ARTICLE

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

Joshua Robison

Department of Political Science, Universiteit Leiden, The Hague, The Netherlands
E-mail: 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. This article addresses a critical limitation in the literature on elite explanation giving and accountability: the absence of attention to conflicting information regarding the official's behavior. The study shows across three pre-registered survey experiments that explanations are ineffective when other speakers offer counter-explanations that focus on the official's potential ulterior motives. It further demonstrates that this occurs even when the counter-explanation comes from a partisan source with low credibility. These results imply that elected officials enjoy less leeway for their actions than existing work allows, and highlight important tensions concerning the relationship between elite behavior and accountability processes.

**Keywords:** accountability; elite explanations; experiments; source credibility; blame

Elected officials generally wish to remain in office (Mayhew 1974). One way they can do so is by taking popular policy actions (for example, Ansolabehere and Jones 2010). However, they 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), for example by *explaining* why they acted (Fenno 1978; Kingdon 1989; McGraw 2001). Prior studies show 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 who wish to maximize their chances of re-election while acting as they see fit, it suggests a tension with accounts of democratic legitimacy that are grounded in officials' responsiveness to public preferences (Dahl 1971).

Previous work on explanation giving has compared the important counterfactual of receiving an explanation vs. not receiving an explanation (for example, Grose, Malhotra and Van Houweling 2015; McGraw, Best and Timpone 1995). However, explanation giving frequently occurs in contexts in which 'an "instigator" (such as 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: they also introduce *counter-explanations* about why the official acted, which often highlight ulterior and self-serving goals that may have motivated the politician. For 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. Yet Burr's opponent 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 in a quid pro quo. We cannot apply findings from previous studies to predict how Burr's constituents will respond in this type of competitive environment. This is a crucial gap in the literature, 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.

I integrate theories on framing and the psychology of blame and suspicion to argue that elite explanations will be less effective in communication environments in which rival accounts focus attention on the politician's potential ulterior motives. 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 where they are considered to be 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 own explanation and a rival account suggesting the presence of ulterior motivations, I expect that many will find the latter more persuasive and thereby reject the focal politician's explanation.

I find support for this argument across three pre-registered survey experiments. While the politicians in these experiments were often (if not always) evaluated more positively when they explained their controversial actions, this effect disappeared when a counter-explanation was offered. I consider two scope conditions for this effect. First, I 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 focal politician's explanation is *also* undermined when the counter-explanation is offered by a *low*-credibility source. Secondly, I consider the role of partisanship – and particularly the behavior of co-partisans – to the focal politician. Here, I 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.

This study contributes to the broader attempt to understand public reactions to competing messages from elites by focusing on a type of message that is 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 elites' ability 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 politicians' pre- *and* post-decision-making behavior (McGraw 2001). I focus 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). My 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 that coincides with an attempt to recast it in a more positive light by highlighting the benefits that may follow from the action. Prior work shows that justifications are powerful blame management tools provided that the justification is considered 'satisfactory' – that is, audience members accept it (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 'positive' intentions motivated the action. In the political domain, this means believing that officials are attempting to help constituents and make good policy rather than trying to help themselves 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".'

**Hypothesis 1:** 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 reduce the evaluative costs of their controversial actions (but see Peterson and Simonovits 2017 for a notable exception). However, this work shares a common limitation: 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, Best and Timpone 1995; McGraw, Timpone and Bruck 1993). 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. They also provide conflicting interpretations of this behavior. This conflict can occur along two dimensions, as implied by the above discussion on blame. First, instigators may provide policy arguments to dispute the focal politician's claims that their actions will generate positive consequences. Previous 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 politicians. It is unclear whether a conflict over the merits of an action would mitigate the blame-avoidant properties of justifications. Justifications may also influence beliefs about politicians' motives, which is important because beliefs about agents' motives appear to be central to impression formation processes (McGraw 2003).

I focus on a second dimension of interpretative conflict, which has not been investigated in the broader literature on competition and opinion formation. Justifications provide an account of why the politician acted – that is, 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). For instance, instigators may provide counter-explanations that highlight ulterior and self-serving reasons that may be 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 the result of learning new information about the consequences of gun violence, her political opponents highlighted her desire to win support from party elites to advance her career (Nilsen 2019). Prior work fails to assess how Gillibrand's constituents, or potential new voters, will respond when faced with these conflicting messages.

I expect that counter-explanations focused on the ulterior motives of the politician will lead to worse evaluations compared to a justification only counter-factual. My argument integrates two separate literatures.[1] First, individuals tend to shift to a more cognitively active form of information processing if they have reason to suspect that ulterior goals motivate an agent's actions (Fein, Hilton and Miller 1990; Mayo 2015; Priester and Petty 1995). In so doing, suspicious perceivers consider the relative plausibility of rival interpretations of the agent's behavior to avoid being misled.

Suspicion entails considering the possibility that ulterior motives guided the politician's behavior. I further argue that many will consider this type of counter-explanation to be persuasive and hence find the focal politician's explanation unsatisfactory for two reasons. 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). Secondly, 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 2020; Iyengar, Norpoth and Hahn 2004). Likewise, beliefs about decision makers' motives influence evaluations of both the decision maker and the decision itself (Bøggild 2016; Hibbing and Alford 2004). Information about the intentions of political candidates, individuals and social groups has the greatest influence on impression formation, reflecting the importance of these considerations in making credit and blame attributions (Bittner 2011; Cuddy, Fiske and Glick 2008; Laustsen and Bor 2017). Thus counter-explanations focused on a politician's ulterior ends 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 of a politician's behavior than their own explanation, which would blunt the latter's impact.

