Public Fears of Terrorism, Partisan Rhetoric, and the Foundations of American Interventionism

Pre-Analysis Plan

Daniel Silverman, dmsilver@andrew.cmu.edu
Daniel Kent, kent.249@osu.edu
Christopher Gelpi, gelpi.10@osu.edu

General Background

Introduction

Since the attacks of September 11th, 2001, counterterrorism has become the central goal that underpins U.S. foreign security policy. According to one key estimate, the U.S. spent $4.79 trillion from 2001 to 2016 on counterterror policies, including the wars in Afghanistan and Iraq, the conflicts in Syria and Pakistan, and homeland security (Crawford 2016).

In one sense, the centrality of terrorism for American foreign policy over the last 17 years seems appropriate in light of the public's grave concern over this issue. As recently as June 2017, for example, Gallup (2017) found that 60% of the American public believed that it was "very" or "somewhat" likely that there would be terrorist attacks in the U.S. "over the next several weeks." On the other hand, this public fear persists despite the fact that the actual risk of death that Americans face from terrorist attacks is remarkably low, with an average of 87 deaths on U.S. soil per year from 1970-2007 – far fewer than many other threats to public safety on which we spend far less resources (Mueller, 2006; Mueller and Stewart, 2015). Meanwhile, according to Department of Defense statistics, more Americans have died in U.S. military operations touted as combating terrorism (6,950) than have been killed by terrorists.

In fact, such observations are not new – scholars of political science and other disciplines have written at length about the American public's overreaction to terrorism. Empirical research has investigated how basic cognitive errors (Sunstein, 2003), emotions like fear or anger (Huddy et al., 2005), widespread religious stereotypes (Sides and Gross, 2013), media coverage of terrorism (Nacos et al., 2011), and the prevailing political discourse around the threat (Friedman et al., 2010; Winkler, 2006) contribute to this stark gap between perception and reality. However, there has been surprisingly little work on a simple – and yet profoundly important – question: can this costly misperception be corrected? If so, how?

In this study, we explore the degree to which public fear of terrorism can be diminished by the information people are given regarding the risks of terror as well as the political rhetoric surrounding the delivery of that information. In doing so, we investigate whether *changes* in the elite discussion of terrorism could alleviate public fears surrounding terrorist attacks and offer a more realistically grounded foundation for American foreign security policy and grand strategy. Our effort thus adds to the few extant studies that do aim to diminish overreactions to terrorism (Bausch et al., 2013; Hoffman and Shelby, 2017). While these studies are useful, we attempt to go beyond them by (1) testing the impact of *political* cues – specifically, endorsements by partisan and non-partisan political elites – in our corrections, and (2) fielding the survey on a nationally representative sample in order to gauge its potential to change minds among the general public (as opposed to opt-in or convenience samples). Meanwhile, our study also speaks to the literature on the correction of misperceptions about politics more generally (Nyhan and Reifler, 2010; Wood and Porter, 2018) by exploring the extent to which citizens update their beliefs after being exposed to facts about the risk of terrorism in different rhetorical contexts.

In a pilot experiment conducted on Amazon's Mechanical Turk (MTurk) in late 2017, we found encouraging results suggesting that the American public's exaggerated fear of terrorism is surprisingly correctable when respondents are provided factual information about the threat reinforced by affirmations from co-partisan elites. With support from Time-Sharing Experiments for the Social Sciences (TESS), in this study we replicate and extend these initial results on a nationally representative sample of Americans with a wider range of elite cues that allow us to substantially build on our findings. We will also field the same revised design simultaneously on MTurk in order to provide additional evidence and investigate the ability of MTurk to recover the results of experiments like ours as compared to high-quality, representative platforms (Coppock, 2018). Moreover, while the design of the revised experiment will be exactly the same on MTurk as on TESS, we will take advantage of the low cost of MTurk samples to field a follow-up wave two weeks after subjects receive the treatments in order to evaluate their persistence. This replicates something we successfully implemented in the pilot study (as discussed below), adding another key layer to our results.

Ultimately, we will use the results of this study to produce three papers: (1) a substantive paper contributing to the literature on terrorism, counterterrorism, and American foreign policy, (2) a substantive paper contributing to the literature on public misperceptions and the correction of false beliefs, and (3) a methods piece adding to research on different survey platforms and the external validity of MTurk in particular.

