Political Science Review* 97 (4): 515–28.

Mayhew, David. 1974. *Congress: The Electoral Connection*. New Haven, CT: Yale University Press.

Mayo, Ruth. 2015. "Cognition Is a Matter of Trust: Distrust Tune Cognitive Processes." *European Review of Social Psychology* 26 (1): 283–327.

McDermott, Rose. 2011. "Internal and External Validity." In *Cambridge Handbook of Experimental Political Scence*, edited by James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia, 27–41. Cambridge: Cambridge University Press.

McGraw, Kathleen M. 1990. "Avoiding Blame: An Experimental Investigation of Political Excuses and Justifications." *British Journal of Political Science* 20 (1): 119–1131.

———. 1991. "Managing Blame: An Experimental Test of the Effects of Political Accounts." *The American Political Science Review* 85 (4): 1133–57.

———. 2001. "Political Accounts and Attribution Processes." In *Citizens and Politics: Perspectives from Political Psychology*, edited by James H Kuklinski, 160–98. Cambridge: Cambridge University Press.

———. 2003. "Political Impressions: Formation and Management." In *Oxford Handbook of Political Psychology*, 394–433. Oxford: Oxford University Press.

McGraw, Kathleen M., Samuel Best, and Richard Timpone. 1995. "'What They Say or What They Do?' The Impact of Elite Explanation and Policy Outcomes on Public Opinion." *American Journal of Political Science* 39 (1): 53–74.

McGraw, Kathleen M., Richard Timpone, and Gabor Bruck. 1993. "Justifying Controversial Political Decisions: Home Style in the Laboratory." *Political Behavior* 15 (3): 289–308.

Mullinix, Kevin J., Thoams L. Leeper, Jeremy Freese, and James N. Druckman. 2015. "The Generalizability of Survey Experiments." *Journal of Experimental Political Science* 2 (2): 109–38.

Nilsen, Ella. 2019. "Kirsten Gillibrand Used to Have Moderate Views on Immigration and Guns. Voters Want to Know Why She's Changed." *Vox*. https://www.vox.com/2019/4/9/18303526/kirsten-gillibrand-cnn-town-hall-immigration-guns.

Peterson, Erik, and Gabor Simonovits. 2017. "Costly Values: The Limited Benefits and Potential Costs of Targeted Policy Justifications." *Journal of Experimental Political Science*. doi:10.1017/XPS.2017.3P.

Pierson, Paul. 1994. *Dismantling the Welfare State? Reagan, Thatcher, and the Politics of Retrenchment*. Cambridge: Cambridge University Press.

Priester, Joseph R., and Richard E. Petty. 1995. "Source Attributions and Persuasion: Perceived Honesty as a Determinant of Message Scrutiny." *Personality and Social Psychology Bulletin* 21 (6): 637–54.

Robison, Joshua. 2017. "The Role of Elite Accounts in Mitigating the Negative Effects of Repositioning." *Political Behavior* 39 (3): 609–28. doi:10.1007/s11109-016-9372-6.

Tomz, Michael, and Robert P. Van Houweling. 2009. "The Electoral Implications of

Candidate Ambiguity." *American Political Science Review* 103 (1): 83–98.

Weaver, R. Kent. 1986. "The Politics of Blame Avoidance." *Journal of Public Policy* 6
(4): 371–98.

Weitz-Shapiro, Rebecca, and Matthew S. Winters. 2016. "Can Citizens Discern?
Information Credibility, Poiltical Sophistication, and the Punishment of Corruption
in Brazil." *The Journal of Politics* 79 (1): 60–74.

**Acknowledgements**

**Data Availability Statement**

Replication data for this paper can be found at the British Journal of Political Science's Dataverse.

**Supplementary Material**

A supplementary material file is available for this manuscript that contains the full model results, details on the samples used, pre-tests, and pre-analysis plans.

1. Online Appendix A: In-Text Models
2. Online Appendix OB: Further Experiment 1 Analyses and Treatment Wordings
3. Online Appendix OC: Further Experiment 2 Analyses and Treatment Wordings
4. Online Appendix D: Further Experiment 3 Analyses and Treatment Wordings
5. Online Appendix E: Counter-Explanations On Their Own: Supplementary Experiment
6. Online Appendix F: Pre-Analysis Plans

**Financial Support**

**Ethical Standards**

All participants provided written consent and were compensated for their participation.

**Supplementary Materials**

Can Elites Escape Blame by Explaining Themselves?
Suspicion and the Limits of Elite Explanations

**Contents**

**Online Appendix A: In-Text Models**

**Table OA1:** Sample Characteristics

| | Experiment 1 (MTurk) | Experiment 2 (MTurk) | Experiment 3 (Lucid) | ANES 2016 | ANES 2018 Pilot |
|---|---|---|---|---|---|
| **% Female** | 53.64 | 50.56 | 50.97 | 51.97 | 51.23 |
| **Age** | | | | | |
| Mean | 38.03 (12.29) | 37.66 (12.45) | 52.21 | 47.37 (SE: 0.36) | 47.16 (SE:0.47) |
| 18-24 | 8.98 | 10.48 | 11.24 | 11.98 | 10.62 |
| 25-34 | 39.70 | 40.26 | 13.42 | 16.53 | 19.43 |
| 35-44 | 25.06 | 24.07 | 14.26 | 14.52 | 16.62 |
| 45-54 | 12.67 | 12.10 | 9.65 | 17.27 | 16.99 |
| 55-64 | 9.31 | 8.81 | 9.31 | 19.48 | 17.23 |
| 65+ | 4.30 | 4.28 | 42.11 | 20.23 | 19.11 |
| **Education** | | | | | |
| <HS | 0.33 | 0.56 | 3.80 | 9.1 | 12.72 |
| HS | 9.88 | 10.07 | 28.02 | 28.86 | 27.5 |
| Some College/Associates | 37.53 | 37.44 | 35.12 | 30.15 | 31.0 |
| Bachelor | 37.69 | 37.31 | 20.58 | 18.76 | 18.2 |
| Post-Bachelor | 14.57 | 14.62 | 12.48 | 13.12 | 10.6 |
| **Race/Ethnicity** | | | | | |
| White | 77.35 | 73.15 | 72.55 | 69.71 | 63.96 |
| African American | 7.20 | 8.95 | 10.70 | 10.99 | 12.03 |
| Hispanic | 7.31 | 9.76 | 10.20 | 11.91 | 15.91 |
| Asian | 6.04 | 6.15 | 3.57 | 2.8 | 5.31 |
| Other | 2.10 | 1.99 | 2.99 | 4.59 | 2.78 |
| **Household Income (Median Category)** | $50,000 to $59,999 | $50,000 to $59,999 | $40,000 to $49,999 | $55,000-$59,999 | $40,000 - $49,999 |
| **PID** | | | | | |
| Mean | 3.48 (2.00) | 3.33 (2.06) | 3.86 (2.27) | 3.77 (SE:0.04) | 3.74 (se: 0.06) |
| % Democrat | 54.69 | 59.63 | 44.27 | 46.68 | 46.13 |
| % Republican | 30.54 | 30.19 | 38.75 | 39.49 | 35.53 |
| % Pure Independent | 14.77 | 10.19 | 16.98 | 13.82 | 18.33 |
| **Symbolic Ideology [a]** | | | | | |
| Mean | 3.55 (1.73) | 3.47 (1.77) | 3.97 (1.88) | 4.15 (SE: 0.04) | 3.93 (SE: 0.05) |
| % Liberal | 48.34 | 53.72 | 35.29 | 32.1 | 35.63 |
| % Moderate | 22.59 | 16.98 | 28.58 | 26.87 | 30.6 |
| % Conservative | 29.08 | 29.3 | 36.12 | 41.03 | 33.78 |

**Notes:** ANES estimates are based on weighted analyses. [a] The MTurk and ANES TS ideology measures use the same fully labeled scales with "extremely" liberal/conservative as the ends and moderate as the middle. The Lucid item uses the same wording as the ANES 2018, which asks whether one is very liberal, somewhat liberal, closer to liberals, neither liberal nor conservative, closer to conservatives, somewhat conservative, or very conservative.

**Table OA2**: Experiment 1 Models

| | (1)<br>Figure 1 | (2)<br>Figure 2 | (3)<br>Figure 3 |
|---|---|---|---|
| No Information Baseline | -0.12$^{**}$<br>(0.02) | -0.12$^{**}$<br>(0.02) | -0.05$^{+}$<br>(0.03) |
| No Justification | -0.16$^{**}$<br>(0.02) | -0.16$^{**}$<br>(0.02) | -0.16$^{**}$<br>(0.03) |
| Justification w/Counter | -0.19$^{**}$<br>(0.02) | | |
| Teacher Counter | | -0.20$^{**}$<br>(0.02) | -0.22$^{**}$<br>(0.03) |
| Non-Partisan Counter | | -0.22$^{**}$<br>(0.02) | -0.25$^{**}$<br>(0.03) |
| Partisan Counter | | -0.16$^{**}$<br>(0.02) | -0.13$^{**}$<br>(0.03) |
| **Legislator Partisanship** | | | |
| Democratic Legislator | 0.11$^{**}$<br>(0.01) | 0.11$^{**}$<br>(0.01) | |
| **Respondent Partisanship** | | | |
| Opposing Partisan | | | -0.20$^{**}$<br>(0.03) |
| Resp=Ind. | | | -0.06<br>(0.04) |
| **Interactions** | | | |
| No Information Baseline #<br>Opposing Partisan | | | -0.14$^{**}$<br>(0.04) |
| No Information Baseline #<br>Resp=Ind. | | | -0.12$^{*}$<br>(0.06) |
| No Justification # Opposing<br>Partisan | | | 0.00<br>(0.04) |
| No Justification # Resp=Ind. | | | -0.08<br>(0.06) |
| Teacher Counter # Opposing<br>Partisan | | | 0.06<br>(0.04) |
| Teacher Counter # Resp=Ind. | | | -0.06<br>(0.06) |
| Non-Partisan Counter #<br>Opposing Partisan | | | 0.07$^{+}$<br>(0.04) |
| Non-Partisan Counter #<br>Resp=Ind. | | | -0.03<br>(0.06) |

| | | | |
|---|---|---|---|
| Partisan Counter # Opposing Partisan | | | -0.01 |
| | | | (0.04) |
| Partisan Counter # Resp=Ind. | | | -0.08 |
| | | | (0.05) |
| Constant | 0.59** | 0.59** | 0.75** |
| | (0.01) | (0.01) | (0.02) |
| Observations | 1813 | 1813 | 1813 |
| Adjusted $R^2$ | 0.119 | 0.123 | 0.231 |

Standard errors in parentheses; (Baseline Categories: Justification Condition; Republican Legislator [Models 1 & 2] or Co-Partisan Legislator [Model 3])
[+] $p < 0.1$, [*] $p < 0.05$, [**] $p < 0.01$

**Table OA3**: Experiment 2 Models

| | (1) Fig 1 (In Text) | (2) Fig 1 (No Control) | (3) Fig 2 (In Text) | (4) Fig 2 (No Control) | (5) Fig 3 (In Text) | (6) Fig 3 (No Control) |
|---|---|---|---|---|---|---|
| No Justification | -0.01 (0.01) | -0.00 (0.02) | -0.01 (0.01) | -0.00 (0.02) | -0.01 (0.02) | -0.01 (0.02) |
| Just. w/Counter | -0.04** (0.01) | -0.04** (0.01) | | | | |
| HC Counter | | | -0.05** (0.01) | -0.05** (0.02) | -0.07** (0.02) | -0.06** (0.02) |
| LC Counter | | | -0.02* (0.01) | -0.03 (0.02) | -0.03* (0.02) | -0.05* (0.02) |
| Republican Legislator | 0.01 (0.01) | -0.03** (0.01) | 0.01 (0.01) | -0.03** (0.01) | | |
| Thermometer (Pre-Test) | 0.73** (0.02) | | 0.73** (0.02) | | 0.73** (0.02) | |
| **Partisanship** Opposing Partisan | | | | | -0.00 (0.02) | -0.16** (0.02) |
| Independent | | | | | -0.04 (0.03) | -0.16** (0.04) |
| **Interactions** No Justification # Opposing Partisan | | | | | -0.02 (0.02) | 0.01 (0.03) |
| No Justification # Independent | | | | | 0.06 (0.04) | 0.06 (0.06) |
| HC Counter # Opposing Partisan | | | | | 0.02 (0.02) | 0.03 (0.03) |
| HC Counter # Independent | | | | | 0.06+ (0.04) | 0.03 (0.05) |
| LC Counter # Opposing Partisan | | | | | 0.01 (0.02) | 0.04 (0.03) |
| LC Counter # Independent | | | | | 0.05 (0.04) | 0.08 (0.06) |
| Constant | 0.09** (0.01) | 0.49** (0.01) | 0.09** (0.01) | 0.49** (0.01) | 0.10** (0.02) | 0.56** (0.02) |
| Observations | 1610 | 1610 | 1610 | 1610 | 1609 | 1609 |
| Adjusted $R^2$ | 0.577 | 0.010 | 0.578 | 0.011 | 0.578 | 0.088 |

Standard errors in parentheses; Baseline Categories: Justification (all models); Democratic Legislator (model 1-4); Co-Partisan legislator (model 5-6); $^+ p < 0.1$, $^* p < 0.05$, $^{**} p < 0.01$

**Table OA4**: Experiment 3 Models, Figure 1 Analyses

| | (1) Gillibrand | (2) Gillibrand | (3) Corker | (4) Corker |
|---|---|---|---|---|
| No Justification | -0.06* | -0.07** | -0.01 | -0.00 |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| Just. w/Counter | -0.04+ | -0.04* | -0.04* | -0.05* |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| **Prior Politician:** | | | | |
| Unfavorable | | -0.26** | | -0.19** |
| | | (0.02) | | (0.02) |
| No Opinion | | -0.13** | | -0.11** |
| | | (0.02) | | (0.02) |
| **Co-Partisanship:** | | | | |
| Opposing Partisan | | -0.06** | | -0.04+ |
| | | (0.02) | | (0.02) |
| Independent | | -0.09** | | -0.06* |
| | | (0.02) | | (0.02) |
| **Issue Proximity:** | | | | |
| Lose Proximity | | -0.19** | | -0.13** |
| | | (0.02) | | (0.02) |
| No Attitude | | -0.10** | | -0.07** |
| | | (0.03) | | (0.03) |
| **Treatment Order:** | | | | |
| Gillibrand First | | 0.00 | | 0.00 |
| | | (0.01) | | (0.02) |
| Constant | 0.50** | 0.73** | 0.46** | 0.66** |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| Observations | 1211 | 1201 | 1212 | 1200 |
| Adjusted $R^2$ | 0.003 | 0.286 | 0.003 | 0.160 |

Standard errors in parentheses
Baseline Categories: Justification (all models); Favorable Prior; Gained Proximity; Co-Partisan; Corker First
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

**Table OA5**: Experiment 3 Models, Figure 2 Analyses

| | (1) Gillibrand | (2) Gillibrand | (3) Corker | (4) Corker |
|---|---|---|---|---|
| No Justification | -0.06* | -0.07** | -0.01 | -0.00 |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| LW Counter | -0.03 | -0.04+ | -0.02 | -0.03 |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| RW Counter | -0.05* | -0.05* | -0.06** | -0.07** |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| **Prior Politician:** | | | | |
| Unfavorable | | -0.26** | | -0.19** |
| | | (0.02) | | (0.02) |
| No Opinion | | -0.14** | | -0.11** |
| | | (0.02) | | (0.02) |
| **Co-Partisanship:** | | | | |
| Opposing Partisan | | -0.06** | | -0.03+ |
| | | (0.02) | | (0.02) |
| Independent | | -0.09** | | -0.06* |
| | | (0.02) | | (0.02) |
| **Issue Proximity:** | | | | |
| Lose Proximity | | -0.19** | | -0.13** |
| | | (0.02) | | (0.02) |
| No Attitude | | -0.10** | | -0.07** |
| | | (0.03) | | (0.03) |
| **Treatment Order:** | | | | |
| Gillibrand First | | 0.00 | | 0.00 |
| | | (0.01) | | (0.02) |
| Constant | 0.50** | 0.73** | 0.46** | 0.66** |
| | (0.02) | (0.02) | (0.02) | (0.02) |
| Observations | 1211 | 1201 | 1212 | 1200 |
| Adjusted $R^2$ | 0.003 | 0.286 | 0.004 | 0.162 |

Standard errors in parentheses
Baseline Categories: Justification (all models); Favorable Prior; Gained Proximity; Co-Paritsan; Corker First
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

