

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

Daniel Silverman<sup>1</sup> *Carnegie Mellon University*

Daniel Kent<sup>2</sup> *The Ohio State University*

Christopher Gelpi<sup>3</sup> *The Ohio State University*

---

An American's yearly chance of being killed by a terrorist attack sits at about 1 in 3.5 million. Yet over 40% of the public have been consistently found to believe they or a family member are likely or very likely to be the victim of a terror attack. Can these heightened estimates of the risks posed by terrorism be brought closer to reality? With trillions of dollars spent on the War on Terror, this question is not just theoretically but practically important. To investigate, we use an experimental approach, testing whether participants update beliefs when presented with factual information about the risks posed by terrorism. In a pilot experiment conducted on Mechanical Turk in late 2017, we found that people did update their risk estimates to more accurate levels when given factual information, but that the information only had a strong and enduring effect when it was reinforced by a cue from a co-partisan political elite. In this paper, we first present these pilot results and then lay out a design for a revised and expanded study which will build on them. In particular, the revised design will: (1) conduct the survey experiment on a nationally representative sample, and (2) test the impact of a richer set of elite cues as treatments. Overall, this project helps show how the American public's overblown fear of terrorism – a particularly costly and sticky misperception in contemporary world politics – can best be mitigated and reduced to more realistic levels.

---

## Introduction

Since the attacks of September 11th, 2001, counter-terrorism has become the central goal that underpins American foreign security policy. According to one prominent estimate, the United States spent \$4.79 trillion from 2001 to 2016 on counter-terrorist policies including the wars in Afghanistan and Iraq, the conflicts in Syria and Pakistan, and American homeland security (Crawford, 2016).

In one sense, the centrality of terrorism for American foreign policy over the past 17 years seems appropriate in light of the public's ongoing grave concern over this issue. As recently

---

<sup>1</sup>dmsilver@andrew.cmu.edu

<sup>2</sup>kent.249@osu.edu

<sup>3</sup>gelpi.10@osu.edu

as June 2017, for example, Gallup found that 60% of the American public said it was “very” or “somewhat” likely that there would be acts of terrorism in the United States “over the next several weeks.” Even more dramatically, nearly 45% of respondents in the same poll stated that they were “very” or “somewhat” worried that they or someone in their family would become a victim of terrorism.<sup>4</sup> Moreover, this level of public anxiety over terrorism has changed very little since late 2001.

On the other hand, the public’s fear of terrorism persists despite the fact that the risk of death that Americans face from terrorism is remarkably low – even when taking into account the devastating but highly unusual attacks of 9/11. From September 11, 2001 through December 31, 2016, 3,081 Americans died in terrorist attacks, with 2,902 of those deaths occurring on 9/11 and 179 Americans dying over the following 15 years.<sup>5</sup> This amounts to an expenditure of \$1.55 billion for each American killed by terrorism from 2001 through 2016. This expenditure far outweighs American spending on any other cause of death. Moreover, according to U.S. Department of Defense statistics, more Americans have died in US military operations touted as combatting terrorism (6,950) than have actually been killed by terrorists.<sup>6</sup>

What could bring these fears back in line with reality? In fact, there has been a substantial body of research on the fear-generating properties of terrorism and the sources of the American public’s overreaction to the threat. For example, scholars have highlighted the nature of terrorism itself and peoples’ tendency to misjudge the risk of extremely low probability events (Sunstein, 2003); the propensity of humans to have an exaggerated response to risks over which they have little control (Slovic, 2000); and the fearful political rhetoric used by American leaders to justify military interventions in Afghanistan, Iraq, and elsewhere around the world (Gelpi et al., 2009). However, while there has been significant attention to the drivers of this American overreaction, there has been almost no empirical research examining whether

---

<sup>4</sup>See “Historical Trends: Terrorism,” *Gallup*, 2016. Available at <http://news.gallup.com/poll/4909/terrorism-united-states.aspx?version=print>

<sup>5</sup>See “American Deaths from Terrorist Attacks, 1995-2016,” *START Fact Sheet*, *University of Maryland*, 2016. Available at [http://www.start.umd.edu/pubs/START\\_AmericanTerrorismDeaths\\_FactSheet\\_Nov2017.pdf](http://www.start.umd.edu/pubs/START_AmericanTerrorismDeaths_FactSheet_Nov2017.pdf)