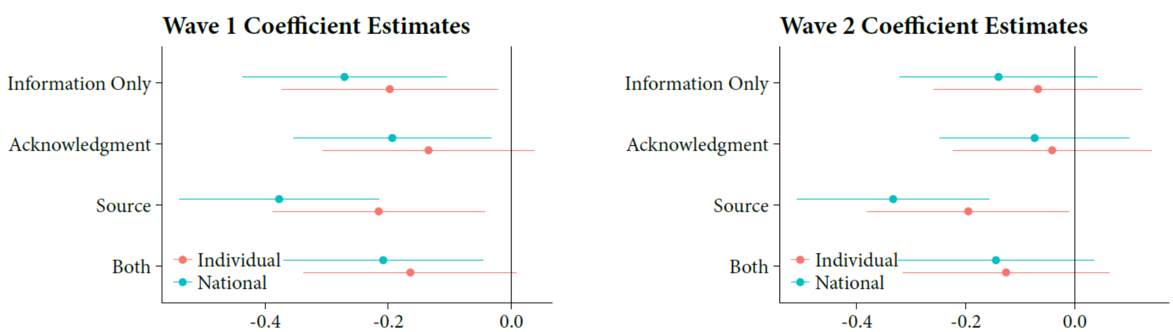
Pilot Study – Overview and Results

In November 2017, we conducted a pilot study on MTurk (N=800) in which we provided subjects with factual information about the risks of death from terrorism relative to other sources of mortal risk. We varied the way in which the information was presented along two dimensions. First, we randomly assigned some subjects to conditions that *acknowledged* the understandable persistence of public fear despite the low levels of risk to explore whether acknowledging their concerns might cause respondents to be less defensive about updating their beliefs. Second, we randomly assigned some people to receive a message from a co-partisan elected official *endorsing* the information regarding the low risk of death from terror attack. The study thus entailed a fully randomized 2x2 (plus control group) experimental design, with a planned interaction term for participants' partisan identification. Moreover, the study included a second wave two weeks later in which we re-contacted participants (84.25% response rate) in order to examine the persistence of any effects over time.

To measure respondents' beliefs about the threat posed by terrorism, we used two types of outcome questions – those capturing beliefs about (1) threat to the individual participant and (2) threat to the nation as a whole. We view both of these as key layers of the public's overreaction to the threat of terrorism. For each of these types of DVs (personal and national), we asked two questions. Principal-factors factor analysis shows that – across both survey waves – responses to the questions in each category hang closely together, with α scores above 0.7 for both pairs of items. Thus, we averaged participants' responses to the individual threat questions as well as the national threat questions and used these as our two primary DVs in the study. In each wave of the survey, we then simply estimated OLS regression models to capture the average impact of each treatment on these two threat scales.

Figure 1 presents the main results for both the first and second waves. As can be seen, the acknowledgment condition did little to alleviate fears of a terrorist attack. However, endorsement of the information regarding public safety from terrorism by a co-partisan politician had a robust and enduring effect on alleviating public fear. In particular, even subjects who were re-contacted two weeks after the study continued to state that they were significantly less concerned about the likelihood of a terror attack after reading factual information about the risk of terrorism relative to other threats, in combination with a co-partisan elite endorsement of these facts. These results show that elite rhetoric surrounding the threat of terrorism is an important source of the problem of exaggerated public fears. But just as importantly, they suggest that politicians have the opportunity to be part of the solution as well.

Figure 1: Treatment effects for the initial and follow-up surveys.

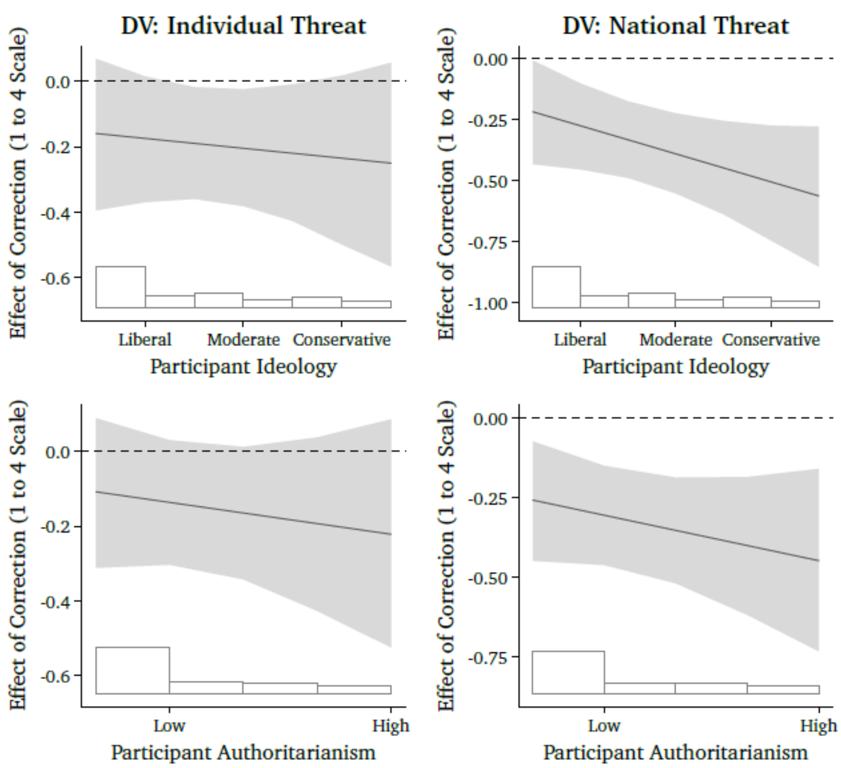


The x-axis represents the magnitude of a treatment effect: the difference between the average response in the control and treatment group. Responses are on a four-point scale where 1 corresponds to assigning terrorism the lowest threat level and 4 the highest threat level, meaning negative coefficients represent a decrease in the average fear of terrorism. The 'both' treatment condition includes the acknowledgment and source. 95% confidence intervals are included. The light red coefficient estimates and confidence intervals are from models where participants are asked about terrorism's threat to their individual security. The light blue coefficient estimates and confidence intervals are from models where participants are asked about the threat posed by terrorism to national security.