**Table OA6**: Experiment 3 Models, Figure 4 Analyses

| | (1) Gillibrand | (2) Gillibrand | (3) Corker | (4) Corker |
|---|---|---|---|---|
| No Justification | -0.10** | -0.08* | -0.05 | -0.05 |
| | (0.03) | (0.03) | (0.04) | (0.03) |
| LW Counter | -0.03 | -0.03 | -0.07* | -0.07* |
| | (0.04) | (0.03) | (0.03) | (0.03) |
| RW Counter | -0.06+ | -0.06+ | -0.08* | -0.08* |
| | (0.04) | (0.03) | (0.04) | (0.03) |
| **Co-Partisanship** | | | | |
| Opposing Partisan | -0.23** | -0.07* | -0.21** | -0.09** |
| | (0.04) | (0.03) | (0.03) | (0.03) |
| Independent | -0.19** | -0.08+ | -0.14** | -0.06 |
| | (0.05) | (0.04) | (0.05) | (0.05) |
| **Interaction Terms** | | | | |
| No Justification # Opposing Partisan | 0.05 | 0.01 | 0.11* | 0.11* |
| | (0.05) | (0.05) | (0.05) | (0.05) |
| No Justification # Independent | 0.06 | 0.01 | -0.01 | 0.02 |
| | (0.07) | (0.07) | (0.06) | (0.06) |
| LW Counter # Opposing Partisan | 0.02 | 0.01 | 0.10* | 0.09+ |
| | (0.05) | (0.05) | (0.05) | (0.05) |
| LW Counter # Independent | -0.02 | -0.05 | 0.01 | 0.01 |
| | (0.07) | (0.06) | (0.07) | (0.06) |
| RW Counter # Opposing Partisan | 0.02 | 0.01 | 0.06 | 0.03 |
| | (0.05) | (0.05) | (0.05) | (0.05) |
| RW Counter # Independent | 0.02 | -0.00 | 0.01 | -0.00 |
| | (0.07) | (0.06) | (0.06) | (0.06) |
| **Politician Prior** | | | | |
| Unfavorable | | -0.26** | | -0.19** |
| | | (0.02) | | (0.02) |
| No Opinion | | -0.14** | | -0.11** |
| | | (0.02) | | (0.02) |
| **Proximity Status** | | | | |
| Lose Proximity | | -0.19** | | -0.13** |
| | | (0.02) | | (0.02) |
| No Attitude | | -0.10** | | -0.07** |
| | | (0.03) | | (0.03) |
| **Treatment Order** | | | | |
| Gillibrand First | | 0.01 | | 0.00 |
| | | (0.01) | | (0.02) |
| Constant | 0.62** | 0.73** | 0.58** | 0.69** |
| | (0.03) | (0.03) | (0.02) | (0.03) |
| Observations | 1210 | 1201 | 1211 | 1200 |
| Adjusted $R^2$ | 0.103 | 0.283 | 0.066 | 0.163 |

Standard errors in parentheses
Baseline Categories: Justification (all models); Favorable Prior; Gained Proximity; Co-Paritsan; Corker First
$^+ p < 0.1,\ ^* p < 0.05,\ ^{**} p < 0.01$

**Figure OA1:** Figure 3 (Experiment 2) Analyses sans Controls
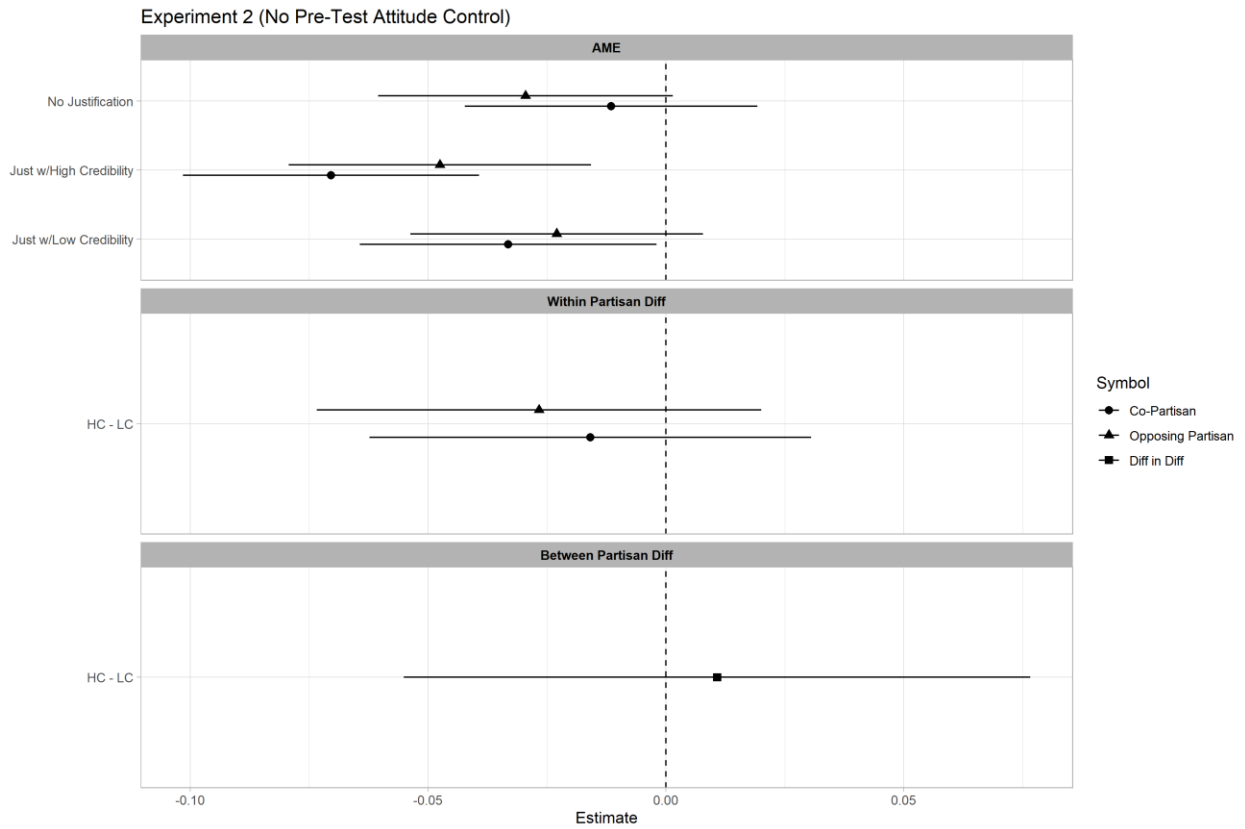


**Notes:** The top facet provides the average difference in evaluations based on treatment condition with separate markers for co-partisans (circle) and opposing partisans (triangles). The middle facet provides the difference *within* partisan group between the counter-explanation coefficients. The final facet provides a difference in difference.

**Figure OA2:** Figure 4 Analyses sans Controls
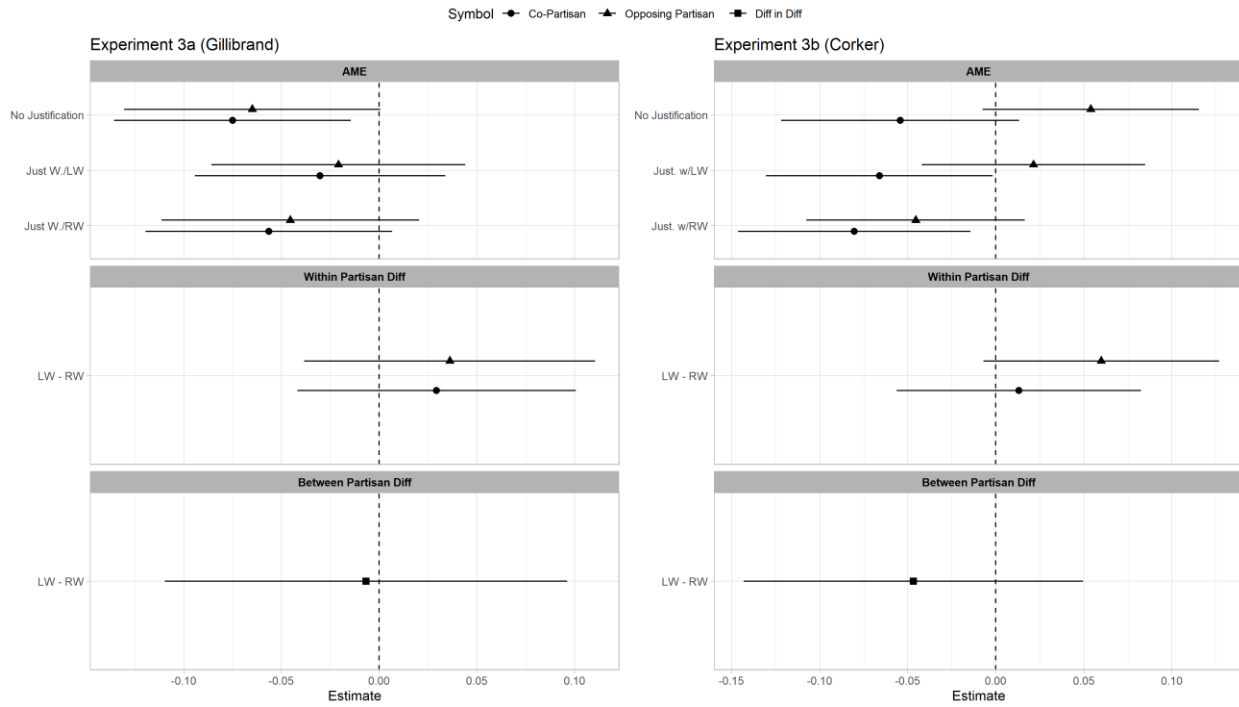


**Notes:** The top facet provides the average difference in evaluations based on treatment condition with separate markers for co-partisans (circle) and opposing partisans (triangles). The middle facet provides the difference *within* partisan group between the counter-explanation coefficients. The final facet provides a difference in difference (e.g., Co-Partisan[LW-RW]– Opposing Partisan[LW-RW]).

**Justifications, Our Study and the Literature**

One intriguing question raised by a reviewer concerns the comparability of the effects of the justifications in our study with those of the broader literature. Do we find significantly weaker effects of justifications than existing work?

We begin to address this question by comparing the effect sizes of receiving a justification (versus receiving none) from our four experiments with those found in experiments in Levendusky and Horowitz (2012), the two studies (involving four explanation treatments) reported in Robison (2017), and from Grose et al. (2015). We focus on these articles for pragmatic purposes. First, they each contain a no justification counter-factual – this is not true, for instance, of Butler and Broockman's (2017) study where individuals were not informed of the politician's disagreeable policy position sans justification. Second, these studies all make their underlying data publicly available which enabled us to calculate effect sizes. Perhaps the obvious omission here are the landmark studies by McGraw and her co-authors as these unfortunately neither make available the underlying data nor report the statistics needed to calculate effect sizes (e.g. mean evaluations and standard deviations by treatment condition) given their focus on the effects of account satisfaction (McGraw 1990, 1991; McGraw, Best, and Timpone 1995; McGraw, Timpone, and Bruck 1993).[1] Likewise, Peterson and Simonovits (2017) are missing from below for similar reasons. This is thus obviously not a systematic meta-analysis and should not be read as such, but rather as a first look at this literature.

The average effect size for these other studies is 0.37 (fixed effects), whereas it is 0.2 for our four studies combined. Figure OA3 provides a forest plot of all the experiments; combing our experiments with these experiments yields an average effect size of 0.31 (fixed effects model). The estimates for the studies that find a significant effect of the explanation tend to be fall within a rather narrow range of effect sizes around this 0.3 mark with our Experiment 1 and Levendusky and Horowitz clearly to the right of these effects. Three of the effects are not statistically significant with two of these stemming from our current study (Experiment 2 and 3b).

Our studies thus do show a lower average effect size due to the null effects in Experiments 2 and 3b. This could imply that we have shown that the effects of justifications are smaller than one might presuppose based on existing work. However, we are cautious in making this claim. The effect size for the three studies we consider here is likely an over-estimate. Peterson and Simonovits, for instance, are missing from our analysis and they find null effects of explanation giving in their study. Meanwhile, it is not clear that any explanations had a significant effect in McGraw (1991) judging from Table 2 of that study. While it is clearly the case that evaluations are more positive in that study among individuals who are highly satisfied with the explanation than among those who are unsatisfied, these evaluations nevertheless remain about the same level, or lower, than those in the Control group. If we include these as 'null' results, then the estimated effect of justifications outside of our studies would naturally come down.

Ultimately, we believe that the available evidence shows that justifications *can* work, but that we should also expect some heterogeneity in this effect (sans counter-explanation). As McGraw argues, explanations are likely to work insofar as they are deemed 'satisfactory' (much as frames are likely to be effective if they are 'strong'). Not all explanations will work. Likewise, not all actions may be 'explainable' as we discuss in the conclusion of the manuscript.

---

[1] These studies generally focus on regressions of post-test evaluations on explanation condition and (post-test) account satisfaction without first reporting the overall mean for those in the various conditions. McGraw (1991) does report means, but not standard deviations, for her experimental conditions, which would be required to calculate Cohen's *d* statistics.

**Figure OA3:** Forest Plot for Explanation Giving



| Source | SMD (95% CI) |
|---|---|
| Experiment 1 | 0.67 [ 0.50; 0.83] |
| Experiment 2 | 0.00 [-0.13; 0.14] |
| Experiment 3a (Gillibrand) | 0.20 [ 0.04; 0.36] |
| Experiment 3b (Corker) | 0.04 [-0.12; 0.20] |
| Levendusky & Horowitz | 0.52 [ 0.44; 0.61] |
| Robison 2017, Study 1, Soc Fairness | 0.19 [ 0.02; 0.37] |
| Robison 2017, Study 1, Comparison | 0.07 [-0.10; 0.25] |
| Robison 2017, Study 2, New Info | 0.42 [ 0.23; 0.61] |
| Robison 2017, Study 2, P. Fairness | 0.29 [ 0.10; 0.48] |
| Grose et al. 2015, Targeted Toward | 0.30 [ 0.16; 0.44] |
| Total (fixed effect) | 0.31 [ 0.26; 0.36] |
| Total (random effects) | 0.27 [ 0.13; 0.42] |
| Prediction interval | [-0.26; 0.80] |

Heterogeneity: $\chi_9^2 = 82.98$ ($P < .001$), $I^2 = 89\%$

Standardised Mean Difference (95% CI)

**Power and Combined Analyses**

Let us summarize the analyses we perform in text. We first examine the influence of justifications and counter-explanations, where we found that they led to significantly more positive evaluations in two of the four cases and that receiving a counter-explanation led to worse evaluations (relative to the justification condition) in all four experiments. We then consider the influence of credibility; here we find that counter-explanations were more damaging when from a high than low credible source in Experiment 1 and 2. We finally consider the role of partisanship, where we found at best inconsistent evidence that partisans paid greater attention to credibility.

One important question here concerns power: do we have sufficient power to detect significant differences?[2] This is an especially pertinent question regarding the analyses involving partisanship and counter-explanation credibility where we move from rather straightforward comparisons to ones involving not just one difference (i.e. are evaluations worse in the high credibility condition than in the justification condition) but two or more (i.e. is the difference in high credibility and justification conditions itself difference from the difference between low credibility and justification; does this difference in difference vary by respondent partisanship?). While the inclusion of pre-test covariates increases the power of our design, it may nevertheless be the case that we fail to find more consistent evidence for H4 because of an inability to reliably parse signal from noise

It is not immediately clear what type of power test to conduct for the complicated analyses of Hypothesis 4. As a beginning, then, we can consider power of our most straightforward analyses. In particular, we will consider what the minimum detectable effect is given our sample size, which is the smallest "true effect" obtainable for a given power and significance level (Bloom 1995). Table OA7 below provides the results of this exercise.[3] There, we show the sample size for each comparison relevant to the hypothesis, the observed Cohen's $d$ statistics when comparing the two conditions r, and the minimum detectable effect given a power of 0.8 and an alpha of 0.05. Note that the observed $d$ statistic is obtained from a t-test and thus does not include any covariates, which should increase the power of our analyses in Experiments 2 and 3a/3b. Broadly, we see that we have more than sufficient power in Experiment 1, and just about sufficient power in the remainder of the experiments to detect the effect we see in-text, although perhaps just barely in the latter cases (although in these latter cases the inclusion of covariates in our models should amplify power).