**Hypothesis 2:** Counter-explanations will lead to worse evaluations compared to when only the justification is present, all else equal.

## Credibility, partisanship and Scope Conditions

I consider two scope conditions for the argument outlined above. First, sources with different degrees of perceived credibility may offer counter-explanations. Political rivals, for instance, may be perceived as having a political motivation to impugn the focal politician and thus be deemed less credible, whereas non-partisan actors may be perceived as more trustworthy.[2] I expect counter-explanations to be more effective if they originate from more credible sources (Chong and Druckman 2007a; Weitz-Shapiro and Winters 2016). However, I do not make

---

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

[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. 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).

any specific predictions about whether counter-explanations are *only* effective when offered by a highly credible source.

Secondly, 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 have a partisan motivation to discount the counter-explanation (Lodge and Taber 2006), this does not mean they are immune to counter-argumentation as strong arguments can overcome partisan interpretations in some contexts (Boudreau and MacKenzie 2014; Bullock 2011). I argue that co-partisans should be particularly attuned to source credibility when evaluating counter-explanations. Partisans from a party that is in competition with that of the focal politician may be motivated to accept derogatory information regardless of the source of the message. Co-partisans, however, may be more resistant and only buckle when the incoming information is highly credible and thus more difficult to counter-argue.

**Hypothesis 3:** Counter-explanations offered by credible sources will harm evaluations more than those offered by less credible sources, all else equal.

**Hypothesis 4:** Credibility will matter more for co-partisans to a politician than for opposing partisans, all else equal.

## Study Design

I test my hypotheses in three pre-registered survey experiments. This allows me to replicate and build on the initial results with new samples and stimuli, thereby increasing the external validity of the argument (McDermott 2011). The experiments move from fictional politicians acting on non-salient issues to real politicians acting on salient issues in which the (counter)-explanations offered reflect actual political rhetoric. Table 1 provides an overview of some of the key elements of these studies. This section describes all three experiments and the logic behind the design choices.

### Choice of Sample

Experiments 1 and 2 were conducted using samples recruited from Amazon's Mechanical Turk. Experiment 3 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 via Lucid utilized quotas on education, race and age. Both sources yield samples that better approximate the demographic profile of the US adult population than other convenience sampling methods (see Table A1 for an overview of sample characteristics). Importantly, validation studies show that treatment effects generated by either Turk or Lucid resemble those generated using probability-based sampling designs (Coppock and McClellan 2019; Mullinix et al. 2015).

### Choice of Politician

I progressively build in greater external realism via the choice of politician. In Experiment 1, subjects read about 'Representative A', a politician who they have 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 in which people possess prior attitudes toward the politician. I address this issue in Experiment 2 by first providing respondents with information about a politician and measuring their initial impression; see Appendix C for

**Table 1.** Overview of experiments

|  | Experiment 1 | Experiment 2 | Experiment 3a | Experiment 3b |
|---|---|---|---|---|
| Sample | MTurk ($n = 1,816$) | MTurk ($n = 1,611$) | Lucid Fulcrum ($n = 1,214$) | |
| 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 |

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

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 this experiment closer to approximating real-world contexts, it nevertheless focuses on a politician who respondents have only learned about once and will never hear about again. Thus in Experiment 3 I randomly assigned 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 that resulted 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. I chose these actions because they are ones that a legislator may need or choose to take for a variety of reasons, but ones that are likely to be unpopular because they impose 'costs' on constituents.[3] Respondents are also unlikely to have a prior attitude about these issues.

Experiment 3 focuses 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 focuses on his change from opposing to supporting the Republican tax cuts passed in late 2018. Politicians may have a particularly difficult time explain-ing away disagreements with constituents on these types of salient political issues (Peterson and

---

[3]The spending cut manipulation was based on a similar scenario used in McGraw, Timpone and Bruck (1993).

Simonovits 2017). Moreover, they involve 'flip-flopping', a behavior that may generate negative impressions on its own (Doherty, Dowling and Miller 2016).

### Experimental Procedure

In each experiment I randomly assigned respondents to conditions that varied 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, the politician grounded 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 local teachers, non-partisan experts or partisans from the opposite party as Representative A. Pre-test results reported in Appendix B show that respondents considered local teachers and non-partisan experts to be more credible than the more explicitly partisan source. However, while this manipulation should provide variation in credibility, on average, it does have a drawback. Respondents who belong to a different party than Representative A and were 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. I 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'. I adopted this language to stimulate respondents to form a 'strong' attitude about 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 need for cognition and affect.[6]

Respondents were assigned to one of four conditions after this buffer. In all four conditions they read a news article detailing how 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. 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