We also checked for the presence of theoretically motivated interaction effects between the treatment and both (1) political ideology as well as (2) right-wing authoritarianism (RWA). We included ideology as a plausible moderator to test for backfire effects among conservatives (e.g., Nyhan and Reifler 2010) given that we would expect them to be most skeptical of reassuring information about terrorism. RWA is also included as a key moderator as it is deeply linked to aggressive responses to threat in general and terrorism in particular (Hetherington and Suhay, 2011). Due to multicollinearity concerns, the two potential moderating factors were modeled as interaction terms with the treatment in separate OLS regressions. For each one, we interacted either ideology or RWA with only the elite source cue treatment, as it was found throughout to be the most robust and effective.

As seen in Figure 2, we found surprising moderating effects with respect to perceptions of national (though not individual) threat. In particular, not only did we observe no evidence of backfire among conservatives, we found that the treatment effects were actually *greater* among conservatives than liberals. Moreover, we found the same effects around authoritarianism, with authoritarians updating their national threat perceptions *more* in response to the correction than non-authoritarians. A closer look into the data reveals that this is likely the result of a "floor effect," whereby conservatives (and authoritarians) in the control group perceive terrorism as far more threatening than liberals (and non-authoritarians) in the first place, leaving more room for them to update their beliefs. This preliminary finding highlights the need for scholars to pay attention to floor (or ceiling) effects when comparing the correctability of beliefs across different sub-groups, while also generally reinforcing recent evidence showing that backfire is rare in contemporary political life (Wood and Porter, 2018).

Figure 2: Initial survey interaction effects



Interaction effects of ideology and authoritarianism upon treatment effect. Each curve represents the effect of treatment by ideological or authoritarian categories. Ideology is measured on a seven-point scale where 1 *corresponds to liberal and 7 to conservative. Authoritarianism is measured on a four-point scale where 1 corresponds to the lowest and 4 to the highest level of authoritarianism. 95% confidence intervals and histograms of data distributions are included. The effect denoted on the y-axis is the same as the x-axis in figure 1: negative estimates indicate a decrease in threat perceptions on a four-point scale.*

Revised Study – Motivation and Design

While intriguing, the results of the pilot study point toward at least three key areas for continued investigation. First, as is often the case, our MTurk sample differed from a representative sample of the public in several key ways. In particular, our sample skewed more Democratic, educated, and comfortable with numbers than the American public overall. We have some reason to be concerned that the efficacy of our treatments may be conditioned by all three of these factors. For example, the Republican party's longstanding rhetoric about the "War on Terror" may insulate rank-and-file Republican's from the effect of the treatment. We instead found some suggestive evidence that conservatives (and authoritarians) might be particularly responsive to these facts when affirmed by trusted co-partisan elites. However, these intriguing findings require further exploration with a nationally representative sample. Fortunately, we have obtained support for fielding this study on a national probability sample through TESS and the National Opinion Research Center (NORC) at the University of Chicago.

Second, we will expand on the treatments in the pilot study by exposing participants to a richer set of elite cues reinforcing the factual information provided about the threat of terrorism. Specifically, while the treatments in our pilot study always matched subjects with rhetoric from a fictional co-partisan elite, we plan to randomly expose people to endorsements from Democrats, Republicans, and military leaders. This will allow us to (1) see *who* is the most effective source for corrective information among realistic American elite messengers about the terror threat, and (2) test the impact of several different mechanisms for *why* such cues might effectively reinforce these corrections, thus contributing to more theoretical debates about what facilitates successful belief correction. In particular, literature on elite cues highlights three key attributes that make them work: people's perceptions that the cue sender (1) shares their interests, (2) has expertise about the issue at hand, and (3) incurs costs by sending the cue (Lupia and McCubbins, 1994). In our study, we can look at different cue-respondent combinations to get at all three mechanisms. In particular, the co-partisan cues best capture the "shared interests" mechanism, while surprising cross-partisan messages (e.g., Republican elites advocating for lower fear about terrorist attacks) tap into the "costly signaling" mechanism, and cues from military elites about the threat represent the mechanism of "relevant expertise." We outline our plans to isolate these three mechanisms in greater detail in the section on Paper #2 below.

Third, our follow-up study will be fielded simultaneously on NORC and MTurk in order to offer a more direct comparison of the treatment effects estimated on these two samples. Survey experiments conducted on convenience samples of the public (including our pilot study) often face difficult questions about external validity. But the collection of nationally representative samples is expensive, and requiring the execution of an experiment on such samples would substantially limit the number and variety of experimental studies of public opinion. By fielding the same experiment simultaneously on these two platforms, we hope to contribute to a growing literature on the robustness of treatment effects for studies of political attitudes on convenience samples such as those collected on MTurk or through market research panels such as those maintained by Qualtrics (Coppock, 2018).