On the one hand, this is somewhat reassuring. On the other, however, the results for Experiments 2, 3a, and 3b, suggest that we may be running into issues of power for the more complex analyses implied by Hypothesis 3 and 4. If so, then what can be done? We obviously cannot go back in time to recruit more participants, but we can take one (non pre-registered) step available to us: combine the experiments into a single omnibus regression model. Doing so will naturally increase the number of observations for each test per cell and thus should improve the power of our analyses. However, this does have a drawback in that we cannot control for covariates in Experiment 1 and thus cannot do so in the combined analyses. Likewise, the credibility treatments in Experiments 3a/3b do not neatly map onto high and low credibility divisions which makes testing Hypotheses 3 and 4 more complicated.

Table OA8 provides analyses from three models that focus on Hypotheses 1-3. The first model regresses post-test evaluations on a three-category indicator for experimental treatment condition (Baseline: Justification; No Justification; Justification with any type of counter) and fixed effects for the experiment.[4] As in text we find a significant effect of the justification (as indicated by the *negative* coefficient for the No Justification condition) and likewise a significant negative effect of the counter-explanation. The difference between these two coefficients is also statistically significant (difference = -0.03 [-0.04, -0.01], F = 25.67, p < 0.01) indicating that evaluations were significantly worse when the counter-explanation was provided than when not.

The second two models focus on Hypothesis 3: high credibility counters are more influential than low credibility ones. Model 2 restricts the data to just Experiments 1 and 2 where we have a clear demarcation between high and

---

[2] We thank an anonymous Reviewer for calling attention to this question.
[3] Results were obtained using the "WebPower" package for R
[4] We also investigating fitting this model as a multilevel model with respondents nested within experiments and the effect of the treatment indicators allowed to vary across experiment. Doing so led to the same results.

low credibility sources. Consistent with Hypothesis 3 we find that the high credibility counter-explanation had a significantly stronger effect than the low credibility counter-explanation, although both undermined evaluations relative to the Justification condition. In Model 3 we add in data from Experiment 3. Here, we sorted respondents into high and low credibility based on the type of counter-explanation they received and their pre-test partisanship; thus, Democrats [Republicans] exposed to a left-wing [right-wing] source are coded as receiving a high credibility message while the inverse combinations are coded as low credibility messages. This transformation means that we lose Independents from Experiment 3 in these latter analyses. Adding Experiment 3 leads to smaller coefficients for the two counter-explanation conditions, but both remain negative and statistically significant and, as the bottom of Table OA8 shows, the difference between the two remains significant as well.

Tables OA9 and OA10 provide the results for tests of Hypothesis 4. Table OA9 regress evaluations on treatment condition, partisanship, and their interaction again with separate models for analyses conducted with just Experiments 1 and 2 versus all experiments. Table OA10, meanwhile, provides the marginal effect of assignment to the counter-explanation conditions for co- and opposing partisans; the difference in marginal effects within partisan grouping; and, finally, the between party differences. We see quite similar results to in-text when we just focus on Experiments 1 and 2; the high credibility counter-explanation effect had a significantly greater effect for co-partisans but a smaller and less precise one for opposing partisans, although the resulting difference in difference is not itself statistically significant. Adding Experiments 3a and 3b into the mix leads to estimates of the influence of the low credibility message that are greater in size (due to the effectiveness of the right-wing message among Democrats in Experiment 3a) and thus to a more precisely measured but now small difference in difference estimate. Combining the data sources thus does not provide additional evidence in favor of Hypothesis 4, suggesting that our failure to find more consistent evidence in support of this claim is not being driven solely by power considerations.

**Table OA7:** Minimum Detectable Effects (Hypotheses 1 & 2)

| Experiment | Condition 1 | Condition 2 | Hypothesis | n | Observed d | MDE |
|---|---|---|---|---|---|---|
| Exp. 1 | Justification | No Justification | H1 | 605 | 0.66 | 0.11 |
| Exp. 1 | Justification | Teachers | H2 | 599 | 0.84 | 0.11 |
| Exp. 1 | Justification | Non-Partisan | H2 | 601 | 0.92 | 0.11 |
| Exp. 1 | Justification | Partisan | H2 | 605 | 0.67 | 0.11 |
| Exp. 1 | Teachers | Partisan | H3 | 606 | 0.19 | 0.11 |
| Exp. 1 | Non-Partisan | Partisan | H3 | 608 | 0.25 | 0.11 |
| Exp. 2 | Justification | No Justification | H1 | 806 | 0.01 | 0.1 |
| Exp. 2 | Justification | High Credibility | H2 | 807 | 0.22 | 0.1 |
| Exp. 2 | Justification | Low Credibility | H2 | 810 | 0.11 | 0.1 |
| Exp. 2 | High Credibility | Low Credibility | H3 | 805 | 0.11 | 0.1 |
| Exp 3a. | Justification | No Justification | H1 | 603 | 0.2 | 0.11 |
| Exp 3a. | Justification | Left Wing | H2 | 604 | 0.1 | 0.11 |
| Exp 3a. | Justification | Right Wing | H2 | 609 | 0.16 | 0.11 |
| Exp 3a. | Left Wing | Right Wing | H3 | 608 | 0.07 | 0.11 |
| Exp. 3b | Justification | No Justification | H1 | 611 | 0.04 | 0.11 |
| Exp. 3b | Justification | Left Wing | H2 | 609 | 0.08 | 0.11 |
| Exp. 3b | Justification | Right Wing | H2 | 614 | 0.21 | 0.11 |
| Exp. 3b | Left Wing | Right Wing | H3 | 602 | 0.13 | 0.11 |

**Table OA8**: Combining the Experiments H1-H3

| | (1) Hyp 1 & 2 | (2) Hyp 3, Exp 1 & 2 | (3) Hyp 3, All Exp |
|---|---|---|---|
| No Justification | -0.0522** (0.0103) | -0.0670** (0.0126) | -0.0522** (0.0103) |
| Justification w/Counter | -0.0778** (0.00892) | | |
| High Credibility | | -0.126** (0.0119) | -0.0890** (0.0102) |
| Low Credibility | | -0.0821** (0.0127) | -0.0539** (0.0107) |
| Experiment 2 | -0.0497** (0.00856) | -0.0560** (0.00859) | -0.0525** (0.00859) |
| Experiment 3a | -0.0450** (0.0108) | | -0.0420** (0.0113) |
| Experiment 3b | -0.0667** (0.0103) | | -0.0643** (0.0107) |
| Constant | 0.558** (0.00930) | 0.581** (0.0104) | 0.557** (0.00940) |
| Observations | 5544 | 3121 | 5319 |
| | | Difference in Credibility Effects | |
| High Credibility – Low Credibility | | -0.04 [-0.07, -0.02] | -0.04 [-0.06, -0.02] |

Standard errors in parentheses
$^+ p < 0.1$, $^* p < 0.05$, $^{**} p < 0.01$

**Table OA9**: Combining the Experiments: Hypothesis 4

| | (1)<br>Exp 1 & 2 | (2)<br>All Exp |
|---|---|---|
| No Justification | -0.0703** | -0.0747** |
| | (0.0167) | (0.0140) |
| High Credibility | -0.142** | -0.108** |
| | (0.0162) | (0.0138) |
| Low Credibility | -0.0823** | -0.0775** |
| | (0.0171) | (0.0147) |
| Opposing Partisan | -0.171** | -0.193** |
| | (0.0181) | (0.0147) |
| Independent | -0.0860** | -0.132** |
| | (0.0293) | (0.0239) |
| No Justification # Opposing Partisan | 0.00477 | 0.0404+ |
| | (0.0251) | (0.0212) |
| No Justification # Independent | -0.0415 | -0.00273 |
| | (0.0386) | (0.0323) |
| High Credibility # Opposing Partisan | 0.0371 | 0.0330 |
| | (0.0241) | (0.0203) |
| High Credibility # Independent | -0.0290 | -0.0115 |
| | (0.0361) | (0.0315) |
| Low Credibility # Opposing Partisan | 0.0115 | 0.0429* |
| | (0.0254) | (0.0211) |
| Low Credibility # Independent | -0.0170 | 0.0297 |
| | (0.0392) | (0.0350) |
| Experiment 2 | -0.0606** | -0.0585** |
| | (0.00817) | (0.00818) |
| Experiment 3a | | -0.0543** |
| | | (0.0107) |
| Experiment 3b | | -0.0661** |
| | | (0.0103) |
| Constant | 0.672** | 0.666** |
| | (0.0130) | (0.0114) |
| Observations | 3120 | 5316 |

Standard errors in parentheses

$^+ p < 0.1$, $^* p < 0.05$, $^{**} p < 0.01$

**Table OA10:** Marginal Effects and Difference in Difference Tests

| | (1)<br>Just Exp 1 & Exp 2 | (2)<br>All Experiments |
|---|---|---|
| | High Credibility Marginal Effects | |
| Co-Partisan | -0.142$^{**}$<br>[-0.173, -0.110] | -0.108$^{**}$<br>[-0.135, -0.0805] |
| Opposing<br>Partisan | -0.104$^{**}$<br>[-0.140, -0.0692] | -0.0745$^{**}$<br>[-0.104, -0.0449] |
| | Low Credibility Marginal Effects | |
| Co-Partisan | -0.0823$^{**}$<br>[-0.116, -0.0488] | -0.0775$^{**}$<br>[-0.106, -0.0486] |
| Opposing<br>Partisan | -0.0708$^{**}$<br>[-0.108, -0.0339] | -0.0346$^{*}$<br>[-0.0647, -0.00438] |
| | Within Partisan Difference | |
| Co-Partisan | -0.06$^{**}$<br>[-0.09, -0.03] | -0.03$^{*}$<br>[-0.06, -0.002] |
| Opposing<br>Partisan | -0.03$^{+}$<br>[-0.07, 0.01] | -0.04$^{**}$<br>[-0.07, -0.01] |
| | Between Partisan Difference | |
| Co-Partisan [HC<br>– LC] –<br>Opposing<br>Partisan<br>[HC – LC] | -0.03<br>[-0.07, 0.02] | 0.01<br>[-0.03, 0.05] |
| *N* | 3120 | 5316 |

95% confidence intervals in brackets
$^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$

**Online Appendix OB: Experiment 1**

**Treatment Wordings**

**Table OB1:** Treatment Wordings, Experiment 1

[Background Information; Received by All:
Representative A is a member of the [Democratic/Republican] Party and has been in the state legislature for twelve years. During this time, Representative A has served on the Appropriations Committee, which oversees budgeting. Representative A has also served as Chair of the Energy Committee. Each term a group called the Americans for Democratic Action provides a score concerning the voting record of legislators. The score ranges from 0-100 with higher scores indicating a more liberal voting record. In its most recent publication, Representative A received a score of [80/20] from the ADA.

[Policy Treatment:
During the previous session, the Appropriations Committee voted on an amendment to the annual budget plan. The amendment proposed a change to the criteria used to distribute education funds to local communities, resulting in cuts in funding for some communities, including Representative A's district, but increases for others. Representative A voted in favor of this amendment.

[Justification:
When asked about the vote, Representative A said: "I voted for the amendment because I believe that it will make the distribution of school funding across the state fairer for all citizens than under the present budget, with more money going to those who need the funding the most."

[Counter-Narrative:
However, [local teachers / non-partisan experts on education policy / [Republican/Democratic] legislators in the state house] quoted in media reports about the cuts say that campaign donors of Representative A benefit financially from the amendment and that this is really why the Representative voted for it.

**Notes:** The Democratic version always had an ideological score of 80, while the Republican always received a score of 20. The partisan counter-narrative source was paired with the partisanship of Representative A; when the latter was a Democrat, then the partisan treatment read "Republican legislators in the state house", etc.

**Pre-Test: Anger and Cutting Economic Spending**

One question may be whether these voting to cut education funding is a controversial action. We thus fielded a separate sample of 301 respondents on MTurk. Respondents to this survey were restricted from participating in the experiments. We asked respondents the following question:

> There has been a good deal of discussion recently about how local and state governments can best avoid budget problems. One potential way to avoid such issues is by cutting spending. Below is a list of areas where governments could cut spending. How angry would you be if your local government cut spending in these areas?

Respondents could then indicate their level of anger on a 1 (not angry at all) to five (very angry) fully labeled scale for the following areas of spending: animal control, community events, sanitation, the fire department, parks and recreation, the school department, the policy department, transportation, and libraries. The order of the items was randomly assigned per respondent. Table OB2 provides the mean levels of anger per program. Cuts to the education spending earns a great deal of ire including from all respondents and from both Democrats and Republicans, albeit more from the former the latter.

**Table OB2:** Mean Anger for Program Cuts

|  | Means | Democrats | Republicans | t-test |
|---|---|---|---|---|
| Animal Control | 2.63 (1.12) | 2.68 (1.08) | 2.5 (1.33) | t=1.27; p = 0.20 |

| | | | | |
|---|---|---|---|---|
| Community Events | 2.24 (1.11) | 2.37 (1.14) | 2.13 (1.01) | t=1.66; p = 0.10 |
| Sanitation | 3.47 (1.16) | 3.55 (1.10) | 3.28 (1.18) | t=1.85; p = 0.07 |
| Fire Dept. | 3.72 (1.19) | 3.67 (1.18) | 3.82 (1.20) | t=0.99; p = 0.32 |
| Parks & Recreation | 2.98 (1.19) | 3.19 (1.15) | 2.69 (1.17) | t = 3.28; p = 0.001 |
| School Dept. | 3.78 (1.28) | 3.98 (1.15) | 3.58 (1.37) | t = 2.53; p = 0.01 |
| Police Dept. | 3.41 (1.23) | 3.21 (1.33) | 3.87 (1.18) | t=3.95; p = 0.0001 |
| Transportation | 2.99 (1.23) | 3.16 (1.23) | 2.76 (1.18) | t=2.52; p = 0.01 |
| Libraries | 3.17 (1.23) | 3.36 (1.20) | 2.85 (1.19) | t=3.28; p = 0.001 |
| N = | 301 | 163-4 | 92 | |

**Pre-Test: Explanation Strength and Source Credibility**

Prior to Study 1 we completed pre-tests to identify a 'strong' justification as well as to identify credible and less credible counter-narrative providers.

*Justification Strength Pre-Test Results*

We first recruited 322 respondents from MTurk to assess justification strength. [As with all of our pre-tests, these respondents are restricted from participating in our experiments.] The survey begins with the following introduction:

> On the next few pages we will ask you to imagine a scenario where your state legislature has voted on an amendment to the annual budget plan. In addition, you will be provided with a prospective explanation that a legislator might offer for taking a particular position on this amendment. Please read this account as if it was provided by your own state representative in this situation

Respondents were then (randomly) provided with a vignette where the legislator either voted *for* cuts to spending for his/her district or voted *against* an amendment that would have led to increased spending for his/her district. The legislator could offer one of four randomly presented arguments: one based in fairness; a tailored explanation; one based on personal conscience; and one based either on costs (against spending) or benefits (for cuts). The wording for each justification is available on request and will be posted with the replication materials.

After receiving the vignette, respondents were asked how satisfied they were with the explanation (1 = extremely dissatisfied, 7 = extremely satisfied); how effective they found it (1 = definitely not effective, 7 = definitively effective); how angry they would be to their state legislator in this situation (1 = not at all, 5 = extremely); and how much credit or blame the legislator deserves for taking the position (1 = great deal of blame, 5 = great deal of credit). We provide summary statistics for all four measures as well as a summary measure of 'warmth' toward the vignette legislator formed from a factor analysis of the four items in Table OB3 below. Two things were apparent to us. First, the vignette legislator was liked less when cutting spending than when obtaining benefits; we thus went with this behavior in our experiment to provide a stronger test for the initial explanation effect hypothesis. Second, the fairness account appeared to be the most effective of the four, hence its use in our experiment.