---

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

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

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

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

*Low-Credibility* version, 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 were from the same party. Respondents from a different party than the representative should thus also be somewhat distrusting of this other speaker given this difference in partisanship. Yet co-partisans to Representative Williams may also feel that this speaker is untrustworthy given that he has electoral incentives to impugn Williams' motives.[8] Regardless of the politician's background, the counter-explanation always had the speaker suggesting that 'lobbyists for Burlington Electric Department had 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 that in 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 the need for cognition was used as a short buffer before respondents completed the first politician experiment, after which they 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 in which a counter-explanation was also provided.[9] The justifications used in the experiment were the actual rationales the two politicians publicly offered for their behavior. Senator Gillibrand's justification focused on how becoming a senator had presented her with opportunities to learn about the consequences of gun violence after 'meeting with the families of gun violence victims'. Senator Corker 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 came from real discourse. The counter-explanation in the Gillibrand experiment was taken from a Republican National Committee press release suggesting that her change of heart was motivated by 'political opportunism'. The counter-explanation in the Corker experiment focused on insinuations that he changed his vote due to the inclusion of a provision in the bill that benefited him financially (for example, Kim 2017). Respondents could receive these messages from either a left-wing or a right-wing source.[10] Variation in credibility thus depends on the audience member's political affiliations. For example, 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. I regress this variable on indicators of the experimental treatment conditions to which the respondent was assigned. I use respondents assigned to the Justification condition as the baseline in these analyses given this condition's central role in all hypotheses.[11]

---

[8]Co-partisans of 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). Appendix C provides pre-test analyses pertaining to source credibility that also suggest that this speaker should have reduced credibility.

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

[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]I thank an anonymous reviewer for suggesting this way of presenting the analyses.

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 (Clifford, Sheagley and Piston 2020; Gerber and Green 2012, Ch. 4). In Experiment 2 I 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, I 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 expected based on prior work on the role of prior attitudes, partisanship and policy agreement in shaping evaluations (for example, Ansolabehere and Jones 2010; Lodge and Taber 2006).

Appendix A reports the results of analyses omitting these covariates. Appendix B provides 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 improved in the post-test. In both cases, the substantive conclusions remain the same as those reported in the main findings.[12]

## Results

### *Do Explanations Work Less Well When Counter-Explanations are Present?*

I begin with Figure 1 and an examination of Hypotheses 1 and 2. Figure 1 provides ordinary least squares coefficients from four regression models. For these initial analyses I merge all respondents who received 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 Hypotheses 1 and 2.

Figure 1 communicates two key points. First, providing a justification is no panacea. In Experiments 1 and 3a (Gillibrand) respondents reported 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 did offering a justification improve evaluations. As I discuss below, 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 be related to the specific action of the politician, although this is purely speculative. In Experiment 2 the politician casts a tie-breaking vote for a policy that will impose financial costs on constituents. Approximately 42 per cent of respondents in the no explanation baseline condition reported more negative evaluations regarding the politician 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 culpable for imposing financial costs on constituents.

Secondly, respondents assigned to a counter-explanation condition reported significantly worse evaluations than those in the justification condition in all cases, as Hypothesis 2 predicts. Moreover, the impact of the counter-explanation did not simply balance out 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.

---

[12]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 and 2) and motive attributions (all experiments). The appendices contain pre-analysis plans. I deviate from this plan in one respect: while I pre-registered analyses that combined the different counter-explanation treatments together for Experiments 2 and 3 (e.g., Figure 1), I did not do so for Experiment 1.

[13]I tested the equivalence of the two coefficients (*No Justification* and *Justification w/Counter*) with Wald tests to ascertain this. In three cases I 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.
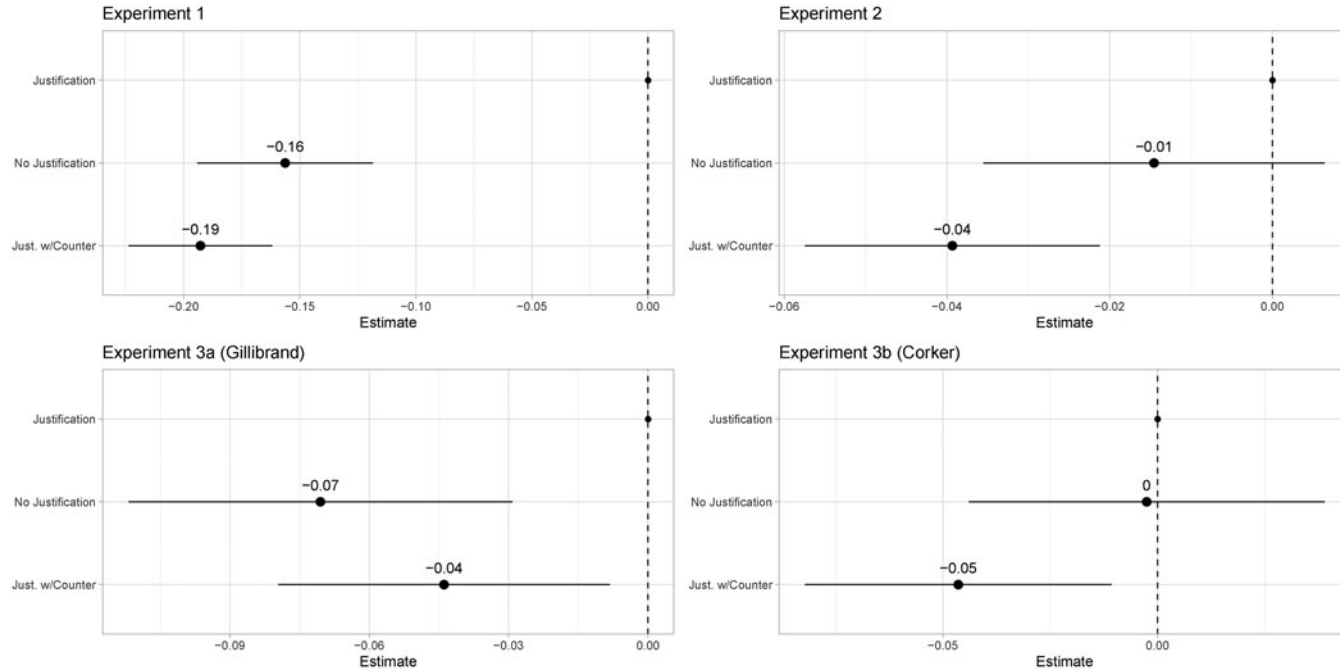
**Figure 1.** Explanations only work when they are unopposed

*Note*: markers provide the average difference in post-test evaluations (with 95 per cent confidence intervals) relative to respondents assigned to the Justification condition (represented by a marker at 0). See Tables A1, A2 and A4 for model results.