Experimental Conditions:

Our study has 5 randomly assigned experimental conditions. The conditions include a control group and 4 treatment groups. The treatment groups vary across a single dimension: the identity of the elite endorsement of the corrective information regarding the risks of terrorism. Specifically, the experimental conditions (and their associated labels) are:

1) Control group: no corrective information (*Control*)
2) Treatment 1: corrective information, no elite endorsement (*Correct_Only*)
3) Treatment 2: corrective information, Democratic endorsement (*Correct_Democrat*)
4) Treatment 3: corrective information, Republican endorsement (*Correct_Republican*)
5) Treatment 4: corrective information, military endorsement (*Correct_Military*)

As seen above, the revised study thus employs a fully randomized design with four treatment groups and one control group. The control group only receives a vignette that captures the general perception of terrorism as a threat (*Control*). The treatment groups each read the control vignette and then are provided a corrective vignette with information on the risk of terrorism relative to various other types of public safety hazards. The first treatment group's vignette only includes these risk statistics (*Correct_Only*). The remaining three treatment group vignettes also include an endorsement of the risk statistics by one of three fictitious elite actors: a Democratic congressman (*Correct_Democrat*), a Republican congressman (*Correct_Republican*), or a military general (*Correct_Military*). This design allows us to not only test the effects of co-partisan, cross-partisan, and military cues, but also to examine how these effects vary across Democratic, Republican, and Independent respondents. We will also replicate this new design on MTurk in a two-wave format in which the second wave simply includes asking our six post-treatment questions to participants again.

Paper #1:

Summary: This piece aims to speak chiefly to literature on terrorism, counterterrorism, and U.S. foreign policy. Here our main interest is in whether the corrections work, which one works best, and how they impact not just beliefs but foreign policy preferences.

DVs: As in the pilot study, the two items about individual threat perceptions and two items about national threat perceptions are the primary DVs in the survey. We average individuals' responses to each of these pairs to create individual and national threat perception scales. In addition, given our interest in this paper in the *foreign policy consequences* of terrorism misperceptions, we also use an item about the degree to which citizens think the U.S. should focus on the threat of China (a more traditional foreign policy priority) as a DV to see whether fear of terrorism is distracting people from other foreign threats. The items used are as follows:

Individual Threat Scale:
1) How worried are you that you or someone in your family will become a victim of terrorism?
-[Very worried / somewhat worried / not too worried / or not worried at all]
2) Here is a list of potential threats to your own personal safety over the next 10 years. In your view, how serious is each threat? (Terrorist attacks)
-[Very serious / fairly serious / not too serious / not serious at all]

National Threat Scale:
1) Here is a list of potential threats to our country's national security over the next 10 years. In your view, how serious is each threat? (Terrorism)
-[Very serious / fairly serious / not too serious / not serious at all]
2) Here is a pair of potential goals that the U.S. can pursue in its foreign policy. In your view, how important is each goal? (Preventing future acts of terrorism on U.S. soil)
-[Very important / Fairly important / Not too important / Not important at all]

China Policy Priority:
1) Here is a pair of potential goals that the U.S. can pursue in its foreign policy. In your view, how important is each goal? (Dealing with great power rivals like China)
-[Very important / Fairly important / Not too important / Not important at all]

Analyses: The primary models will simply be OLS regressions of these three DVs on indicators for each treatment group to explore their average effect on the three outcomes. While we do not delve deeply into conditional effects in this paper, because of the foreign policy relevance of determining whether these cues work among different sub-groups of American society, we will estimate each model not just for the full sample but also separately for Democrats, independents, and Republicans. Our expectations here are simple: that the two partisan cues will be stronger on co-partisans, but that the military cue will be strongest overall (because it has bipartisan appeal). The main models to be estimated are thus as follows:

Individual threat models:
[FULL SAMPLE] Individual threat scale = $\beta_0 + \beta_1(\mathit{Correct\_Only}) + \beta_2(\mathit{Correct\_Democrat}) + \beta_3(\mathit{Correct\_Republican}) + \beta_4(\mathit{Correct\_Military})$
[DEMOCRATS ONLY] Individual threat scale = $\beta_0 + \beta_1(\mathit{Correct\_Only}) + \beta_2(\mathit{Correct\_Democrat}) + \beta_3(\mathit{Correct\_Republican}) + \beta_4(\mathit{Correct\_Military})$
[INDEPENDENTS ONLY] Individual threat scale = $\beta_0 + \beta_1(\mathit{Correct\_Only}) + \beta_2(\mathit{Correct\_Democrat}) + \beta_3(\mathit{Correct\_Republican}) + \beta_4(\mathit{Correct\_Military})$
[REPUBLICANS ONLY] Individual threat scale = $\beta_0 + \beta_1(\mathit{Correct\_Only}) + \beta_2(\mathit{Correct\_Democrat}) + \beta_3(\mathit{Correct\_Republican}) + \beta_4(\mathit{Correct\_Military})$

National threat models:
[FULL SAMPLE] National threat scale = $\beta_0 + \beta_1(\mathit{Correct\_Only}) + \beta_2(\mathit{Correct\_Democrat}) + \beta_3(\mathit{Correct\_Republican}) + \beta_4(\mathit{Correct\_Military})$
[DEMOCRATS ONLY] National threat scale = $\beta_0 + \beta_1(\mathit{Correct\_Only}) + \beta_2(\mathit{Correct\_Democrat}) + \beta_3(\mathit{Correct\_Republican}) + \beta_4(\mathit{Correct\_Military})$
[INDEPENDENTS ONLY] National threat scale = $\beta_0 + \beta_1(\mathit{Correct\_Only}) + \beta_2(\mathit{Correct\_Democrat}) + \beta_3(\mathit{Correct\_Republican}) + \beta_4(\mathit{Correct\_Military})$
[REPUBLICANS ONLY] National threat scale = $\beta_0 + \beta_1(\mathit{Correct\_Only}) + \beta_2(\mathit{Correct\_Democrat}) + \beta_3(\mathit{Correct\_Republican}) + \beta_4(\mathit{Correct\_Military})$