**Table OB3:** Justification Strength Statistics

| | Satisfaction | Effectiveness | Anger (Higher = more) | Blame (Higher = Blame) | Factor: Warmth |
|---|---|---|---|---|---|
| **Scenario** | | | | | |
| Against; Tailored | 4.10 (1.43) | 4.05 (1.43) | 2.00 (0.95) | 3.18 (0.97) | -0.19 (0.78) |
| Against; Conscience | 4.31 (1.83) | 4.23 (1.77) | 1.81 (0.98) | 2.88 (1.21) | 0.02 (1.01) |

| | | | | | |
|---|---|---|---|---|---|
| Against; Fairness | 5.47 (1.29) | 5.24 (1.41) | 1.53 (0.88) | 2.22 (1.05) | 0.67 (0.84) |
| Against; Costs | 4.45 (1.44) | 4.54 (1.43) | 2.06 (1.06) | 3.84 (1.0) | 0.05 (0.83) |
| For; Conscience | 4.11 (1.47) | 3.71 (1.54) | 2.06 (1.00) | 3.2 (1.16) | -0.27 (0.93) |
| For; Fairness | 4.76 (1.53) | 4.51 (1.53) | 1.93 (1.07) | 2.8 (1.18) | 0.15 (1.09) |
| For; Tailored | 3.57 (1.50) | 3.66 (1.56) | 2.32 (1.12) | 3.30 (0.97) | -0.50 (1.03) |
| For; Benefits | 4.33 (1.69) | 4.28 (1.60) | 2.02 (1.05) | 2.93 (1.16) | -0.04 (0.95) |
| | | | | | |
| **Policy** | | | | | |
| Against Spending Increase | 4.68 (1.56) | 4.60 (1.57) | 1.82 (0.97) | 2.72 (1.11) | 0.20 (0.92) |
| For Spending Cut | 4.16 (1.59) | 4.03 (1.59) | 2.10 (1.07) | 3.06 (1.12) | -0.18 (1.03) |
| | | | | | |
| **Explanation** | | | | | |
| Tailored | 3.79 (1.49) | 3.83 (1.53) | 2.18 (1.06) | 3.25 (0.97) | -0.36 (0.94) |
| Conscience | 4.2 (1.63) | 3.93 (1.65) | 1.95 (0.99) | 3.07 (1.18) | -0.14 (0.97) |
| Fairness | 5.14 (1.44) | 4.90 (1.50) | 1.72 (0.99) | 2.49 (1.14) | 0.44 (1.14) |
| Benefits/Costs | 4.38 (1.57) | 4.40 (1.53) | 2.04 (1.05) | 2.89 (1.10) | -0.001 (0.87) |

*Credibility Pre-Test Results*

Later on in the same survey we asked respondents to answer questions to tap the credibility of various speakers.

(1) how much [local teachers; the National Education Association; the American Federation of Teachers; the Democratic Party; and the Republican Party] knows will happen if the level of education funding is changed (1-4 scale, from nothing to a lot);

(2) how favorable or unfavorable they feel toward each group (1-5, from very unfavorable to very favorable)

(3) how trustworthy the group, or elected officials from the two parties, are (1-5 from very untrustworthy to very trustworthy).

Table OB4 breaks down the describe statistics for all four variables both overall and by respondent partisanship. Local teachers were highly rated on all four measures, hence our choice of them as 'highly credible' sources. Meanwhile, the Democratic and Republican parties do worse overall than the other three groups.

**Table OB4: Credibility**

| | Overall | Democrats | Republicans |
|---|---|---|---|
| **Knowledge (Range: 1-4)** | | | |
| Local Teachers | 2.98 (0.90) | 3.14 (0.90) | 3.03 (0.85) |
| NEA | 3.11 (0.81) | 3.24 (0.80) | 3.11 (0.79) |
| ATF | 3.04 (0.79) | 3.13 (0.81) | 3.01 (0.79) |
| Dem Party | 2.78 (0.87) | 2.90 (0.79) | 2.65 (0.95) |
| Rep Party | 2.61 (0.92) | 2.39 (0.96) | 2.95 (0.80) |
| | | | |
| **Favorability (Range: 1-5)** | | | |
| Local Teachers | 4.11 (0.90) | 4.31 (0.78) | 3.94 (0.90) |
| NEA | 3.25 (0.93) | 3.41 (0.91) | 3.09 (0.98) |
| ATF | 3.46 (0.90) | 3.65 (0.89) | 3.31 (1.01) |
| Dem Party | 2.93 (1.18) | 3.70 (0.87) | 2.16 (1.04) |

| | | | |
|---|---|---|---|
| Rep Party | 2.34 (1.20) | 1.63 (0.79) | 3.68 (0.82) |
| | | | |
| **Trustworthiness** | | | |
| Local Teachers | 3.95 (0.93) | 4.20 (0.80) | 3.79 (0.99) |
| NEA | 3.30 (0.92) | 3.46 (0.92) | 3.19 (0.98) |
| ATF | 3.43 (0.94) | 3.63 (0.95) | 3.24 (0.97) |
| Dem Party | 2.94 (1.14) | 3.54 (0.85) | 2.26 (1.11) |
| Rep Party | 2.41 (1.22) | 1.86 (1.05) | 3.46 (0.91) |

We were somewhat worried that local teachers might nevertheless be discounted given their potential material interest in the issue of education cuts. We thus performed a follow up pre-test with a different sample of 181 respondents from MTurk. In varying order, we provided the following question stem:

> "News stories concerning cuts in education funding often feature quotes from [local teachers; representatives of non-partisan research groups that specialize in education policy, such as the National Education Policy Center; politicians that support the proposed funding cuts; politics that oppose the proposed funding cuts]".

For each target we asked respondents to agree or disagree on a 1-7 scale to the following four statements:

1. they [the speaker] is well informed on the issue and understands the consequences of funding cuts;
2. they have a personal interest in whether funding cuts are passed;
3. they are trustworthy concerning the desirability of funding cuts;
4. they are a good source of information concerning why the cuts were passed or not passed.

Table OB5 below provides the summary statistics. Again, local teachers appear highly credible, being highly informed and trustworthy, although their personal interest in the matter may cut against this credibility somewhat. Meanwhile, the non-partisan source had a lower personal interest, while still being regarded as informative if not definitively trustworthy about the desirability of the cuts. More specifically political sources, meanwhile, were deemed less trustworthy and not particularly informed. We thus included the non-partisan source as a secondary highly credible source in the experiment.

**Table OB5:** Credibility

| | Local Teachers | Non-Partisan | Supporters | Opposes |
|---|---|---|---|---|
| Informed | 5.40 (1.53) | 5.08 (1.38) | 3.94 (1.76) | 4.52 (1.66) |
| Personal Interest | 5.99 (1.40) | 4.32 (1.69) | 5.14 (1.78) | 5.04 (1.56) |
| Trustworthy | 5.03 (1.66) | 4.31 (1.61) | 3.25 (1.65) | 3.96 (1.62) |
| Process | 4.72 (1.62) | 4.36 (1.64) | 3.88 (1.79) | 4.34 (1.66) |

Higher = agree with statement

**Additional Measures**

We asked respondents some further items about the legislator and their behavior related to our discussion of the psychology of blame. On the one hand, we asked respondents whether they agreed (+1), disagreed (-1) or neither (0) with the vote taken by the representative. On the other, we asked respondents to rate the importance of a variety of "possible reasons for Representative A's vote on the education-funding amendment to the budget plan". (Those in the No Policy Information condition were instead asked to rate the importance of the motives for explaining the representative's "voting behavior while in the state legislature.") We will examine two indices created via factor analysis of these items (M=0, SD=1). First, an index of 'good representative' motives formed from the items "desire to help constituents," "desire to help all state residents", and "desire to make good policy" ($\alpha = 0.84$). Second, we examine an index of negative 'political' motives formed from the items "pandering to voters," "influence of special interests", and "winning re-election" ($\alpha = 0.69$). It should be noted that the order in which respondents answered the policy approval and motives questions was randomly varied, i.e. some answered the policy approval item before the motives while others received the inverse order.

Tale OB6 provides the mean scores for these items by condition for all respondents and by respondent/legislator partisanship along with 95% confidence intervals. The patterns seen in Table OB6 closely resemble those in text. Those assigned to the Justification condition report greater agreement with the policy change and place more importance on "Good Representative" motives in explaining the legislators behavior than those in the No Policy Info or No Explanation counter-factuals. Ratings on these variables, however, are more in line with the two no explanation conditions when a counter-narrative is present. Again, credibility matters as the message from the Partisan Opponents is less effective at undermining evaluations than the counter-message from the two more credible speakers.

**Table OB6**: Policy Agreement and Motive Attributions

| | **All Respondents** | | | | | |
|---|---|---|---|---|---|---|
| | No Policy Info | No Expl. | Expl. | Teachers | Experts | Partisan |
| Policy Agreement | N/A | -0.08 (-0.16, 0.01) | 0.51 (0.43, 0.59) | -0.02 (-0.11, 0.08) | -0.06 (-0.16, 0.03) | 0.16 (0.07, 0.25) |
| "Political" Motives | 0.37 (0.26, 0.47) | -0.26 (-0.37, -0.14) | -0.31 (-0.44, -0.18) | 0.03 (-0.07, 0.14) | 0.06 (0.05, 0.17) | 0.12 (0.01, 0.22) |
| "Good Representative" Motives | 0.16 (0.05, 0.27) | -0.03 (-0.14, 0.09) | 0.28 (0.18, 0.37) | -0.14 (-0.26, -0.02) | -0.25 (-0.38, -0.13) | -0.02 (-0.12, 0.09) |
| | | | | | | |
| | **Co-Partisans** | | | | | |
| Policy Agreement | N/A | 0.12 (-0.01, 0.25) | 0.65 (0.54, 0.77) | 0.20 (0.04, 0.35) | 0.10 (-0.05, 0.24) | 0.49 (0.36, 0.62) |
| "Political" Motives | 0.28 (0.14, 0.43) | -0.34 (-0.51, -0.17) | -0.65 (-0.84, 0.45) | -0.10 (-0.26, 0.07) | 0.04 (-0.13, 0.21) | -0.04 (-0.23, 0.14) |
| "Good Representative" Motives | 0.49 (0.36, 0.62) | 0.24 (0.11, 0.38) | 0.51 (0.38, 0.64) | 0.18 (0.02, 0.34) | -0.02 (-0.19, 0.15) | 0.36 (0.21, 0.51) |
| | | | | | | |
| | **Opposing Partisans** | | | | | |
| Policy Agreement | N/A | -0.30 (-0.44, -0.17) | 0.38 (0.25, 0.51) | -0.14 (-0.29, 0.002) | -0.24 (-0.39, -0.09) | -0.08 (-0.22, 0.06) |
| "Political" Motives | 0.47 (0.30, 0.63) | -0.27 (-0.45, -0.09) | -0.09 (-0.28, 0.09) | 0.02 (-0.14, 0.19) | 0.14 (-0.02, 0.30) | 0.18 (0.03, 0.33) |
| "Good Representative" Motives | -0.13 (-0.32, 0.05) | -0.26 (-0.45, -0.07) | 0.08 (-0.08, 0.25) | -0.31 (-0.51, -0.11) | -0.42 (-0.62, -0.22) | -0.24 (-0.41, -0.08) |

**Notes:** Cells provide condition means alongside 95% confidence intervals.

## Online Appendix OC: Experiment 2

**Treatment Wordings**

**Figure OC1**: Initial Vignette about Politician

**Candidate Background**

Dennis Williams is a member of the Democratic Party and has served in the Vermont state legislature for fourteen years. During this time, Representative Williams has served on the Appropriations Committee, which oversees budgeting. Representative Williams has also served as Chair of the House Committee on Energy and Technology where he has a reputation for working across the aisle to get things done. Representative Williams won his last re-election campaign by a margin of 67% to 33%.

**Policy Positions**

Representative Williams lists the following policy positions on his office website:

- Supports increasing taxes for those making over $250,000 a year
- Opposes legalizing the purchase and possession of small amounts of marijuana
- Opposes the state government providing parents with vouchers to send their children to any school they choose
- Supports making laws covering the sale of firearms stricter than they currently are

**Endorsements**

In his most previous re-election campaign, Representative Williams received endorsements from the editorial board of the largest newspaper in his legislative district, the American Association of Retired Persons, and the Chamber of Commerce.

**Candidate Background**

Dennis Williams is a member of the Republican Party and has served in the Vermont state legislature for fourteen years. During this time, Representative Williams has served on the Appropriations Committee, which oversees budgeting. Representative Williams has also served as Chair of the House Committee on Energy and Technology where he has a reputation for working across the aisle to get things done. Representative Williams won his last re-election campaign by a margin of 67% to 33%.

**Policy Positions**

Representative Williams lists the following policy positions on his office website:

- Opposes increasing taxes for those making over $250,000 a year
- Opposes legalizing the purchase and possession of small amounts of marijuana
- Supports the state government providing parents with vouchers to send their children to any school they choose
- Supports making laws covering the sale of firearms stricter than they currently are

**Endorsements**

In his most previous re-election campaign, Representative Williams received endorsements from the editorial board of the largest newspaper in his legislative district, the American Association of Retired Persons, and the Chamber of Commerce.

**Notes**: This image provides the background information subjects received concerning the legislator as it was seen by respondents. The left-hand image shows the information received by those randomly assigned to read about a Democratic legislator, the right shows the Republican legislator.

**Figure OC2**: Initial Impressions of the Politician
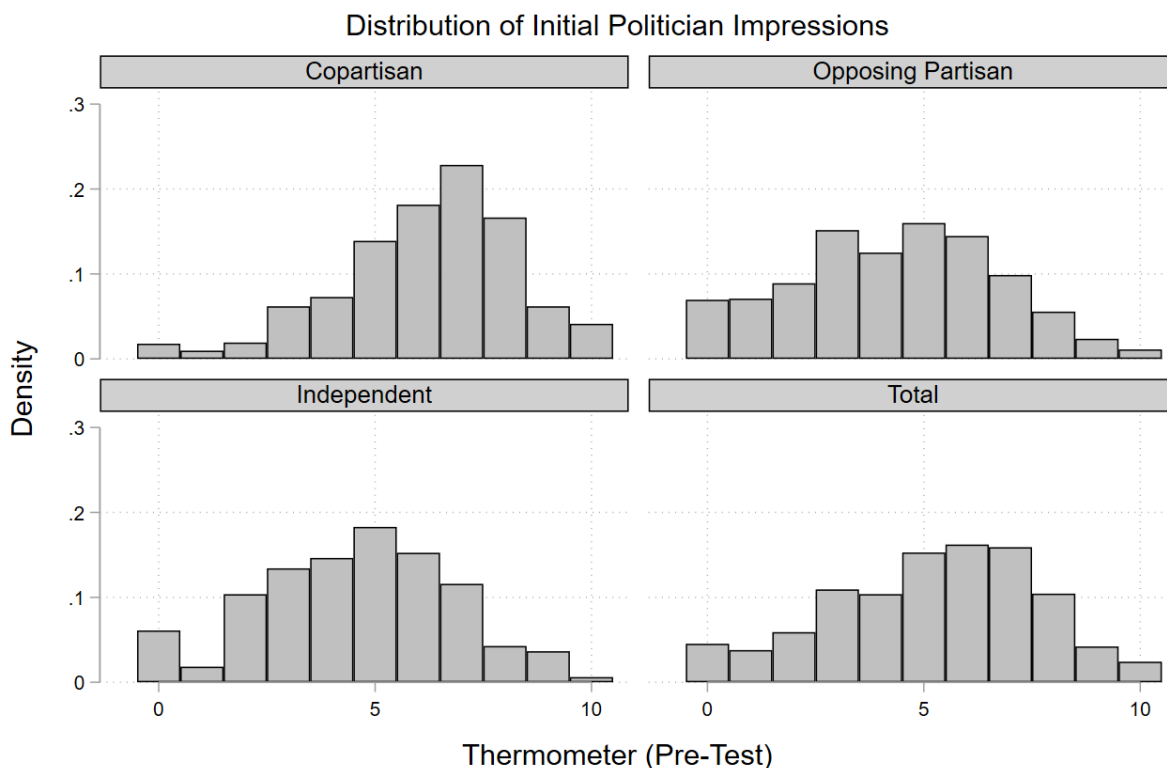


Note: The overall mean is 5.20 [5.09, 5.32]. Not surprisingly, this mean is significantly larger among co-partisans (6.21 [6.06, 6.36]) than both opposing partisans (4.33 [4.16, 4.51]) and Independents (4.57 [4.24, 4.93]). ANOVA: F = 135.34, p < 0.001.

**Article Wordings, Justification Portion of Experiment**

*Baseline (No Justification or Counter)*

MONTPELIER – Higher electric rates for Burlington Electric Department customers starting September 1 are one step closer to being imposed.

Tuesday night, the House Committee on Energy and Technology voted 4-3 to narrowly approve an amended version of the budget that includes the rate increase among other provisions. The deciding vote was cast by the Chair of the Committee, Representative Dennis Williams.

Under the increase, customers of the publicly-regulated Burlington Electric that use 1,000 kilowatt hours of electricity a month – what officials say is the average home – would see their monthly bills go up by an average of $10.42 to $107.20. That is a 10.77 percent jump. However, some customers would see higher-than-average rate hikes while others lower-than-average increases depending on their rate classes and overall electricity use.