Ultimately, Figure 1 provides some supportive evidence in favor of Hypotheses 1 and much more solid evidence in favor of Hypotheses 2.

### Does the Source of the Counter-Explanation Matter?

Figure 1 showed that politicians who hope to avoid negative evaluations by justifying themselves had better hope that another actor does not insinuate that they have ulterior motives. Hypothesis 3 predicts that counter-explanations should be more effective when they are voiced by more credible sources. I 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 separated by source. The bottom facet 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. Experiments 1 and 2 provide the clearest differentiation in source credibility at this level of respondent aggregation.

Figure 2 displays two key results. 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, and the difference between the high- and low-credibility effects in the bottom facet are 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, since the coefficient for the low-credibility source is in line with, and not statistically distinct from, the coefficient for those assigned to the No Justification condition.

Secondly, Figure 2 shows that even a source with low credibility can undermine the effectiveness of the politician's explanation in these first two experiments, 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 that a weak argument does not influence opinions, and that 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, my 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 (2007a, 2007b), thereby enabling the justification to still influence evaluations (particularly in Experiment 1). In other words, my finding that the counter-explanation from the low-credibility source continued to fully balance out the effect of the otherwise substantially effective explanation demonstrates 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, I do not expect to see significant 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 weak 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). I return to this point in the following section. For now, the clear message for politicians is to hope that media coverage of their behavior focuses attention on their partisan rivals rather than on 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?

I now turn to the role of partisanship. I am especially interested in co-partisans' reactions to the politician offering a justification. Hypothesis 4 predicts that these co-partisans will be especially
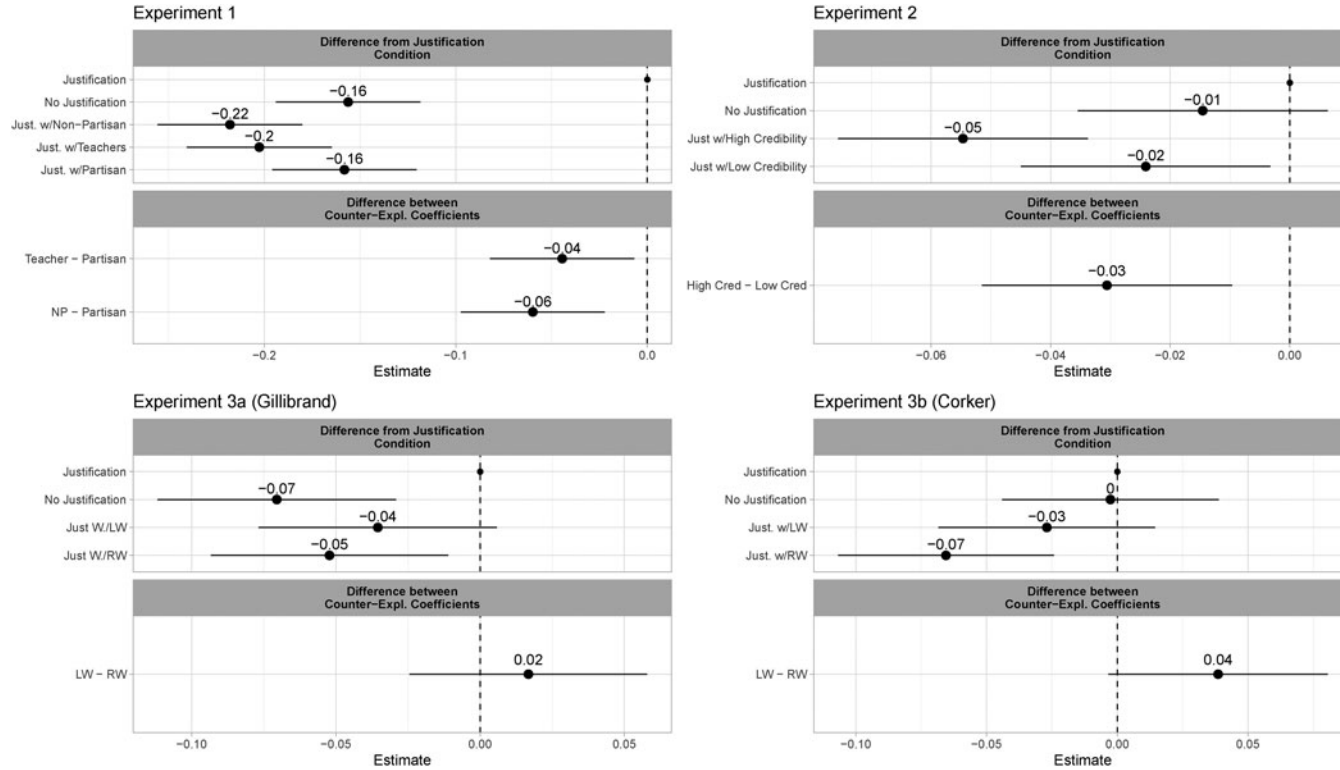
**Figure 2.** Both partisan and non-partisan sources harm evaluations