China priority models:
[FULL SAMPLE] China policy priority = $\beta_0 + \beta_1(Correct\_Only) + \beta_2(Correct\_Democrat) + \beta_3(Correct\_Republican) + \beta_4(Correct\_Military)$
[DEMOCRATS ONLY] China policy priority = $\beta_0 + \beta_1(Correct\_Only) + \beta_2(Correct\_Democrat) + \beta_3(Correct\_Republican) + \beta_4(Correct\_Military)$
[INDEPENDENTS ONLY] China policy priority = $\beta_0 + \beta_1(Correct\_Only) + \beta_2(Correct\_Democrat) + \beta_3(Correct\_Republican) + \beta_4(Correct\_Military)$
[REPUBLICANS ONLY] China policy priority = $\beta_0 + \beta_1(Correct\_Only) + \beta_2(Correct\_Democrat) + \beta_3(Correct\_Republican) + \beta_4(Correct\_Military)$

We will estimate all of these models on the MTurk as well as the TESS data. In the MTurk case, we will repeat all of these analyses for the second wave as well as the first. To ensure robustness, we will also include replications of all models with ordered logit regressions as well as with a set of standard demographic covariates (age, gender, education, income, and urbanity) in the paper's online Appendix.

Ultimately, this design will allow us to investigate whether the treatments were effective (and, if so, which ones) on the American public in terms of correcting their misperceptions about the risk of terrorism and shaping their ensuing foreign policy priorities.

Paper #2:

Summary: This piece aims to speak chiefly to literature on the correction of misperceptions and false beliefs. Here our main interest is in *why* different corrections work and which *mechanisms* lead to factual updating among citizens about contentious political issues like terrorism.

DVs: The scales of individual and national threat perceptions are the primary DVs in the paper, as described above.

Derivation of Hypotheses: In order to develop a theoretically grounded set of expectations for which cues work, we turn to the rich literature on elite cues in public opinion formation. Indeed, numerous studies have documented that citizens generally have limited factual knowledge about policy issues, including foreign policy (e.g., Dropp et al., 2014), and often seek to compensate for this lack of empirical knowledge by relying on informational cues that can guide them toward attitudes consistent with their values and interests. In order for a cue to be useful in constructing attitudes conform with values and interests, the literature suggests that several criteria must be met. First, an individual must believe that the cue-giver has some knowledge of the true state of the world. Otherwise, there is no informational content in the cue that would lead individuals to update their beliefs. But expert knowledge is not sufficient for successful persuasion because listeners must be concerned about whether the speaker will deceive them. Lupia and McCubbins (1998) argue that knowledgeable speakers will be persuasive when at least one of two additional conditions are met: 1) the listener perceives him or herself to have common interests with the speaker; or 2) the speaker faces external constraints such as verification of the truthfulness of their message, penalties for lying and or paying a cost to send their message.

With regard to the first criterion, one of the most important indices that individuals use to judge a speaker's interests in the realm of politics is their partisan identification (Page and Shapiro, 1982; Sniderman et al., 1991; Berinsky, 2007). For this reason, much of the literature on opinion formation and survey response has emphasized partisan labels as the central cues that individuals use to construct their attitudes on a variety of issues (Feldman and Zaller, 1992; Bartels, 2002; Zaller et al., 1992; Berinsky, 2007; Achen and Bartels, 2006). Thus we would expect that cues from elites who share a common partisan affiliation with the recipient could be influential in causing them to alter their attitudes or beliefs.

With the second criterion, it is difficult to construct external constraints or penalties for lying in the context of political speech. Politicians regularly make statements that lack empirical support or are even contradicted by the available facts and face few – if any – consequences. However, speech can be costly in a political context if it contradicts the partisan interests of the speaker (Baum and Groeling, 2009). Consequently, elite cues that contradict the partisan predispositions of the speaker should be credible because the speaker would not send the cue unless it were true. This argument suggests that elite cues that contradict the known partisan predispositions of the cue sender should be influential in causing recipients to update their attitudes and beliefs.

Building on the insights in the literature, our revised study will investigate the impact of these three key criteria of elite cues – policy expertise, shared interests, and costly signals – in shaping the efficacy of our corrections. In order to do so, we have embedded co-partisan, cross-partisan, and non-partisan expert cues in our treatments, randomly exposing participants to endorsements from either (1) Democratic, (2) Republican, or (3) military leaders. This design allows us to tap into the mechanisms in the literature in several ways. In particular, co-partisan cues will generally be effective due to the presumption of shared interest between cue giver and recipient. Cross-partisan cues should be influential only when the cue clashes with the opposing party's expected interests. And cues from military officers will generally depend upon perceptions of their expertise and access to information about the effectiveness of military operations. Yet the alignment is not always perfect: for example, a cue endorsing the low risk of terrorism from a Republican politician to a Republican recipient would combine both the shared interest and costly signal mechanisms. Thus, the complementary nature of these mechanisms substantially complicates evaluating and comparing their impact.