The budget process now moves on to the full House for a final vote next week before the State Senate takes up its own budget bill. The key debate in the Senate will concern revenue. One potential amendment likely to be discussed is an increase in the state's sales tax from 7% to 7.3%. However, it is unclear whether there will be enough support for the amendment and most experts expect the Senate to adopt the House's budget before sending it on to the Governor for final approval.

Below are the wordings of the justification and counter-narratives. These always came after the "under the increase" paragraph and before the "the budget process now moved" paragraph; the latter paragraph always ended the article.

*Justification (4th paragraph in article, when it appears):*

Williams defended his vote to a group of constituents after the vote by noting the need for additional spending to keep electricity costs down in the long run. "This was a difficult decision," Williams said, "but if we do not make an investment now then prices will increase even more dramatically over time. Sometimes you have to make tough choices that you believe are in the best long-term interests of the community."

*High Credibility Counter-Narrative (5th paragraph in article when it appears):*

However, some were not convinced. Gary Allison, an economist at the non-partisan Center on Budget and Policy Priorities, noted that lobbyists for Burlington Electric Department had extensively lobbied committee members to vote for the rate increase. "The rate increase means a big cash inflow for Burlington Electric," Allison said, "and I'm sure they were holding out future campaign donations as a carrot to vote for the hike."

*Low Credibility Counter-Narrative (5th paragraph in article when it appears):*

However, some were not convinced. Gary Allison, a [Democratic/Republican] city council member trailing Williams in the polls in their upcoming primary election, noted that lobbyists for Burlington Electric Department had extensively lobbied committee members to vote for the rate increase. "The rate increase means a big cash inflow for Burlington Electric," Allison said, "and I'm sure they were holding out future campaign donations as a carrot to vote for the hike."

**Figure OC3:** Example of Policy Vote Treatment



**Burlington**
**Free Press**

# Vermont Legislature Passes Controversial Electricity Amendment

**Alex Jackson** | Free Press Staff Writer
Published 6:22 p.m. UTC April 11, 2018

MONTPELIER – Higher electric rates for Burlington Electric Department customers starting September 1 are one step closer to being imposed.

Tuesday night, the House Committee on Energy and Technology voted 4-3 to narrowly approve an amended version of the budget that includes the rate increase among other provisions. The deciding vote was cast by the Chair of the Committee, Representative Dennis Williams.

Under the increase, customers of the publicly-regulated Burlington Electric that use 1,000 kilowatt hours of electricity a month – what officials say is the average home – would see their monthly bills go up by an average of $10.42 to $107.20. That is a 10.77 percent jump. However, some customers would see higher-than-average rate hikes while others lower-than-average increases depending on their rate classes and overall electricity use.

**Notes:** A partial screengrab of what the newspaper article looked like to respondents. This specific example comes from the No Justification condition.

**Alternative Specifications**

In text we regressed respondents' post-test evaluations on treatment conditions and their pre-test evaluation. In Table OC1 we provide an alternative way of using this pre-test information: a difference score (post-test – pre-test; rescaled to range from 0-1). We see the same results as in-text: providing a justification (versus not) did not lead to better evaluations, while the counter-explanations undermined these evaluations. The high credibility account was again more effective (difference = -0.02 [-0.03, -0.01], F = 8.21, p < 0.01). And, as in text, we find that a more reliable difference in effectiveness based on credibility for co-partisans (-0.02 [-0.04, -0.004], F = 5.36, p < 0.05) than for opposing partisans (-0.02 [-0.04, 0.005], F = 2.3, p =0.13), but again with an insignificant difference in difference (-0.01 [-0.04, 0.02]).

We also report below the results of a final specification: one where we simplify the difference score to three categories: a negative change in evaluations (post-test < pre-test), no change in evaluations (post-test = pre-test), and a positive change in evaluations (post-test > pre-test). Table OC2 focuses on an analogue to our Figure 2 analyses where we do not combine the two credibility conditions. Respondents assigned to the justification condition were more likely to report a positive change in evaluations than where those assigned to the no explanation condition as indicated by the significant negative coefficient for No Justification in Table OC2. Meanwhile, those assigned to the High Credibility counter were the most likely to report a negative change and were significantly more likely to do so than those in the Justification condition.

Table OC2 provides the interaction between treatment assignment and co-partisanship while Figure OC4 plots the results of the interaction models. The markers provide the difference in probability of giving each response by treatment assignment (y-axis) and partisanship (circles = co-partisan, triangles = opposing partisans). Co-partisans were more likely than those in the justification condition to report 'worse' evaluations both when a low and high credibility source offered the counter-explanations, whereas opposing partisans only did so when the high credibility source was involved. This cuts against our prior results using the full scale items. The multinomial models thus provide a bit more evidence in favor of H1, and less for H4, than do the in-text models.

**Table OC1**: Difference Score

| | (1) Figure 1 | (2) Figure 2 | (3) Figure 3 |
|---|---|---|---|
| No Justification | -0.01 | -0.01 | -0.01 |
| | (0.01) | (0.01) | (0.01) |
| Just. w/Counter | -0.02$^{**}$ | | |
| | (0.01) | | |
| Republican Legislator | 0.01 | 0.01 | |
| | (0.00) | (0.00) | |
| Thermometer (Pre-Test) | -0.17$^{**}$ | -0.17$^{**}$ | -0.17$^{**}$ |
| | (0.01) | (0.01) | (0.01) |
| HC Counter | | -0.03$^{**}$ | -0.04$^{**}$ |
| | | (0.01) | (0.01) |
| LC Counter | | -0.02$^{*}$ | -0.02$^{*}$ |
| | | (0.01) | (0.01) |
| Opposing Partisan | | | -0.00 |
| | | | (0.01) |
| Independent | | | -0.03 |
| | | | (0.02) |
| No Justification # Opposing Partisan | | | -0.01 |
| | | | (0.01) |
| No Justification # Independent | | | 0.04 |
| | | | (0.03) |
| HC Counter # Opposing Partisan | | | 0.01 |
| | | | (0.01) |
| HC Counter # Independent | | | 0.04$^{+}$ |
| | | | (0.02) |
| LC Counter # Opposing Partisan | | | 0.01 |
| | | | (0.01) |
| LC Counter # Independent | | | 0.03 |
| | | | (0.02) |
| Constant | 0.68$^{**}$ | 0.68$^{**}$ | 0.69$^{**}$ |
| | (0.01) | (0.01) | (0.01) |
| Observations | 1610 | 1610 | 1609 |
| Adjusted $R^2$ | 0.160 | 0.164 | 0.163 |

Standard errors in parentheses
Baseline Categories: Justification (all models); Democratic Legislator (model 2); Co-Partisan legislator (model 3)
$^{+} p < 0.1$, $^{*} p < 0.05$, $^{**} p < 0.01$

**Table OC2**: Multinomial Logit Models, Experiment 2

| | Negative Change | No Change [Baseline Category] | Positive Change |
|---|---|---|---|
| No Justification | -0.02 | 0.00 | -0.49* |
| | (0.15) | (.) | (0.21) |
| HC Counter | 0.57** | 0.00 | -0.30 |
| | (0.16) | (.) | (0.22) |
| LC Counter | 0.25 | 0.00 | -0.17 |
| | (0.16) | (.) | (0.21) |
| Opposing Partisan | -0.37** | 0.00 | 0.33+ |
| | (0.11) | (.) | (0.17) |
| Independent | -0.55** | 0.00 | 0.34 |
| | (0.19) | (.) | (0.26) |
| Constant | 0.11 | 0.00 | -1.04** |
| | (0.12) | (.) | (0.17) |
| Observations | | 1609 | |
| Pseudo $R^2$ | | 0.017 | |

Predicted Probabilities

| | | | |
|---|---|---|---|
| No Justification | 0.41 [0.36, 0.46] | 0.46 [0.42, 0.51] | 0.12 [0.09, 0.16] |
| Justification | 0.39 [0.34, 0.43] | 0.43 [0.38, 0.47] | 0.19 [0.15, 0.22] |
| HC Counter | 0.55 [0.50, 0.60] | 0.34 [0.30, 0.39] | 0.11 [0.08, 0.14] |
| LC Counter | 0.46 [41, 0.51] | 0.39 [0.35, 0.44] | 0.15 [0.11, 0.18] |

Standard errors in parentheses
Reference category is 'No Change in Thermometer'
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

**Table OC3**: Multinomial Logit Models: Interaction, Experiment 2

| | Baseline Category: No Change in Evaluation | |
| --- | --- | --- |
| | Negative Change | Positive Change |
| No Justification | 0.32 | -0.10 |
| | (0.22) | (0.34) |
| HC Counter | 0.94$^{**}$ | 0.09 |
| | (0.23) | (0.37) |
| LC Counter | 0.66$^{**}$ | -0.01 |
| | (0.23) | (0.36) |
| Opposing Partisan | 0.02 | 0.70$^{*}$ |
| | (0.23) | (0.30) |
| Independent | 0.75$^{+}$ | 0.74 |
| | (0.42) | (0.56) |
| No Justification # Opposing Partisan | -0.38 | -0.68 |
| | (0.32) | (0.45) |
| No Justification # Independent | -2.34$^{**}$ | -0.64 |
| | (0.67) | (0.77) |
| HC Counter # Opposing Partisan | -0.47 | -0.59 |
| | (0.33) | (0.48) |
| HC Counter # Independent | -1.65$^{**}$ | -0.79 |
| | (0.54) | (0.75) |
| LC Counter # Opposing Partisan | -0.70$^{*}$ | -0.37 |
| | (0.33) | (0.45) |
| LC Counter # Independent | -1.25$^{*}$ | 0.01 |
| | (0.58) | (0.76) |
| Constant | -0.16 | -1.25$^{**}$ |
| | (0.16) | (0.23) |
| Observations | 1609 | |
| Pseudo $R^2$ | 0.024 | |

Standard errors in parentheses
Reference category is 'No Change in Thermometer'
$^{+} p < 0.1$, $^{*} p < 0.05$, $^{**} p < 0.01$

**Figure OC4:**



Difference from Justification Condition in Response Probability
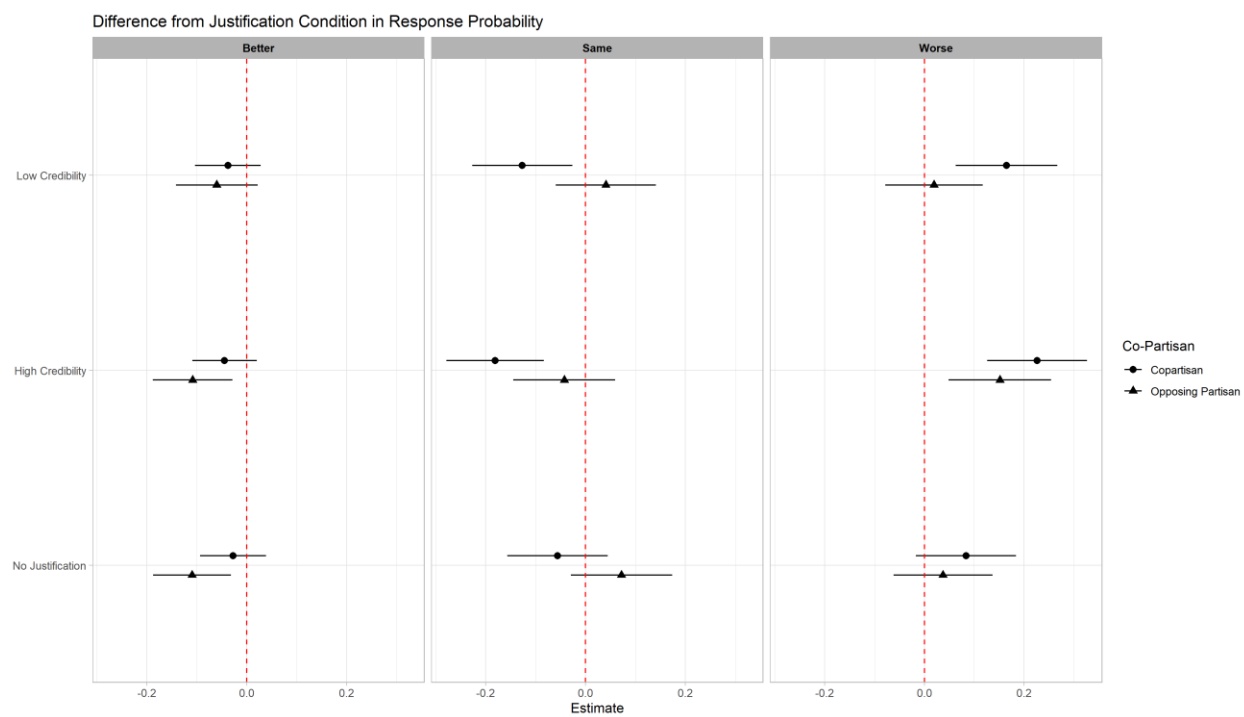
**Pre-Test Analyses**

Prior to conducting Experiment 2 we conducted pre-tests to examine the credibility of speakers and potential justifications. Our goal was to identify a 'consensual' low credibility source in the former case and a satisfactory justification in the latter.

*Credibility*

We recruited 302 subjects from MTurk. We provided these respondents with the Justification version of the legislator vignette from Study 1. We then provided them with one of the four following question stems (randomly either before or after evaluating the legislator in the vignette);

- A Democratic candidate running against Representative A in an upcoming primary was quoted in media reports saying that campaign donors of Representative A benefit financially from the amendment and that this is why the Representative voted for it.
- A Republican candidate running against Representative A in an upcoming primary was quoted in media reports saying that campaign donors of Representative A benefit financially from the amendment and that this is why the Representative voted for it.
- School administrators whose jobs are threatened by the cuts were quoted in media reports saying that campaign donors of Representative A benefit financially from the amendment and that this is why the Representative voted for it.

We then asked them to agree/disagree with whether the speaker would be well informed, have a personal interest in the cuts being passed, trustworthiness, and a personal interest in how Representative A is seen by voters. Table OC4 provides the descriptive statistics from this exercise separately for the Republican or Democratic version of Representative A and combined across legislator partisanship. The rival partisan tended to be seen as a bit less trustworthy, as having a stronger personal interest in what voters think, a greater personal interest in the cuts, and is slightly less informed. We thus chose this source as our low credibility source in Study 2.

**Table OC4:** Credibility Pre-Test Results

|  | Dem:Rival | Rep:Rival | Rival | Dem:Admin | Rep:Admin | Admin |
|---|---|---|---|---|---|---|
| Informed | 4.67 (1.30) | 4.15 (1.29) | 4.41 (1.42) | 4.45 (1.49) | 4.94 (1.39) | 4.70 (1.46) |
| Interest in Cuts | 3.27 (1.30) | 3.51 (1.33) | 3.39 (1.31) | 2.38 (1.55) | 2.21 (1.25) | 2.29 (1.40) |
| Trustworthy | 3.89 (1.24) | 3.57 (1.42) | 3.73 (1.34) | 3.91 (1.50) | 4.17 (1.33) | 4.04 (1.42) |
| Interest in Voters | 2.45 (1.54) | 2.61 (1.55) | 2.53 (1.54) | 2.97 (1.58) | 2.57 (1.32) | 2.77 (1.47) |

Note: Higher = *more* credible (i.e. more informed, more trustworthy, *less* of a personal interest). For the interest measures, lower scores = more agreeing with the items; hence, we should see lower scores on the interest in cuts for the administrators than the rivals, but lower scores on the voters in the reverse direction.