*Note*: 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 and 2; Left Wing – Right Wing for Experiments 3a and 3b). Negative coefficients indicate that the former source had a larger negative impact on evaluations than the latter source. See Tables A1, A2, and A5 for model results.
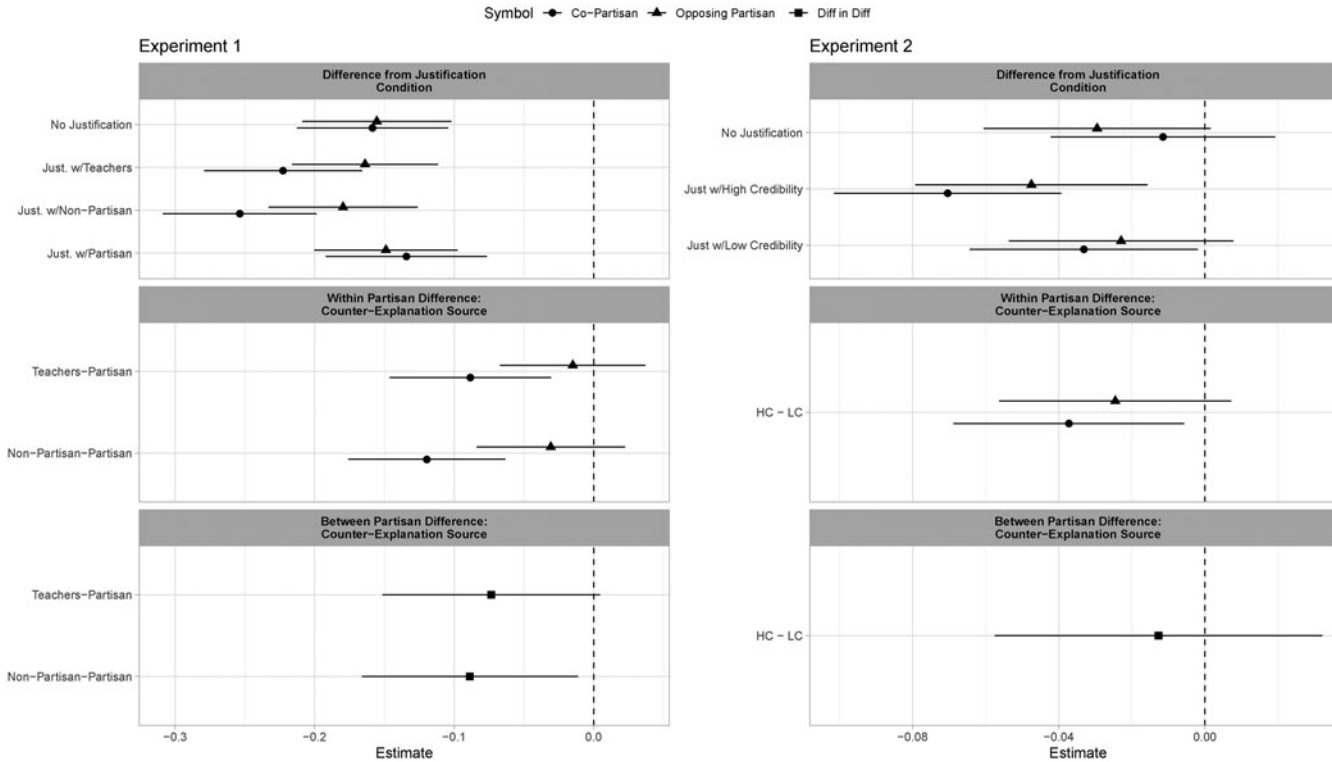
**Figure 3.** Co-Partisans pay more attention to credibility in Experiments 1 and 2

*Note*: 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 (for example, Co-Partisan[HC-LC]). The final facet provides a difference in difference (for example, Co-Partisan [HC-LC] – Opposing Partisan[HC-LC]).

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? Secondly, are the differences observed within the partisan group also significant from one another? That is, 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.

Figure 3 focuses on Experiments 1 and 2. The results it displays are based on a model in which evaluations are regressed on the information condition, co-partisanship status and their interaction (along with pre-test evaluations for Experiment 2; see Tables A1, A2 and A6). 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 (for example, Teachers – Partisan for Co-Partisans), while the bottom panel provides the difference-in-differences estimate (for example, [Co-Partisan: Teachers – Partisan] – [Opposing Partisan: Teachers – Partisan]) Significant negative estimates in these two latter panels would support Hypothesis 4.

The 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 types of partisans when the counter-explanation came 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 is out-partisans. Co-partisans were thus not immune to argumentation impugning a politician from their side.

The middle and bottom facets of Figure 3 consider whether co-partisans were more sensitive to source credibility; the evidence was mixed. 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 the results from Experiment 1 should be interpreted with caution, 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-differences estimate is only statistically significant in Experiment 1 when the high-credibility source was a non-partisan.[14] These results thus fail to produce consistent evidence in support of Hypothesis 4.[15]

Figure 4 provides the same style of graph but for Experiments 3a and 3b. It demonstrates the role of partisanship in shaping reactions to the politician's justification. While both types of partisans reacted positively to the Gillibrand justification, the Corker experiment yielded more 'traditional' partisanship effects, as partisans were 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.