While this presents a key challenge, we believe that examining carefully chosen comparisons and patterns of our treatment effects will allow us to tease apart the different mechanisms at work. Starting with the shared interest mechanism, our best opportunity to isolate this comes in the case of Democratic subjects responding to cues from Democratic politicians. This is because terrorism has been widely seen as a "Republican" issue in the United States at least since the attacks of September 11th, 2001, and the Republican party has been more vocal in highlighting the threat of terrorism to the public (e.g., Winkler, 2006). Consequently, a Republican endorsement of the inflated risks of terrorism will likely be seen as more surprising – and thus costly – than a Democratic one. This means that a Republican endorsement to a Republican citizen would likely be interpreted as an indication of shared interests as well as costly signaling. In contrast, Democratic subjects responding to cues from Democratic politicians are likely to be motivated by their sense of shared interests with the cue giver rather than a belief that such a cue would be highly costly to the sender.

Moving to the costly signaling mechanism, isolating this requires a more sophisticated strategy. This is because, while a Republican elite's endorsement of our correction sends a costly signal, it will be mixed with perceptions of shared interest when it is received by a Republican respondent (and perceptions of opposing interest when it is received by a Democratic one). Thus, there is no cue that "purely" captures surprise or costliness alone. The solution to this problem, we propose, is a difference-in-differences strategy. The logic is as follows. As noted, a Republican cue to a Republican citizen (denoted R → R) mixes surprise with shared interest. In contrast, a Democratic cue to a Democratic citizen (denoted D → D) simply taps into shared interest. Thus, the difference between the former and latter (R → R – D → D) gives us a measure of any "surprise effect" net of the impact of shared interest between sender and receiver. Here, like in any difference-in-difference design, we assume parallel trends – that in-group trust is on average equivalent in both Democrats and Republicans. Any significant remaining difference represents the impact of surprise.

Finally, we propose a similar approach to separate the impact of expertise from the other factors. Specifically, as noted above, a Republican cue sent to a Republican citizen (R → R) represents the effect of shared interests – or trust – and surprise. Meanwhile, a military cue sent to a Republican citizen (M → R), we would argue, taps into all three factors: trust, surprise, and expertise. Indeed, given the overwhelming support for military leaders among Republicans in particular (e.g., McDermott and Panagopoulos, 2015), a statement from a four-star general is likely to be as trusted as a co-partisan statement in the eyes of Republican voters. Similarly, given the military's status as the country's main counter-terrorism arm since 9/11 and a source of "strong" and interventionist foreign policy stances, a statement about the inflated risks of terrorism is likely to be similarly surprising when coming from a general as opposed to a Republican elite. Therefore, we argue that the difference between a military-to-Republican cue and a Republican-to-Republican cue (that is, M → R – R → R) allows us to gauge the effect of issue-specific policy expertise net of the effects of trust and of surprise.

Analysis: We formalize the above discussion of estimating these precise mechanisms by drawing upon Angrist and Pischke's (2008) treatment of difference-in-differences designs. In a traditional difference-in-differences model, the researcher measures the within-group changes over time and then looks at how that within-group temporal difference varies across groups, hence the difference-in-differences. In our case, however, the two differences under comparison are not temporal but categorical: treatment effects estimated in a bivariate regression are simply the difference between average control and treatment group responses for relevant combinations. We then take the difference of these categorical differences to isolate our mechanisms of interest.

To begin, our general model can be specified as follows:

$$Y_{irs}|r,s \;=\; \gamma_r \;+\; \beta_j \mathrm{D}_s \;+\; \varepsilon_{irs}$$

where Y_{irs} represents the expected fear in a recipient, given their partisan affiliation (r represents recipient party) and the endorser's identity (s represents the sender's identity). β represents the treatment effect – the average change in the fear of terrorism – given a certain combination of s and r. And γ_r represents the baseline average fear in the control group for each partisan group. Relevant values for r, s, and j are:

$$r = \begin{cases} D & \text{if Dem recipient} \\ R & \text{if Rep recipient} \\ I & \text{if Ind recipient} \end{cases} \quad s = \begin{cases} D & \text{if Dem sender} \\ R & \text{if Rep sender} \\ I & \text{if Mil sender} \\ 0 & \text{if no sender} \end{cases} \quad \text{and } j = \begin{cases} 1 & \text{if Dem} \rightarrow \text{Dem} \\ 2 & \text{if Rep} \rightarrow \text{Rep} \\ 3 & \text{if Mil} \rightarrow \text{Rep} \end{cases}$$

Turning to our treatment effects, though there are more possible sender-receiver combinations for β_j than denoted, we only focus on the following three combinations due to the aforementioned theoretical concerns. β_1 captures the treatment effect for a Democratic recipient and sender. β_2 captures the treatment effect for a Republican recipient and sender. β_3 captures the treatment effect for a Republican recipient and military sender. Also, D_s denotes whether a treatment was applied, and if so the sender's identity. Given this layout, we plan to estimate three models. The first solely measures the impact of a trusted source. This is straightforward with Democratic recipients and Democratic endorsers:

$$E(Y_{0irs}|r,s) = \gamma_r + \beta_j D_s$$

Where D = 0 and $E(Y_{0irs}|r,s)$ = 0, meaning:

$$E(Y_{0irs}|r,s) = \gamma_r$$

From here we are able to get the 'trust only' model:

$$E(Y_{irs}|r = D, s = D) - E(Y_{irs}|\, r = D, s = 0) = \beta_1$$

Next, we consider the 'costly signal' model, in which we look at the difference between a R → R and D → D model:

$$E(Y_{irs}|r = R, s = R) - E(Y_{irs}|\, r = R, s = 0) = \beta_2$$

Drawing upon our 'trust only' model, this gives us a difference-in-differences where our effect of interest is isolated as:

$$\begin{gathered}[E(Y_{irs}|r = D, s = D) - E(Y_{irs}|\, r = D, s = 0)] - \\ [E(Y_{irs}|r = R, s = R) - E(Y_{irs}|\, r = R, s = 0)] = \\ \beta_2 - \beta_1\end{gathered}$$