*Justification*

We fielded two pilot tests of Study 2 to identify a 'strong' justification (n=292 and 257 respectively). The pilot tests varied in two important respects. First, the policy positions of the legislator varied across the two versions. In the first pilot test the Democratic (Republican) legislator supported (opposed) raising taxes on those making over $250,000 a year; supported (opposed) instituting a carbon tax to regulate greenhouse gases; opposed (supported) raising the state minimum age to $15 an hour; and supported (opposed) decriminalizing possess of marijuana. The legislator in the second pre-test had the same positions as we used in the final study. Second, the 'long-term benefits' justification, which we would ultimately use in our study, was slightly modified between the pilots. In the first test, the legislator says: "This was a difficult decision, but prices will keep on rising without additional funding to improve the efficiency of the grid." We ultimately decided that discussions of grid efficiency might come off as an 'environmentalist' message and thus turn off Republicans. We thus kept the spirit of the justification the same

(i.e. planning for the long haul) but described the plan as an 'investment'. A third difference as well is that we also tested a 'contemporary' benefits styled justification: "This was a difficult decision, but the budget also increases spending for our school system and includes supplemental transportation funding." Figures OC5-OC6 below plot the mean post-test thermometer for the legislator (left panel) and its difference from the pre-test (right panel). In both pre-tests the long-term benefits treatment led to a significantly reduced decline in evaluations compared to the No Justification baseline. This, of course, deviates from what we observed in text where we see a weaker effect of the justification.
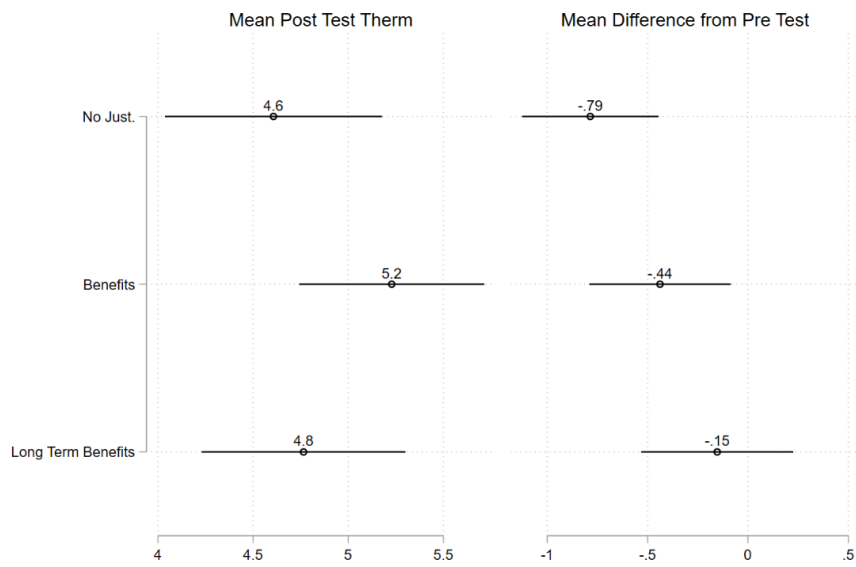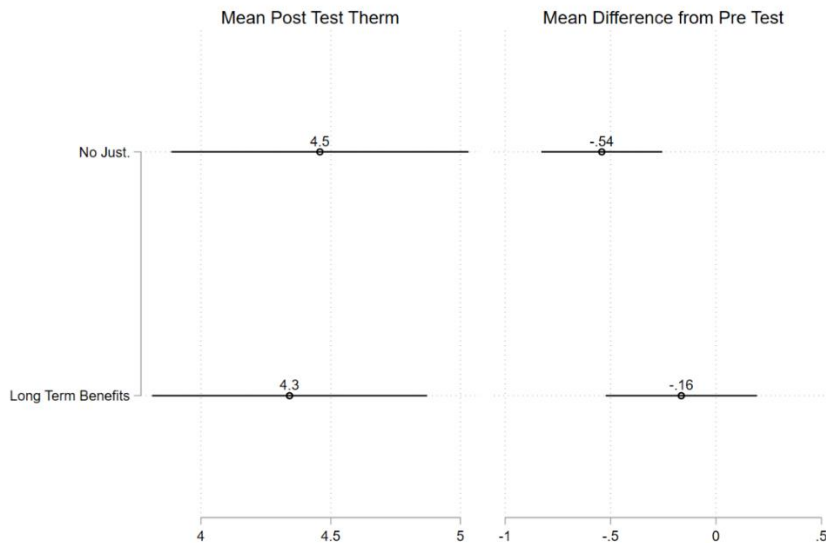
**Figure OC5:** Pre-Test 1



**Figure OC6:** Pre-Test 2

**Additional Measures**

Much as in Experiment 2 respondents also answered questions concerning policy approval and motive attributions. In the former case respondents were asked how strongly they approve or disapprove "of the decision to raise electricity rates that was described in the article" on a 1-7 scale (higher = approve). In the latter case, we again asked respondents to rate the importance of a "list of possible reasons" for the Representative's "vote on the electricity amendment." Much as in Experiment 1 we focus on two indices (M=0, SD=1): Good Representative Motives ($\alpha = 0.81$) and Political Motives ($\alpha=.79$).[5]

To begin with, we see that the Justification had a positive influence on policy approval and motive attributions, albeit more precisely on the former rather than the latter. The difference between the Justification and No Justification conditions are as follows: Policy = 0.37 [0.14, 0.61]; Good Representative Motives: 0.13 [-0.003, 0.27]; Political Motives: -0.10 [-0.25, 0.04]. Cohen's *d* for the three comparisons is 0.22, 0.14, and 0.10. This lack of a greater or more certain influence on motive attributions may be one reason why we did not see a stronger palliative effect of the justification in Study 2. On the other hand, the counter-narrative undermined evaluations on these items relative to the Justification Condition and particularly so when it came from a more credible source. The differences between the two credibility conditions and the Justification (No Counter) condition for the three items are: Policy: HC = -0.46 [-0.68, -0.23], LC = -0.37 [-0.59, -0.14]; Good Representative Motives: HC = -0.23 [-0.37, -0.09], LC = -0.18 [-0.32, -0.04]; Political Motives: HC = 0.16 [0.02, 0.30], LC = 0.19 [0.05, 0.33]. Thus, in all cases we see significantly worse impressions when the counter-narrative is present than when respondents just receive the justification.

**Table OC5:  Motives and Policy Agreement**

| | All Respondents | | | |
|---|---|---|---|---|
| | No Explanation | Explanation | Counter (HC) | Counter (LC) |
| Policy Agreement | 2.94 (2.78, 3.09) | 3.31 (3.14, 3.48) | 2.85 (2.70, 3.01) | 2.94 (2.79, 3.10) |
| "Political" Motives | -0.01 (-0.11, 0.09) | -0.11 (-0.22, -0.01) | 0.05 (-0.05, 0.14) | 0.08 (-0.02, 0.18) |
| "Good Representative" Motives | 0.0004 (-0.10, 0.10) | 0.05 (0.04, 0.23) | -0.09 (-0.19, 0.01) | -0.05 (-0.14, 0.05) |
| | | | | |
| | Co-Partisans | | | |
| Policy Agreement | 3.09 (2.87, 3.31) | 3.58 (3.23, 3.85) | 3.00 (2.76, 3.24) | 2.99 (2.75, 3.23) |
| "Political" Motives | -0.15 (-0.30, -0.001) | -0.20 (-0.35, -0.05) | -0.01 (-0.15, 0.13) | 0.001 (-0.15, 0.16) |
| "Good Representative" Motives | 0.12 (-0.03, 0.27) | 0.28 (0.15, 0.42) | -0.06 (-0.22, 0.09) | 0.05 (-0.10, 0.19) |
| | | | | |
| | Opposing Partisans | | | |
| Policy Agreement | 2.77 (2.52, 3.01) | 3.16 (2.91, 3.41) | 2.77 (2.53, 3.00) | 2,77 (2.66, 3.11) |
| "Political" Motives | 0.15 (0.01, 0.29) | -0.01 (-0.16, 0.15) | 0.10 (-0.04, 0.25) | 0.18 (0.04, 0.32) |
| "Good Representative" Motives | -0.12 (-0.28, 0.03) | 0.05 (-0.09, 0.18) | -0.12 (-0.27, 0.04) | -0.13 (-0.28, 0.02) |

**Notes:** Cells provide condition means alongside 95% confidence intervals.

---

[5] The former index is based on the reasons: "desire to make good public policy", "a desire to help his constituents", "a desire to help all state residents", and "personal values". The latter index draws on the reasons: "the influence of special interests", "political ambition", "loyalty to higher-ups in his political party", and "wining re-election".

**Online Appendix D: Experiment 3**

**Treatment Wordings**

*Gillibrand Treatment*[6]

[No Justification, Baseline:

**Kirsten Gillibrand used to have a moderate position on guns. Voters want to know why she's changed.**

During a Tuesday night CNN town hall, Sen. Kirsten Gillibrand (D-NY) tried to tackle her biggest campaign problem: name recognition.

Gillibrand answered questions on her positions on issues such as climate change and health care during the town hall. She also reiterated her support for victims of sexual misconduct and pledged to continue to fight for legislation to combat assault and harassment.

One issue raised during the town hall that is likely to follow Gillibrand on the campaign trail is her changed stance on gun control. During her tenure in the US House of Representatives from 2007-2009, then-Rep. Gillibrand fought vigorously in defense of gun rights. During her 2008 bid for re-election to the House, Gillibrand campaigned on her pro-gun rights record, touting her "A" rating from the National Rifle Association. However, after she was appointed to the US Senate in January 2009, Gillibrand shifted from her conservative record on gun rights and her NRA grade changed from an "A" to an "F" by September 2010.

One broader challenge for Gillibrand will be cutting through the noise of a crowded field of candidates in the 2020 Democratic primary.

[Justification (Appears before "one broader challenge" ending):

Gillibrand told the audience that her position changed when she started meeting with the families of gun violence victims. "When I was a member of Congress from upstate New York, I was really focused on the priorities of my district. However, what I recognized pretty quickly when I became a senator was that I didn't spend enough time thinking about other people around the state and other families who were really suffering," she said. "When you talk to a mom and a dad who lost their teenage daughter because she was at a party with friend and a stray bullet hit her and killed her, and you meet her whole class, not only do you immediately know that you were wrong, but you know you have to do something about it."

[Counter-Explanation (appeared *before* justification when it was present):

The timing of the shift has opened her up to criticism that she flip-flopped for political reasons. "If you looked up 'political opportunism in the dictionary, Kirsten Gillibrand's photo would be next to it" said [conservative/liberal] columnist Frank Rich in an op-ed last week. "Gillibrand always goes where the political wind blows," he continued.

*Corker Treatments*[7]

----

[6] The following news articles were consulted/used in the construction of this treatment:
- https://www.vox.com/2019/4/9/18303526/kirsten-gillibrand-cnn-town-hall-immigration-guns
- https://www.washingtonpost.com/politics/i-will-stand-up-for-what-i-believe-in-sen-kirsten-gillibrand-says--but-what-she-believes-quickly-changed-as-she-moved-from-house-to-senate/2019/01/19/1534b4ce-1b55-11e9-9ebf-c5fed1b7a081_story.html?noredirect=on&utm_term=.24ea1d6bfa7c
- http://www.msnbc.com/the-cycle/the-flip-flopping-nature-kirsten-gillibran
- https://www.washingtonexaminer.com/opinion/kirsten-gillibrand-shrugs-a-whole-career-of-flip-flopping

[7] The following news articles were used in constructing the treatment:
- https://newrepublic.com/minutes/146338/bob-corkers-flip-flop-tax-reform-sure-seems-fishy
- https://www.politico.com/story/2017/12/18/bob-corker-tax-bill-kickback-republicans-respond-302482

[No Justification Baseline:

**Sen. Bob Corker reverses course, will vote for Republican tax bill**

Sen. Bob Corker announced Friday that he will support the GOP tax bill, a reversal for the lone Republican to vote against the plan in the U.S. Senate.

Corker had long been critical of the proposal for not doing enough to address the national deficit. Corker was the only Republican to vote against the plan in the Senate after GOP leaders failed to satisfy his demands that the package must not increase the federal deficit.

The House and Senate each passed separate versions of the tax bill, and on Friday, GOP negotiators in the two chambers signed off on a final agreement.

A final vote is expected next week and Corker's announcement provides key support for its passage in the Senate.


[Justification (Appears before "the House and Senate each passed…"):

"After many conversations over the past several days with individuals from both sides of the aisle—including business owners, farmers, chambers of commerce and economic development leaders—I have decided to support the tax reform package. I believe this bill accompanied with the significant regulatory changes that are underway could have a significant positive impact on the well-being of Americans and help drive additional foreign direct investment in Tennessee" Corker said in a statement.

[Counter-Explanation (appears after the justification, when present):

[Conservative/Liberal] critics of Corker suggested a change to the bill may have affected his decision. Corker switched his vote after a provision was added that reduces taxes on real estate LLCs—and Corker, a real estate mogul, made $7 million in income from real estate LLCs last year.  "Looked at in the best possible light," Corker's change "illustrated the unseemly haste with which this tax bill was written" said Brett Stevens, a Senior Fellow at the [conservative/liberal] think tank The Tax Policy Center. "At worst, it looks like Corker got bought off."

---

- http://nymag.com/intelligencer/2017/12/the-gop-tax-bill-was-corrupt-before-the-corker-kickback.html
- https://indyweek.com/news/archives/corker-kickback-one-senator-s-going-make-mint-gop-s-tax-reform-bill/

**Table OD1:** Summary Statistics, Covariates

| | Support Stronger Restrictions | Oppose Stronger Restrictions | No Attitude |
|---|---|---|---|
| Gun Control Attitude | 67.93 | 21.85 | 10.22 |
| | Supports | Opposes | No Attitude |
| Tax Cut Attitude | 41.68 | 45.72 | 12.60 |
| | Co-Partisan | Opposing Partisan | Independent |
| Corker | 38.75 | 44.27 | 16.98 |
| Gillibrand | 44.27 | 38.75 | 16.98 |
| | Favorable | Unfavorable | No Opinion |
| Corker | 23.03 | 28.60 | 48.38 |
| Gillibrand | 25.95 | 32.75 | 41.29 |

Notes: For the Corker experiment, gain proximity = those that support the tax cuts, lose proximity = those that oppose. For the Gillibrand experiment, gain proximity = support stronger restrictions, lose proximity = oppose stronger restrictions.

**Additional Measures**

As with Experiments 1 and 2, we asked additional measures on the post-test. In particular, we asked the following item: "The following are a list of possible reasons for why [Senator Gillibrand changed her stance on gun control; Senator Corker voted for the tax reform bill.]. Please rate how important you think each is in explaining the Senator's behavior. The following motives were asked about: winning re-election, political ambition, the influence of special interests, personal gain, a desire to help constituents, and a desire to make good public policy. Factor analyses support a two factor solution in both cases relating to negative motives (the first four) and positive ones (the final two).[8] Table OD2 provides the mean scores for these two dimensions (factor variable, m = 0, sd = 1) for Corker while Table OD3 provides regression results that include the covariates used in text. Tables OD4 & 5 do the same for Gillibrand.

Corker's justification did not significantly impact people's motive attributions relative to the No Justification counterfactual. However, the counter-explanation did influence these perceptions. In particular, the LW counter-explanation led to an increased attribution of negative motives (albeit not significant at $p < 0.05$) and a reduced attribution of positive motives to Corker in the aggregate. The same pattern occurs with the right wing source but with more precisely measured effects on the negative motives index. Co-partisans and opposing partisans look to have taken different lessons from the counter-explanation to some extent. Co-partisan attributions of negative motives are significantly higher when the counter-explanation is present than when it is absent, but no such effect emerges for opposing partisans. On the other hand, while co-partisans attribute less positive intentions to Corker, these effects are noisy, whereas they are a bit more sharply estimated for opposing partisans at least in the case of the right-wing source.

---

[8] Corker: Two factors had eigen values over 1 (F1: 2.64, prop explained = 0.44, F2: 1.57, proportion explained: 0.26). The scale for the four negative motives has an alpha reliability score of 0.81 while the two negative items are correlated at 0.61 (alpha = 0.76). Gillibrand: F1 (EV = 2.76, proportion explained = 0.46); F2 (EV = 1.56, proportion explained = 0.26). The negative items have an alpha of 0.84, the positive one of 0.75 (correlation = 0.61) as well.

Gillibrand's justification, on the other hand, led to a reduced attribution of negative motives (albeit not significantly so) and a significant growth in the attribution of positive motives relative to the no justification counter-factual. Both of these effects are driven by co-partisans. This influence is smaller and noisier when the counter-explanation. These results again speak to the role that justifications play in shaping motive attributions and the role that counter-explanations play in undermining this persuasive impact.