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

---

[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. I consider the role of power in Appendix A, and conclude this consideration is unlikely to be driving the lack of effect here.
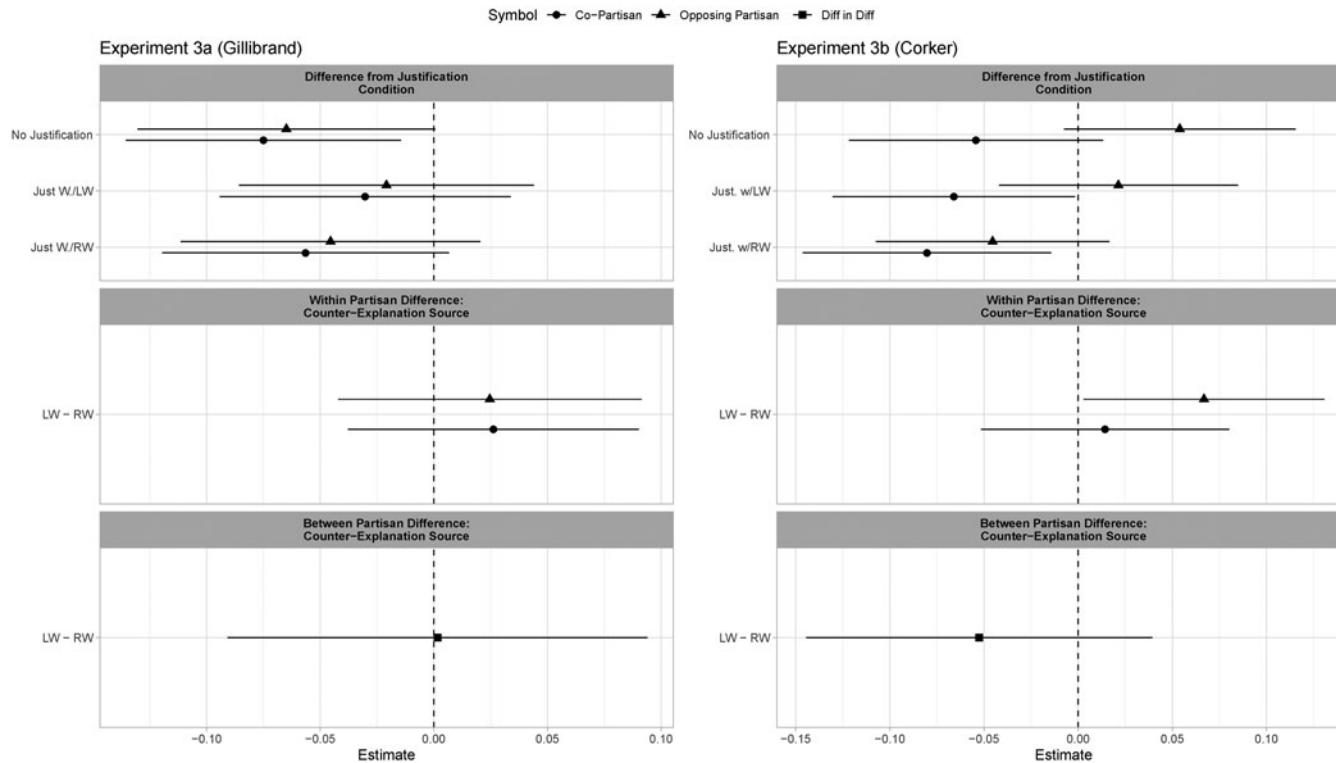
**Figure 4.** Co-Partisanship Matters Less in Experiment 3

*Note*: 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 (for example, Co-Partisan[LW-RW]– Opposing Partisan[LW-RW]).

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).

These results are not consistent with Hypothesis 4. Co-partisans in the Corker experiment reported significantly worse evaluations than for the Justification Only condition regardless of the source of the counter-explanation; the resulting difference between the two coefficients is statistically insignificant. Opposing partisans (Democrats) were 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 in which 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 (for example, Alt, Lassen and Marshall 2015). The resulting difference-in-differences estimate was also insignificant. Partisans in the Gillibrand experiment did not react significantly different based on the source of the counter-explanation in either test.

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 less credible 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 being 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 increase the temptation to shirk mass preferences (Grose, Malhotra, and Van Houweling 2015, 741). This article contributes to the literature on explanation giving, as well as work on elite accountability and democratic representation, by showing that explanations offer 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 hand, I have shown that rival explanations that root a politician's behavior in an ulterior motive can counteract 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 of blame management is a risky strategy that may mostly be effective when the politician's behavior 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. Study participants received both types of rhetoric at the same time, as occurs in most studies of explanation giving and competitive elite communications (for example, 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, and for others it may be the reverse. Future work could explore two potential implications of timing. 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, elected officials would have an incentive to 'go public' with rationales for their policy actions as quickly as possible to inoculate supporters against alternative understandings of their behavior.