For both this model and the following model we plan to simply estimate our standard errors through bootstrapping. Last, our 'expertise' model subtracts the effect of R → R from the effect of M → R. Here we estimate the effect of a military endorsement upon Republicans, β_3 as:

$$E(Y_{irs}|r = R, s = M) - E(Y_{irs}|\, r = R, s = 0) = \beta_3$$

Accordingly, we use the same format as in equations 6 through 8:

$$[E(Y_{irs}|r = R, s = M) - E(Y_{irs}|\ r = R, s = 0)] - [E(Y_{irs}|r = R, s = R) - E(Y_{irs}|\ r = R, s = 0)] = \beta_3 - \beta_2$$

Our planned means of estimating these models is to subset the dataset among Republican and Democrat participants and to then fit the relevant models for each. Just like in the 'surprising cues' model, we are assuming parallel trends regarding the trustworthiness and level of surprise around for both groups. This means that any observed effect is assumed to solely represent expertise.

Ultimately, we can use these three estimation strategies to assess which of the three mechanisms – trust, surprise, and expertise – has the largest effect size and is thus the most impactful. Though all three mechanisms has been empirically supported in the literature, our research design presents a novel opportunity to evaluate which is the most impactful.

Lastly, we will estimate all of these models on the MTurk as well as the TESS data. In the MTurk case, we will repeat all of these analyses for the second wave as well as the first. To ensure robustness, we will also include replications of all models with ordered logit regressions as well as with a set of standard demographic covariates (age, gender, education, income, and urbanity) in the paper's online Appendix. Finally, we will include a robustness check in which we add the pre-treatment level of national terrorism threat (the same item is asked in the pre- and post-treatment modules) to the national threat models in order to rule out baseline differences in the DV as the reason for observed variation across sub-groups.

Paper #3:

<u>Summary:</u> This piece aims to speak to the methodological literature on the relative merits of various survey platforms employed in the social sciences. We will be focused on the generalizability of Amazon's Mechanical Turk (MTurk), conditional on political ideology, relative to a more expensive national probability sample.

Does MTurk recover reliable treatment effect estimates in political studies, and does this external validity hold across the political ideology of participants? Although recent work has demonstrated that average treatment effects tend to be reliable in convenience samples via MTurk (Berinsky et. al. 2012, Mullinix et. al. 2015, Coppock 2018), MTurk samples skew more Democratic than Republican. This raises concerns about MTurk's efficacy in conducting studies which hypothesize that treatment effects might vary by political leaning or orientation. In this study, we aim to explore whether this ideological *conditional* average treatment effect systematically differs on a MTurk study about a political topic, relative to the same exact study hosted simultaneously on a probability based population sample. This is vitally important if MTurk is to be used as a major source of data about mass-level politics and public opinion in political science.

DV: Same as above; the two scales about the salience of terrorism as a threat to participants' national and personal security.

In this paper, we will estimate 4 total models. Once again, we are employing the same two DVs of interest: (1) a participant's perception of terrorism as a threat to national security, and (2) a participant's perception of terrorism as a threat to individual security. For each of these DVs, we will estimate every model twice: once with the MTurk data and once with the TESS data. Each model then interacts all treatment conditions with political ideology.

The models are as follows, each estimated once with MTurk data and once with TESS data:

Individual threat scale = $\beta_0 + \beta_1(\textit{Correct\_Only}) + \beta_2(\textit{Correct\_Democrat}) + \beta_3(\textit{Correct\_Republican}) + \beta_4(\textit{Correct\_Military}) + \beta_5(\textit{Ideology}) + \beta_6(\textit{Correct\_Only} * \textit{Ideology}) + \beta_7(\textit{Correct\_Democrat} * \textit{Ideology}) + \beta_8(\textit{Correct\_Republican} * \textit{Ideology}) + \beta_9(\textit{Correct\_Military} * \textit{Ideology})$

National threat scale = $\beta_0 + \beta_1(\textit{Correct\_Only}) + \beta_2(\textit{Correct\_Democrat}) + \beta_3(\textit{Correct\_Republican}) + \beta_4(\textit{Correct\_Military}) + \beta_5(\textit{Ideology}) + \beta_6(\textit{Correct\_Only} * \textit{Ideology}) + \beta_7(\textit{Correct\_Democrat} * \textit{Ideology}) + \beta_8(\textit{Correct\_Republican} * \textit{Ideology}) + \beta_9(\textit{Correct\_Military} * \textit{Ideology})$

Political ideology is a seven-point scale. This leaves us with four treatment conditions (*Correct_Only*, *Correct_Democrat*, *Correct_Republican*, *Correct_Military*), each of which will be estimated and interacted with ideology for both datasets. Interaction effects will be estimated and presented according to the practices outlined in Brambor, Clark, and Golder (2006) and Braumoeller (2004). In sum, then, because there are seven ideology conditions for four treatment effects across two datasets, we will be estimating 56 total conditional average treatment effects and making 28 comparisons.