**Table OD2:** Motive Attributions, Bob Corker

| | No Just | Just | LW | RW | ANOVA |
|---|---|---|---|---|---|
| | **All Respondents** | | | | |
| Positive | 0.10 [-0.01, 0.21] | 0.13 [0.02, 0.24] | -0.14 [-0.25, -0.02] | -0.10 [-0.21, 0.02] | F = 5.64, p < 0.001 |
| Negative | -0.10 [-0.22, 0.01] | -0.05 [-0.16, 0.06] | 0.06 [-0.05, 0.17] | 0.10 [-0.02, 0.21] | F = 2.60, p < 0.10 |
| | | | | | |
| | **Co-Partisans** | | | | |
| Positive | 0.29 [0.12, 0.46] | 0.45 [0.12, 0.46] | 0.06 [-0.11, 0.24] | 0.18 [-0.004, 0.36] | F = 3.71, p < 0.05 |
| Negative | -0.19 [-0.40, 0.02] | -0.17 [-0.34, -0.01] | 0.09 [-0.07, 0.25] | 0.14 [-0.03, 0.32] | F = 3.64, p < 0.05 |
| | | | | | |
| | **Opposing Partisans** | | | | |
| Positive | 0.02 [-0.15, 0.19] | -0.13 [-0.30, 0.04] | -0.21 [-0.40 [-0.03] | -0.24 [-0.42, -0.07] | F = 1.78, p = 0.15 |
| Negative | 0.08 [-0.07, 0.23] | 0.17 [0.0001, 0.34] | 0.18 [0.03, 0.33] | 0.18 [0.002, 0.35] | F = 0.33, p = 0.80 |

Notes: "Co-P" = co-partisan, "OpP" = Opposing Partisan" Standard errors in parentheses
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

**Table OD3:** Experiment 3 Results, Corker Motive Attributions

| | (1) Neg | (2) Pos | (3) Neg (Co-P) | (4) Pos (Co-P) | (5) Neg (OpP) | (6) Pos (OpP) |
|---|---|---|---|---|---|---|
| Justification | 0.0458 (0.0794) | 0.0187 (0.0778) | 0.0122 (0.130) | 0.168 (0.125) | 0.0350 (0.112) | -0.120 (0.120) |
| LW | 0.152 (0.0803) | -0.265*** (0.0786) | 0.279* (0.129) | -0.215 (0.124) | 0.0689 (0.115) | -0.242 (0.123) |
| RW | 0.222** (0.0798) | -0.230** (0.0782) | 0.350** (0.132) | -0.108 (0.127) | 0.0716 (0.113) | -0.291* (0.121) |
| Unfavorable | 0.155 (0.0833) | -0.527*** (0.0816) | 0.0537 (0.126) | -0.491*** (0.121) | 0.241 (0.126) | -0.499*** (0.134) |
| No Opinion | -0.0552 (0.0761) | -0.226** (0.0745) | 0.0239 (0.103) | -0.269** (0.0988) | -0.0919 (0.123) | -0.127 (0.131) |
| Lose Proximity | 0.155* (0.0728) | -0.318*** (0.0713) | 0.0715 (0.123) | -0.0694 (0.119) | 0.276** (0.106) | -0.508*** (0.113) |
| No Attitude | -0.365*** (0.0991) | -0.155 (0.0970) | -0.367* (0.165) | -0.118 (0.159) | -0.427* (0.176) | -0.364 (0.188) |
| Opposing Partisan | 0.0577 (0.0747) | -0.106 (0.0731) | | | | |
| Independent | -0.243** (0.0902) | -0.249** (0.0883) | | | | |
| Gillibrand First | 0.190*** (0.0564) | 0.0494 (0.0552) | 0.162 (0.0905) | -0.0634 (0.0871) | 0.135 (0.0811) | 0.0886 (0.0868) |
| Constant | -0.229** (0.0857) | 0.606*** (0.0839) | -0.270* (0.120) | 0.546*** (0.115) | -0.184 (0.144) | 0.624*** (0.154) |
| Observations | 1181 | 1181 | 456 | 456 | 521 | 521 |
| Adjusted $R^2$ | 0.070 | 0.104 | 0.026 | 0.049 | 0.070 | 0.090 |

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

**Table OD4:** Motive Attributions, Kirsten Gillibrand

| | No Just | Just | LW | RW | ANOVA |
|---|---|---|---|---|---|
| | **All Respondents** | | | | |
| Positive | -0.11 [-0.22, 0.01] | 0.08 [-0.04, 0.21] | -0.01 [-0.12, 0.10] | 0.04 [-0.08, 0.14] | F = 1.96, p = 0.12 |
| Negative | 0.06 [-0.05, 0.17] | -0.08 [-0.20, 0.04] | -0.05 [-0.16, 0.07] | 0.07 [-0.04, 0.18] | F = 1.70, p = 0.17 |
| | | | | | |
| | **Co-Partisans** | | | | |
| Positive | 0.11 [-0.03, 0.25] | 0.41 [0.25, 0.57] | 0.39 [0.24, 0.54] | 0.30 [0.17, 0.44] | F = 3.50, p < 0.05 |
| Negative | 0.08 [-0.05, 0.20] | -0.25 [-0.41, -0.08] | -0.12 [-0.28, 0.05] | -0.06 [-0.21, 0.08] | F = 3.25, p < 0.05 |
| | | | | | |
| | **Opposing Partisans** | | | | |
| Positive | -0.36 [-0.56, -0.16] | -0.17 [-0.37, 0.02] | -0.21 [-0.38, -0.04] | -0.25 [-0.46, -0.03] | F = 0.65, p = 0.58 |
| Negative | 0.07 [-0.15, 0.28] | 0.23 [0.05, 0.42] | 0.14 [-0.03, 0.32] | 0.31 [0.11, 0.52] | F = 1.17, p = 0.32 |

**Table OD5:** Experiment 3 Results, Gillibrand Motive Attributions

| | (1) Neg | (2) Pos | (3) Neg (Co-P) | (4) Pos (Co-P) | (5) Neg (OpP) | (6) Pos (OpP) |
|---|---|---|---|---|---|---|
| Explanation | -0.146 | 0.219** | -0.343** | 0.252* | 0.156 | 0.189 |
| | (0.0809) | (0.0752) | (0.106) | (0.102) | (0.137) | (0.128) |
| LW Counter | -0.0852 | 0.124 | -0.204 | 0.228* | 0.106 | 0.141 |
| | (0.0812) | (0.0754) | (0.108) | (0.104) | (0.138) | (0.129) |
| RW Counter | 0.00768 | 0.175* | -0.148 | 0.149 | 0.206 | 0.121 |
| | (0.0809) | (0.0752) | (0.106) | (0.102) | (0.141) | (0.132) |
| Unfavorable | 0.0182 | -0.531*** | -0.0180 | -0.455*** | -0.191 | -0.625*** |
| | (0.0840) | (0.0781) | (0.113) | (0.109) | (0.155) | (0.145) |
| No Opinion | -0.143 | -0.260*** | -0.113 | -0.295*** | -0.396* | -0.390* |
| | (0.0739) | (0.0687) | (0.0848) | (0.0813) | (0.163) | (0.153) |
| Lose Proximity | 0.241** | -0.583*** | 0.0139 | -0.449*** | 0.286** | -0.698*** |
| | (0.0756) | (0.0702) | (0.129) | (0.124) | (0.110) | (0.102) |
| No Attitude | -0.186 | -0.231* | -0.0148 | 0.0260 | -0.173 | -0.318* |
| | (0.100) | (0.0932) | (0.177) | (0.170) | (0.168) | (0.157) |
| Opposing Partisan | 0.203** | -0.204** | | | | |
| | (0.0705) | (0.0655) | | | | |
| Independent | -0.104 | -0.332*** | | | | |
| | (0.0862) | (0.0801) | | | | |
| Gillibrand First | -0.00184 | 0.0839 | -0.0172 | 0.0845 | -0.144 | 0.0750 |
| | (0.0571) | (0.0530) | (0.0772) | (0.0741) | (0.0978) | (0.0914) |
| Constant | 0.0116 | 0.393*** | 0.140 | 0.343*** | 0.291 | 0.353* |
| | (0.0810) | (0.0752) | (0.0947) | (0.0909) | (0.173) | (0.161) |
| Observations | 1182 | 1182 | 520 | 520 | 462 | 462 |
| Adjusted $R^2$ | 0.044 | 0.176 | 0.008 | 0.083 | 0.033 | 0.153 |

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

**Online Appendix E**

In text we focus on experiments where participants are randomly assigned to one of the following three basic counterfactuals: a politician does something without explanation; they provide an explanation for their behavior; they provide this explanation while another actor provides a rival account to make sense of the politician's behavior. As we discuss in the Conclusion of the manuscript, an alternative counterfactual of interest is also present: what happens when the counter-explanation is actually the *only* account that people receive? Perhaps it is the case that politicians are not better off when a counter-explanation is present than when it is absent, but they are better off when they provide an explanation in a competitive environment compared to one where they do not explain themselves while some other actors does.

We fielded a small experiment that sheds some light on these considerations. Specifically, we recruited 319 participants from Mechanical Turk and asked them to consider a situation wherein their state representative voted in favor of a recently passed budget amendment that would "would change the criteria used to distribute funds to fire departments in local communities in the state," such that there would be "cuts in funding for some communities, including your own, but increases for others." [See below for full treatment wordings.] One important deviation from Experiment 1 is present: we did not provide information concerning the partisanship of the politician.

We randomly assigned participants to read one of eight such vignettes after which they evaluated the legislator on a 0-10 feeling thermometer measure. Respondents assigned to the baseline condition received information only concerning the legislator's vote for the amendment. Respondents in three of the remaining conditions also received a justification from the legislator for their vote. One focused on fairness motives, i.e. to "make the distribution of funds across the state fairer" (similar to the one used in Experiment 1). In the second justification, the legislator argues that his/her constituents will ultimately benefit because the budget "decreases property taxes and increases spending for the school system in my district". The third type of justification was a 'tailored' account along the lines used by Grose et al. (2015). Here, the legislator highlighted past positive actions on this issue ("last term I sponsored bills to increase funding to improve fire safety and to combat arson") and grounds their vote in the need for a legislative compromise to pass the budget. These three conditions replicate prior work and we expect them elicit to more positive evaluations relative to the baseline condition.

The final four conditions include a counter-narrative for the elite's behavior. Here, respondents read that "non-partisan budget experts quoted in media reports about the cut say that campaign donors" of the representative "benefit financially from the amendment and that this is why the Representative voted for it" (again, similar to Experiment 1). In one of these conditions the counter-explanation was provided by itself, i.e. absent one of the justifications. In the remaining conditions, meanwhile, we paired the counter-narrative with one of the above justifications.

Figure OE1 provides the mean ratings of the representative per condition alongside 95% and 83.5% confidence intervals; the latter are more appropriate for visually approximating whether the difference between two condition means is statistically significant (Bolsen and Thornton 2014; Goldstein and Healy 1995). Four results stand out in Figure 1. First, evaluations of the legislator in the no justification baseline condition are very low. A legislator taking the actions described in the vignette is likely to elicit quite negative reactions from constituents sans explanation. Second, evaluations are substantially more positive in all three conditions where the legislator justifies their actions. This is consistent with Hypothesis 1. Third the counter-explanation undermined the effectiveness of these otherwise effective justifications in all cases as we argued would occur in Hypothesis 2.[9] Finally, the counter-explanation, when presented by itself, did not undermine evaluations relative to the Baseline condition. If anything, then, the providing the justification did not lead the politician to *recover* from a negative situation that would otherwise occur. We hesitate to place too much weight on these analyses, however, given the small cell sizes in each condition (e.g., ~40 per condition). In addition, evaluations of the politician were already quite low in the Baseline which may constrain the ability of the counter-explanation to lead to *even worse* evaluations.

---

[9] The difference between the Explanation (Counter) and Explanation (No Counter) conditions is statistically significant in all three cases: Fairness (difference = -3.48 [-4.52, -2.43]); Benefits (difference = -2.26 [-3.31, -1.21]); Tailored (difference = -1.38 [-2.42, -0.34]).

**Figure OE1:** Evaluations of Representative in Study 1
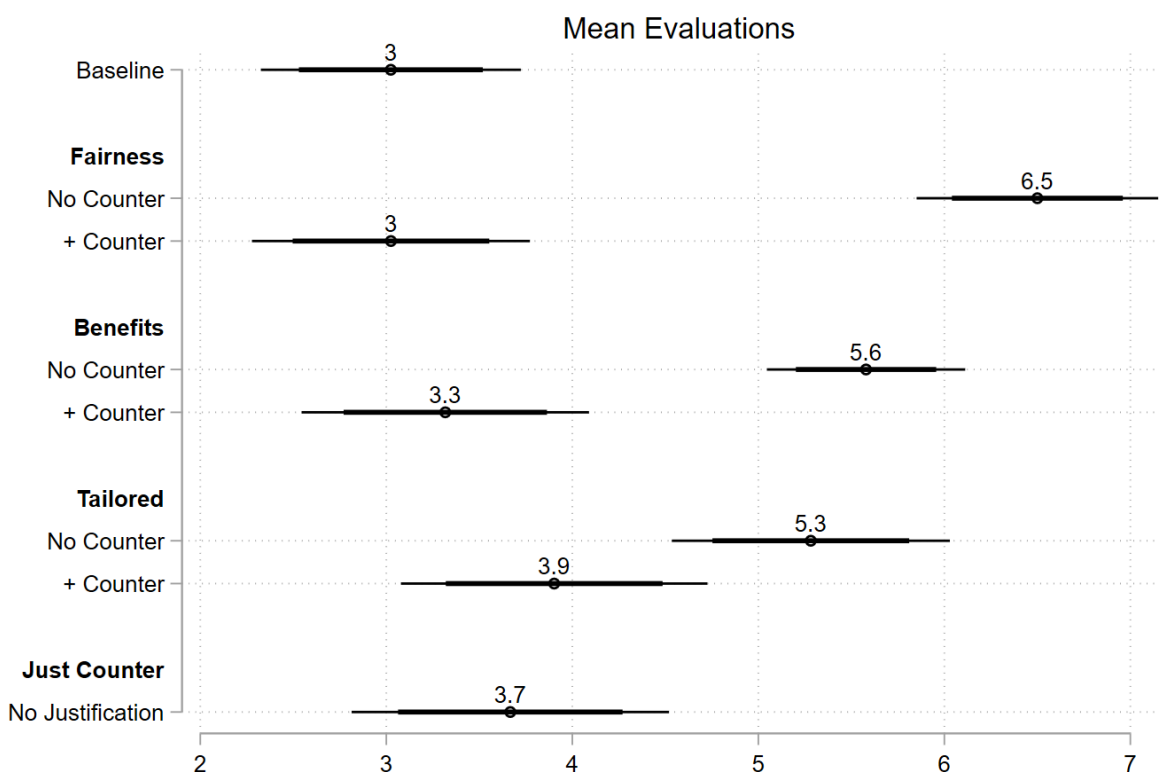


**Notes**: Markers provide the mean thermometer rating for representative by treatment condition along with 95% and 83.5% confidence intervals.

## Wordings of Treatments and Outcome Variable

**Introduction to Vignette (common to all):**

On the next few pages we will ask you to imagine a scenario where your state legislature has voted on an amendment to the annual budget plan. Please read this account as if it was provided by your own state representative in this situation.

**No Explanation Vignette:**

Imagine that your state's legislature recently passed an amendment to the annual budget plan which would change the criteria used to distribute funds to fire departments in local communities in the state. This amendment would result in cuts in funding for some communities, including your own, but increases for others.  Your state representative voted in favor of this amendment.

**Fairness Vignette:**

Imagine that your state's legislature recently passed an amendment to the annual budget plan which would change the criteria used to distribute funds to fire departments in local communities in the state. This amendment would result in cuts in funding for some communities, including your own, but increases for others.  Your state representative voted in favor of this amendment.

When asked about the vote, your representative said: "I voted for the amendment because my community generates enough funding for our local firefighters already and this amendment will make the distribution of funds across the state fairer for all communities with more money going to those who need the funding the most."

**Fairness + Counter Vignette:**

Imagine that your state's legislature recently passed an amendment to the annual budget plan which would change the criteria used to distribute funds to fire departments in local communities in the state. This amendment would result in cuts in funding for some communities, including your own, but increases for others. Your state representative voted in favor of this amendment.

When asked about the vote, your representative said: "I voted for the amendment because my community generates enough funding for our local firefighters already and this amendment will make the distribution of funds across the state fairer for all communities with more money going to those who need the funding the most." However, non-partisan budget experts quoted in media reports about the cuts say that campaign donors of Representative A benefit financially from the amendment and that this is really why the Representative voted for it.

**Benefits Vignette:**

Imagine that your state's legislature recently passed an amendment to the annual budget plan which would change the criteria used to distribute funds to fire departments in local communities in the state. This amendment would result in cuts in funding for some communities, including your own, but increases for others. Your state representative voted in favor of this amendment.

When asked about the vote, your representative said: "A compromise was required to pass this budget. In the end I believe my constituents will benefit from this budget because, while it unfortunately cuts funding to the fire department, it also decreases property taxes and increases spending for the school system in my district."

**Benefits + Counter Vignette:**

Imagine that your state's legislature recently passed an amendment to the annual budget plan which would change the criteria used to distribute funds to fire departments in local communities in the state. This amendment would result in cuts in funding for some communities, including your own, but increases for others. Your state representative voted in favor of this amendment.