Secondly, it may be that the absence of a positive explanation effect seen in these experiments is a best-case scenario for politicians. Previous studies on elite explanations contrast two counter-factuals: (1) no explanation is provided and (2) and explanation is provided. However, the

relevant counterfactuals for politicians may be (1) the *counter-explanation* stands alone and (2) the counter-explanation competes against the politician's own account. Perhaps in such instances a politician's best hope is to tread water by providing an explanation. Appendix F reports the results of a small experiment that sheds some light on this possibility. Participants were randomly assigned to conditions like those in Study 1 but with an additional wrinkle: some received the 'counter'-explanation but not the explanation itself. As reported in the main results, in this small experiment I find that justifications mollify otherwise upset respondents, and that counter-explanations attenuated this effect. Notably, I did not find that the counter-explanation undermined evaluations relative to a baseline condition, although I 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 should fully untangle how people respond to rival accounts of elite behavior over time.

Across the four experiments there were inconsistent effects for the focal politician's justification. Justifications are not panaceas; indeed, the existing literature on explanation giving highlights cases in which 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, since 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, the politician was unable to explain away an action for which he was she was clearly responsible, and the material costs to voters were direct and easy to understand. Perhaps justifications are more effective on 'hard' issues where individuals have weaker prior attitudes and find it more difficult to ascertain the full consequences of elite actions. However, the Corker experiment highlights the potential role of polarization and identity. His justification polarized partisans, leading the justification to have an overall null effect. His actions were also closely tied to President Trump. One possibility is that actions that 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 of gun control (Experiment 3a) cuts against this to a certain extent. Future research should unpack this question.

This study has important potential implications for understanding elite incentives and behavior. On the one hand, the 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 exploit 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). These results may provide one way to address 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 has important implications for elite strategy and broader normative debates. The 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 when delivered by 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

---

[16]I thank an anonymous reviewer for calling attention to this important question.

careful in simply accepting the stance of their elected officials. More pessimistically, however, the counter-explanations explored here worked by changing the substance of the 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 are incentivized to orient their attention to the contents of their character and away from the contents of their policies.

## References

**Aalberg T, Strömbäck J and de Vreese CH** (2011) The framing of politics as strategy and game: a review of concepts, operationalizations and key findings. *Journalism* **13**(2), 162–178.

**Adams J, Ezrow L and Somer-Topcu Z** (2011) Is anybody listening? Evidence that voters do not respond to European parties' policy statements during elections. *American Journal of Political Science* **55**(2), 370–382.

**Alt JE, Lassen DD and Marshall J** (2015) Credible sources and sophisticated voters: when does new information induce economic voting? *The Journal of Politics* **78**(2), 327–342.

**Ansolabehere S and Jones PE** (2010) Constituents' responses to congressional roll-call voting. *American Journal of Political Science* **54**(3), 583–597.

**Bianco WT** (1994) *Trust: Representatives & Constituents*. Ann Arbor: University of Michigan Press.

**Bittner A** (2011) *Platform or Personality? The Role of Party Leaders in Elections*. Oxford: Oxford University Press.

**Bøggild T** (2016) How politicians' reelection efforts can reduce public trust, electoral support, and policy approval. *Political Psychology* **37**(6), 901–919.

**Bøggild T, Aarøe L and Petersen MB** (2020) Citizens as complicits: distrust in politicians and biased social dissemination of political information. *American Political Science Review*. doi: 10.1017/S0003055420000805

**Boudreau C and MacKenzie SA** (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 DE and Butler DM** (2017) The causal effects of elite position-taking on voter attitudes: field experiments with elite communication. *American Journal of Political Science* **61**(1), 208–221.

**Bullock JG** (2011) Elite influence on public opinion in an informed electorate. *American Political Science Review* **105**(3), 496–515.

**Butler DM and Powell EN** (2014) Understanding the party brand: experimental evidence on the role of valence. *The Journal of Politics* **76**(2), 492–505.

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

**Chong D and Druckman JN** (2007a) Framing public opinion in competitive democracies. *American Political Science Review* **101**(4), 637–655.

**Chong D and Druckman JN** (2007b) Framing theory. *Annual Review of Political Science* **10**, 103–126.

**Chong D and Druckman JN** (2010) Dynamic public opinion: communication effects over time. *American Political Science Review* **104**(4), 663–680.

**Clarke N et al.** (2018) *The Good Politician: Folk Theories, Political Interaction, and the Rise of Anti-Politics*. Cambridge: Cambridge University Press.

**Clifford S, Sheagley G and Piston S** (2020) Increasing Precision in Survey Experiments Without Introducing Bias. *APSA Preprints*. doi: 10.33774/apsa-2020-tbd3c.

**Coppock A and McClellan OA** (2019) Validating the demographic, political, psychological, experimental results obtained from a new source of online survey respondents. *Research & Politics* **6**(1), 1–14.

**Cuddy AJC, Fiske ST and Glick P** (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 RA** (1971) *Polyarchy: Participation and Opposition*. New Haven, CT: Yale University Press.

**Doherty D** (2015) Perceived motives in the political arena. *American Politics Research* **43**(3), 363–393.

**Doherty D, Dowling CM and Miller MG** (2016) When is changing policy positions costly for politicians? Experimental evidence. *Political Behavior* **38**(2), 455–484.

**Douglas A** (2016) NC GOP senator says he's dropping Medicare plan attacked by opponent. *McClatchy DC*, 7 September. Available from http://www.mcclatchydc.com/news/politics-government/election/article100489837.html (accessed 2 December 2020).

**Druckman JN, Fein J and Leeper TJ** (2012) A source of bias in public opinion stability. *American Political Science Review* **106**(2), 430–454.

**Eggen D** (2009) Industry cash flowed to drafters of reform; key senator Baucus is a leading recipient. *The Washington Post*, 21 July.