Our null hypothesis is that there is no difference between effects estimated on MTurk and TESS, conditional on ideology. In order to reject the null, according to traditional thresholds for statistical significance, 5% or fewer of the interactive effects can be statistically indistinguishable. Two conditional average treatment effects will be deemed statistically distinguishable if their 95% confidence intervals do not overlap. For a similar visualization to what we will do, see Figure 2 of Wood and Porter (2018).

This study will thus provide a critical contribution to work on MTurk and its generalizability by comparing the treatment effects from a politically-oriented survey experiment – both on the full sample and on key ideological sub-groups – from a nationally representative probability sample of the American public vs. a parallel convenience sample on Mechanical Turk.

Works Cited

C. H. Achen and L. M. Bartels. It feels like we're thinking: The rationalizing voter and electoral democracy. In *Annual Meeting of the American Political Science Association, Philadelphia*, volume 30, 2006.

Joshua D. Angrist, and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2008.

L. M. Bartels. Beyond the running tally: Partisan bias in political perceptions. *Political behavior*, 24(2):117–150, 2002.

M. A. Baum and T. Groeling. Shot by the messenger: Partisan cues and public opinion regarding national security and war. *Political Behavior*, 31(2):157–186, 2009.

A.W. Bausch, J. R. Faria, and T. Zeitzoff. Warnings, terrorist threats and resilience: A laboratory experiment. *Conflict Management and Peace Science*, 30(5):433–451, 2013.

A. J. Berinsky. Assuming the costs of war: Events, elites, and american public support for military conflict. *The Journal of Politics,* 69(4):975–997, 2007.

A. J. Berinksy, Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political analysis*, *20*(3), 351-368.

T. Brambor, Clark, W. R., & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political analysis*, *14*(1), 63-82.

B. F. Braumoeller (2004). Hypothesis testing and multiplicative interaction terms. *International organization*, *58*(4), 807-820.

A. Coppock. (2018). Generalizing from survey experiments conducted on mechanical Turk: A replication approach. *Political Science Research and Methods*, 1-16.

N. Crawford. US Budgetary Costs of Wars Through 2016: $4.79 Trillion and Counting. http://watson.brown.edu/costsofwar/files/cow/imce/papers/2016/Costs%20of%20War%20through%202016%20FINAL%20final%20v2.pdf, 2016.

K. Dropp, J. D. Kertzer, and T. Zeitzoff. The less americans know about ukraine's location, the more they want us to intervene. *Washington Post*, 2014.

S. Feldman and J. Zaller. The political culture of ambivalence: Ideological responses to the welfare state. *American Journal of Political Science*, pages 268–307, 1992.

B. H. Friedman, J. Harper, and C. A. Preble. *Terrorizing Ourselves: Why US Counterterrorism Policy is Failing and How to Fix It*. Cato Institute, 2010.

M. Hetherington and E. Suhay. Authoritarianism, threat, and americans' support for the war on terror. *American Journal of Political Science*, 55(3):546–560, 2011.

A. M. Hoffman and W. Shelby. When the "laws of fear" do not apply: Effective counterterrorism and the sense of security from terrorism. i, 70(3):618–631, 2017.

L. Huddy, S. Feldman, C. Taber, and G. Lahav. Threat, anxiety, and support of antiterrorism policies. *American Journal of Political Science*, 49(3):593–608, 2005.

A. Lupia and M. D. McCubbins. *The democratic dilemma: Can citizens learn what they need to know?* Cambridge University Press, 1998.

M. L. McDermott and C. Panagopoulos. Be all that you can be: The electoral impact of military service as an information cue. *Political Research Quarterly*, 68(2):293–305, 2015.

J. E. Mueller. *Overblown: How Politicians and the Terrorism Industry Inflate National Security Threats, and Why We Believe Them*. Simon and Schuster, 2006.

J. E. Mueller and M. G. Stewart. *Chasing Ghosts: The Policing of Terrorism*. Oxford University Press, 2015.

K. J. Mullinix, Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, *2*(2), 109-138.

B. L. Nacos, Y. Bloch-Elkon, and R. Y. Shapiro. *Selling fear: Counterterrorism, the media, and public opinion*. University of Chicago Press, 2011.

B. Nyhan and J. Reifler. When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2):303–330, 2010.

B. I. Page and R. Y. Shapiro. Changes in americans' policy preferences, 1935–1979. *Public Opinion Quarterly*, 46(1):24–42, 1982.

C. R. Sunstein. Terrorism and probability neglect. *Journal of Risk and Uncertainty*, 26(2-3): 121–136, 2003.

J. Sides and K. Gross. Stereotypes of muslims and support for the war on terror. *The Journal of Politics*, 75(3):583–598, 2013.

P. M. Sniderman, R. A. Brody, and P. E. Tetlock. The role of heuristics in political reasoning: A theory sketch. *Reasoning and choice: explorations in political psychology*, pages 14–30, 1991.

C. Winkler. *In the name of terrorism: Presidents on political violence in the post-World War II era*. SUNY Press, 2006.

T.Wood and E. Porter. The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence. *Political Behavior*, 2018.

J. R. Zaller et al. *The nature and origins of mass opinion*. Cambridge university press, 1992.