When asked about the vote, your representative said: "A compromise was required to pass this budget. In the end I believe my constituents will benefit from this budget because, while it unfortunately cuts funding to the fire department, it also decreases property taxes and increases spending for the school system in my district." However, non-partisan budget experts quoted in media reports about the cuts say that campaign donors of Representative A benefit financially from the amendment and that this is really why the Representative voted for it.

**Tailored Vignette:**

Imagine that your state's legislature recently passed an amendment to the annual budget plan which would change the criteria used to distribute funds to fire departments in local communities in the state. This amendment would result in cuts in funding for some communities, including your own, but increases for others. Your state representative voted in favor of this amendment.

When asked about the vote, your representative said: "I have been a strong supporter of firefighters throughout my

career. Last term I sponsored bills to increase funding to improve fire safety and to combat arson. However, while this amendment has its flaws, we had to pass a budget and a compromise was required."

**Tailored + Counter Vignette:**

Imagine that your state's legislature recently passed an amendment to the annual budget plan which would change the criteria used to distribute funds to fire departments in local communities in the state. This amendment would result in cuts in funding for some communities, including your own, but increases for others. Your state representative voted in favor of this amendment.

When asked about the vote, your representative said: "I have been a strong supporter of firefighters throughout my career. Last term I sponsored bills to increase funding to improve fire safety and to combat arson. However, while this amendment has its flaws, we had to pass a budget and a compromise was required." However, non-partisan budget experts quoted in media reports about the cuts say that campaign donors of Representative A benefit financially from the amendment and that this is really why the Representative voted for it.

**Just Counter Vignette:**

Imagine that your state's legislature recently passed an amendment to the annual budget plan which would change the criteria used to distribute funds to fire departments in local communities in the state. This amendment would result in cuts in funding for some communities, including your own, but increases for others. Your state representative voted in favor of this amendment. Non-partisan budget experts quoted in media reports about the cuts say that campaign donors of Representative A benefit financially from the amendment and that this is why the Representative voted for it.

**Table OE1.** Sample Characteristics – Demographic Variables

| | Sample | ACS | ANES 2016 (Weighted) | ANES 2016 (Unweighted) |
|---|---|---|---|---|
| % Female | 39.81 | 51.3 | 51.96 | 52.90 |
| **Age** | | | | |
| Average | 36.18 | | 47.31 | 49.58 |
| % 18-29 | 31.35 | 21.5 | 15.71 | 15.71 |
| % 30-44 | 51.10 | 25.0 | 25.11 | 25.11 |
| % 45-64 | 15.05 | 33.8 | 36.27 | 36.27 |
| % 65+ | 2.51 | 19.7 | 22.92 | 22.92 |
| | | | | |
| **Education** | | | | |
| % < HS | 0.63 | 12.6 | 9.14 | 6.67 |
| % HS | 10.66 | 27.7 | 28.88 | 19.16 |
| % Some College | 31.66 | 31.0 | 30.97 | 35.49 |
| % Bachelor Degree | 43.26 | 18.3 | 18.33 | 22.59 |
| % Post-Bach. Degree | 13.79 | 10.5 | 12.69 | 16.09 |
| | | | | |
| **Household Income** | | | | |
| Median | $50,000 to $59,999 | 57,617 | | $55,000-$59,999 |
| <10,000 | 3.45 | 6.7 | 9.45 | 9.14 |
| 10,000-39,999 | 29.78 | | 26.69 | 27.16 |
| 40,000-69,999 | 31.66 | | 22.11 | 22.93 |
| 70,000-99,999 | 19.44 | | 16.41 | 16.49 |
| 100,000-149,999 | 10.66 | 14.0 | 13.49 | 13.25 |
| 150,000+ | 5.02 | 12.1 | 11.85 | 11.03 |
| | | | | |
| **Race** | | [see note] | | |
| White | 79.31 | 72.6 | | |
| African American | 6.27 | 12.7 | | |
| Asian | 12.23 | 5.4 | | |
| Other | 2.19 | 9.3 | | |
| | | | | |
| Latino/Hispanic (% Yes) | 8.15 | 17.8 [see note] | | |
| | | | | |
| White (~Hispanic) | 73.35 | | 69.17 | 71.68 |
| Black (~Hispanic) | 4.70 | | 10.92 | 9.39 |
| Asian | 12.23 | | 3.12 | 3.49 |
| Hispanic | 8.15 | | 11.89 | 10.62 |
| Other | 1.57 | | 4.9 | 4.81 |
| | | | | |
| **Party Identification** | | | | |
| Mean PID | 3.50 (2.13) | N/A | 3.82 (SE: 0.04) | 3.86 (2.15) |
| | | | | |
| % Dem | 56.78 | N/A | 45.98 | 45.67 |
| % Ind. | 34.07 | N/A | 14.67 | 13.63 |
| % Rep. | 9.15 | N/A | 39.35 | 40.70 |

**Notes**: ACS Estimates stem from the 2016 ACS 1-year estimate files. Gender and Age statistics were constructed from Table B01001. Education is constructed from Table S1501. Income is constructed from Table S1901; the

income categories for the ACS do not neatly overlap as they begin from 10-14,999 and then move in ten thousand increments from there while our data begins at 10,000-19,999 and then proceeds via ten thousand increments until 100,000. The ACS estimates that do appear represent those categories that overlap with our own. Race & Latino/Hispanic estimates are from Tables B02001 and B03003. Note however that these tables focus on the *total* population including those < 18 years old. As younger Americans are more diverse than older, this yields estimates of higher diversity and particular for Latinos (see comparison between the ACS and ANES).

# The Effectiveness of Elite Explanations in a Competitive Environment (#6530)

**Created:** 11/02/2017 08:33 AM (PT)
**Shared:** 01/07/2019 12:47 AM (PT)

This pre-registration is not yet public. This anonymized copy (without author names) was created by the author(s) to use during peer-review. A non-anonymized version (containing author names) will become publicly available only if an author makes it public. Until that happens the contents of this pre-registration are confidential.

**1) Have any data been collected for this study already?**
No, no data have been collected for this study yet.

**2) What's the main question being asked or hypothesis being tested in this study?**
Previous work suggests that political elites that take unpopular actions (i.e. voting for unpopular bills) can evade punishment for their actions by offering an explanation for why they acted the way they did. However, this work has not explored contexts wherein other actors interject to argue that the elite's actions were actually motivated by negatively valenced motives. We seek to understand whether these strategic counter-explanations undermine the effectiveness of the original explanation.

To answer this question, we will field an experiment with a 6 (Baseline Information [B], Vote for Cuts [V], Vote+Explanation [E], Vote+Explanation+Counter-Explanation1 [CH1], + Counter2 [CH2], + Counter3 [CL1]) x 2 (legislator partisanship: [Democrat, Republican]) design. The three counters have the same content but vary in speaker credibility (two high, one low [e.g. a partisan source]).

We have the following expectations regarding legislator evaluations: (1) B > V (the vote costs the legislator support); (2) E > V [the explanation mitigates evaluative costs]); (3) E > CL1 > (CH1 ≈ CH2) [the counter-explanation undermines the explanation's effects, particularly when given by a highly credible source]. We do not express a strong prior regarding whether the explanation will fully mitigate costs (thus, we expect: B ≥ E > V ) or whether the counter will fully undermine the effects of the explanation (thus: [CH1 ≈ CH2 ] ≥ V).

We also postulate that respondent partisanship will matter in three ways, focusing on co-partisans to the legislator (CP) and opposing partisans (OP): (1) The effect of the elite's explanation for their own behavior will have a more positive effect among CP than OP; (2) the effect of the credible counter-explanation will not differ significantly across partisanship alignment (i.e. effect(CP) ~ effect(OP)); (3) the low credibility counter-explanation will undermine evaluations (relative to the explanation only condition) most strongly among those that share a partisan allegiance with the counter-explanation giver as individuals that share a similar partisan identity as the legislator will discount this strategic charge.

**3) Describe the key dependent variable(s) specifying how they will be measured.**
Our core dependent variable is a general evaluation of the legislator scored on a 0-10 scale where higher = more positive affect.

We will also investigate the following variables. Our theory holds that the elite explanation mitigates the evaluative costs of taking an unpopular position via two potential routes: (1) by persuading some people that the act was the right one and thereby leading to less blame & more credit given for the action; and (2) by leading individuals to believe that the elite was positively (rather than negatively) motivated and thereby undermine blame attributions (albeit without the offering of credit for the behavior). In turn, the counter-explanations should upset these pathways. We will thus measure whether the respondent agrees or disagrees with the elite's action, the level of credit/blame accorded the elite for the action, and perceptions of the elite's motives for taking the action.

**4) How many and which conditions will participants be assigned to?**
12 conditions as noted above.

**5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.**
Our key test for the first set of predictions above is a consideration of difference in condition means across the six treatment conditions; here we will use ANOVA/t-tests and OLS regressions. The second set of expectations are conditioned on respondent partisanship; we will thus focus on OLS regressions where co-partisanship is included as a moderator variable and compare the effect of the relevant treatment both within and across partisan co-identification.

**6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.**
None currently planned

**7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.**
1800 respondents.

**8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses**

**planned?)**

We expect to run three secondary analyses. First, prior work shows that account satisfaction matters; individuals that receive an explanation but find it unsatisfactory, i.e. reject it, should not better evaluate the elite (e.g. McGraw 1990, 1991). We will thus investigate whether subjective satisfaction moderates the reception of the accounts by including this factor as a moderating variable in the analyses described above.

Second, we will ask two manipulation checks wherein the respondents are asked to indicate the party of the legislator and how they voted regarding the amendment. We will generate a robustness check wherein we focus on cases wherein the respondents are correct on these items.

Finally, we will consider whether the elite's explanation is more effective when offered by a Democrat than a Republican. Voting for a cut to education funding is a potentially more surprising or counter-stereotypical action for a Democratic than Republican legislator; prior work suggests that actions such as these may grant the Democratic legislator enhanced credibility as their actions may be inferred as conflicting with their general interests (Alt, Lassen, and Marshall 2015; Berinsky 2017).

# Explanations, Counter-Narratives, and Blame Management (#11852)

**Created:** 06/14/2018 02:37 AM (PT)
**Shared:** 01/07/2019 12:46 AM (PT)

This pre-registration is not yet public. This anonymized copy (without author names) was created by the author(s) to use during peer-review. A non-anonymized version (containing author names) will become publicly available only if an author makes it public. Until that happens the contents of this pre-registration are confidential.

**1) Have any data been collected for this study already?**
No, no data have been collected for this study yet.

**2) What's the main question being asked or hypothesis being tested in this study?**
Our study concerns when elected officials can escape blame for taking controversial actions by justifying the action and whether counter-narratives provided by alternative actors focusing on the ulterior motives of the elite can attenuate this explanation effect.

Our first hypothesis (H1) is that they can, i.e. that evaluations of the elite, the action, and their motives, will be more positive when a justification is offered than when it it absent. Our second hypothesis (H2) is that this positive effect will be attenuated when a counter-narrative suggesting ulterior motives were in play is also present. Our third hypothesis (H3) is that this attenuation will be greater when the counter-narrative is presented by a highly credible source. Our final hypothesis (H4) is that credibility will matter more for co-partisans or those with positive prior attitudes toward the legislator than opposing partisans who will be likely to take negative information about the actor from high and low credible sources.

**3) Describe the key dependent variable(s) specifying how they will be measured.**
We will examine the following dependent variables: (1) the change in evaluations of the legislator on the post-test from a pre-test measure; we will examine the 0-10 scaled post-test measure controlling for the 0-10 scaled pre-test measure, a difference score (i.e. post-pre), and a simplified version of the difference measured (-1 = negative change, 0 = same, +1 = positive change); (2) approval of the policy action on a 1-7 scale from strongly disapprove to strongly approve; (3) positive and negative motive attributions and the difference between them (i.e. positive motives - negative; both measures are formed from a battery of items and reduced via factor analysis); and (4) how much credit or blame the actor receives for taking the action in question.

**4) How many and which conditions will participants be assigned to?**
8 in total, from a 2 x 4 factorial.

Respondents will first take part in an impression formation task where they will be given information about a state legislator. The legislator will be randomly assigned to be either a Democrat or Republican. Later, after a series of buffer items, respondents will be assigned to one of four conditions: (1) No Explanation (legislator does not offer an explanation for their actions); (2) Explanation; (3) Explanation + Counter-Narrative (High Credibility Source); or (4) Explanation + Counter-Narrative (Low Credibility Source).

**5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.**
We will regress our dependent variables on the four treatment condition assignment (base = Explanation Provided). We will also interact this indicator with a measure of co-partisanship to the legislator and with pre-test evaluations to examine moderation by prior evaluations and partisan allegiance.

We will also use t-test and Wald tests to compare the difference in evaluations between High and Low Credibility (both directly against each other and their difference from the Explanation condition, e.g. [(Explanation - High Credibility) - (Explanation - Low Credibility)].Likewise, we will investigate the difference between our outcome variables in the Counter-Narrative conditions to the No Explanation condition using t-tests.

Finally, we will use multinomial logit analyses to estimate the simplified change indicator (i.e. negative, no, and positive change) described above.

**6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.**
The elected official in the experiment is described as a state legislator from Vermont. We will thus replicate our models without any Vermont residents to see if they are driving any results.

**7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.**
1600 respondents, yielding approximately 400 per our four main conditions or 200 per the full 2 x 4 factorial.

**8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)**

# Explanations and Counter-Explanations with Real Politicians (#22342)

**Created:** 04/18/2019 12:39 AM (PT)
**Shared:** 06/17/2019 11:58 PM (PT)

---

This pre-registration is not yet public. This anonymized copy (without author names) was created by the author(s) to use during peer-review. A non-anonymized version (containing author names) will become publicly available only if an author makes it public. Until that happens the contents of this pre-registration are confidential.

---

**1) Have any data been collected for this study already?**

No, no data have been collected for this study yet.

**2) What's the main question being asked or hypothesis being tested in this study?**

Do explanations offered by politicians mitigate blame for controversial actions when competing voices insinuate that the politician has ulterior motives for the action? We expect that explanations are effective in a non-competitive environment but that their influence will be mitigated in this type of competitive environment. Moreover, we expect that this mitigation will be stronger when a more credible source attacks the focal politician and that credibility will matter more for co-partisans to the focal politician.

**3) Describe the key dependent variable(s) specifying how they will be measured.**

Our core DV is a general evaluation of the legislator on a 0-10 scale where higher = more positive. We also assess the perceived motives of the legislator for the actions described in the experimental vignette on a 1-5 scale (higher = more important). We will analyze these both separately and as two separate sub-scales (dependent on scale reliability).

**4) How many and which conditions will participants be assigned to?**

Respondents will read about two politicians (Kirstin Gillibrand and Bob Corker) with the order randomly varied. Within each candidate experiment, they will be assigned to one of four conditions: (1) No Explanation (the legislator changes policy positions without explaining why); (2) Explanation (they explain why); (3) Explanation + Left-Wing Counter (they explain why and a left-wing source questions the explanation); (4) Explanation + Right Wing Counter (the counter information is from a right-wing source). We expect that a LW source should be deemed more credible by Democrats and a RW source as more credible among Republicans, all else equal.

**5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.**

We will compare respondents across experimental treatments using ANOVA/t-tests and OLS regressions. We will first do so with a truncated version of treatment assignment (No Explanation vs Explanation vs Explanation w/Counter) before disaggregating the two counter-explanation conditions. We will perform two sets of regressions: one without covariates and one with (covariates will include: pre-test attitude toward the candidate; whether the respondent has gained/lost issue proximity from the politician's behavior; partisanship/co-partisanship; and the order of the politician experiment). To test whether credibility matters we will use Wald tests comparing the coefficients for the two counter-information treatments and also t-tests. To examine heterogenous treatment effects we will interact treatment assignment with respondent partisanship. For analyses of motives we will pay particular attention to two: personal ambition and personal gain. The counter-explanation in the Gillibrand experiment calls attention to ambition while the one in the Corker treatment focuses on personal gain.

**6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.**

We do not expect any.

**7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.**

We have contracted for 1200 respondents (~400 per treatment condition).

**8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)**

We will also analyze interaction models wherein experimental treatment is interacted with the respondent's pre-test attitude to the politician and whether they gained/lost proximity due to the politician's behavior. Generally, we would expect similar patterns as with co-partisanship, e.g. those with favorable pre-test attitudes/gain proximity should be more wiling to accept the explanation and more willing to counter-argue contrary information with the result being that credibility should matter more for these respondents as well.

Wharton UNIVERSITY of PENNSYLVANIA CREDIBILITY LAB