**Fein S, Hilton JL and Miller DT** (1990) Suspicion of ulterior motivation and the correspondence bias. *Journal of Personality and Social Psychology* **58**(5), 753–764.

**Fenno RF** (1978) *Home Style: House Members in Their Districts*. Glenview, IL: Scott Foresman & Co.

**Gerber AS and Green DP** (2012) *Field Experiments: Design, Analysis, and Interpretation*. New York: W.W. Norton & Company, Inc.

**Grose C, Malhotra N and Van Houweling RP** (2015) Explaining explanations: how legislators explain their policy positions and how citizens react. *American Journal of Political Science* **59**(3), 724–743.

**Hibbing JR and Alford JR** (2004) Accepting authoritative decisions: humans as wary cooperators. *American Journal of Political Science* **48**(1), 62–76.

**Hibbing JR and Theiss-Morse E** (2002) *Stealth Democracy: Americans' Beliefs About How Government Should Work*. Cambridge: Cambridge University Press.

**Iyengar S, Norpoth H and Hahn KS** (2004) Consumer demand for election news: the horserace sells. *The Journal of Politics* **66**(1), 157–175.

**Jerit J** (2009) How predictive appeals affect policy opinions. *American Journal of Political Science* **53**(2), 411–426.

**Kim SM** (2017) Why Corker flipped on the tax bill. *Politico*, 18 December. Available from https://www.politico.com/story/2017/12/18/bob-corker-tax-bill-kickback-republicans-respond-302482 (accessed 2 December 2020).

**Kingdon JW** (1989) *Congressman's Voting Decisions*. Ann Arbor: University of Michigan Press.

**Lau RR, Sigelman L and Rovner IB** (2007) The effects of negative political campaigns: a meta-analytic reassessment. *The Journal of Politics* **69**(4), 1176–1209.

**Laustsen L and Bor A** (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 MS and Horowitz MC** (2012) When backing down is the right decision: partisanship, new information, and audience costs. *The Journal of Politics* **74**(2), 323–338.

**Lodge M and Taber CS** (2006) Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science* **50**(3), 755–769.

**Malle BF** (2011) Time to give up the dogmas of attribution: an alternative theory of behavior explanation. *Advances in Experimental Social Psychology* **44**, 297–352.

**Malle BF, Guglielmo S and Monroe AE** (2014) A theory of blame. *Psychological Inquiry* **25**, 147–186.

**Mansbridge JJ** (2003) Rethinking representation. *American Political Science Review* **97**(4), 515–528.

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

**Mayo R** (2015) Cognition is a matter of trust: distrust tune cognitive processes. *European Review of Social Psychology* **26**(1), 283–327.

**McDermott R** (2011) Internal and external validity. In Druckman JN, Green DP, Kuklinski JH and Lupia A (eds), *Cambridge Handbook of Experimental Political Science*. Cambridge: Cambridge University Press, pp. 27–41.

**McGraw KM** (1990) Avoiding blame: an experimental investigation of political excuses and justifications. *British Journal of Political Science* **20**(1), 119–1131.

**McGraw KM** (1991) Managing blame: an experimental test of the effects of political accounts. *The American Political Science Review* **85**(4), 1133–1157.

**McGraw KM** (2001) Political accounts and attribution processes. In Kuklinski JH (ed.), *Citizens and Politics: Perspectives from Political Psychology*. Cambridge: Cambridge University Press, pp. 160–198.

**McGraw KM** (2003) Political impressions: formation and management. In *Oxford Handbook of Political Psychology*. Oxford: Oxford University Press, pp. 394–433.

**McGraw KM, Best S and Timpone R** (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 KM, Timpone R and Bruck G** (1993) Justifying controversial political decisions: home style in the laboratory. *Political Behavior* **15**(3), 289–308.

**Mullinix KJ et al.** (2015) The generalizability of survey experiments. *Journal of Experimental Political Science* **2**(2), 109–138.

**Nilsen E** (2019) Kirsten Gillibrand used to have moderate views on immigration and guns. Voters want to know why she's changed. *Vox*, 9 April. Available from https://www.vox.com/2019/4/9/18303526/kirsten-gillibrand-cnn-town-hall-immigration-guns (accessed 2 December 2020).

**Peterson E and Simonovits G** (2017) Costly values: the limited benefits and potential costs of targeted policy justifications. *Journal of Experimental Political Science* 4(2), 95–106.

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

**Priester JR and Petty RE** (1995) Source attributions and persuasion: perceived honesty as a determinant of message scrutiny. *Personality and Social Psychology Bulletin* **21**(6), 637–654.

**Robison J** (2017) The role of elite accounts in mitigating the negative effects of repositioning. *Political Behavior* **39**(3), 609–628.

**Robison, J.** (2021), "Replication Data for: Can Elites Escape Blame by Explaining Themselves? Suspicion and the Limits of Elite Explanations", https://doi.org/10.7910/DVN/W14GCV, Harvard Dataverse, V1, UNF:6:yVvwJM/ErgBmhez13hNcyQ== [fileUNF]

**Tomz M and Van Houweling RP** (2009) The electoral implications of candidate ambiguity. *American Political Science Review* **103**(1), 83–98.

**Weaver RK** (1986) The politics of blame avoidance. *Journal of Public Policy* **6**(4), 371–398.

**Weitz-Shapiro R and Winters MS** (2016) Can citizens discern? Information credibility, political sophistication, and the punishment of corruption in Brazil. *The Journal of Politics* **79**(1), 60–74.