<sup>6</sup>See “U.S. Casualty Status,” *U.S. Department of Defense*, 2018. Available at <https://www.defense.gov/casualty.pdf>

it can be corrected.<sup>7</sup>

In this study, we investigate the extent to which public fears of terrorism can be mitigated based on the information that individuals are given regarding the risks of terrorism as well as the political rhetoric surrounding the delivery of such information. In so doing, we examine whether changes in the elite discussion of terrorism could alleviate public fears regarding terrorist attacks and provide a more realistically grounded foundation for U.S. security policy and grand strategy. Additionally, the study also speaks to literature on the correction of factual misperceptions about politics more broadly, where there is heated debate about the extent to which false political beliefs can be corrected and the best strategy to do so (Nyhan and Reifler, 2010; Wood and Porter, 2018). We contribute to this debate by testing the degree to which Americans update their beliefs about the risks of terrorism after being exposed to the facts, thus exploring the malleability of a crucial factual misperception in contemporary world politics while also bringing international security and political violence more squarely into this conversation.

In order to investigate these dynamics, we conducted a survey experiment that provided subjects with accurate information about the risk of death from terrorism in the context of other sources of mortal risk. We varied the context in which this information was delivered across two dimensions. First, we randomly assigned some subjects to conditions that acknowledged the understandable persistence of public fears despite the low levels of risk to investigate whether acknowledging these fears might cause individuals to be less defensive about updating incorrect factual beliefs. Second, we randomly assigned some subjects to receive a message from a co-partisan elected official acknowledging the low risk of death from terrorism

The acknowledgment condition did little to alleviate fears of a terrorist attack. However, endorsement of the information regarding public safety from terrorism had a robust and endur-

---

<sup>7</sup>For recent exceptions, see Bausch et al. (2013) and Hoffman and Shelby (2017). While these studies are valuable contributions, we aim to go beyond them by (1) testing the impact of political cues – in particular, co-partisan and cross-partisan political elites – and (2) fielding the survey on a nationally representative sample in order to gauge its potential to change minds among the general American population (as opposed to opt-in or convenience samples).

ing impact on alleviating public fears. Specifically, subjects who were recontacted two weeks after the study continued to state that they were significantly less concerned about the likelihood of a terrorist attack after receiving factual information about the risk of terrorism relative to other threats in combination with a co-partisan elite endorsement of these facts. These results imply that elite rhetoric surrounding the threat of terrorism is an important source of the problem of exaggerated public fears. But just as importantly, they imply that politicians have the opportunity to be part of the solution as well.

These results imply that elite rhetoric surrounding the threat of terrorism is an important source of the problem of exaggerated public fears. But our pilot study's focus on co-partisan cues only represents a first step in evaluating the role that elite rhetoric that may play in shaping public fears of terrorism. In our follow-up study – which we are preparing to field in the fall of 2018 – we will explore more fully how elite political rhetoric shapes fears of terrorism. Moreover, we will begin to explore the causal mechanisms that may play a key role in this belief correction process. Finally, our follow-up study will evaluate the impact of these treatments on a probability sample of the American public as well as a convenience sample recruited in the same manner as our pilot study.

Specifically, our follow-up study will compare the impact of three distinct types of elite cues in promoting the adoption of new information regarding the true human risks of terrorism for Americans: co-partisan cues, cross-partisan cues, and expert cues. Moreover, we explore three distinct but potentially complementary causal mechanisms identified by the literature on elite cues as potential sources of persuasive leverage: shared preferences with the cue sender, surprising cues that are contrary to the expected preferences of the cue sender, and policy expertise on the part of the cue sender. Finally, through support from Time-Sharing Experiments for the Social Sciences (TESS), we will field our follow-up study simultaneously on a probability sample of the US public as collected by the National Opinion Research Center (NORC) at the University of Chicago and on a parallel convenience sample recruited through Amazon Mechanical Turk in order to investigate any potential biases in average treatment effects as

estimated on commonly used convenience samples such as MTurk.

## **Background Literature**

Scholars have long noted that while terrorists rarely achieve their ultimate political goals, they can often inflict high costs on a target society by spreading a pervasive and disproportionate sense of fear among – that is, by *terrorizing* – the target public (Abrahms, 2006; Kydd and Walter, 2006). This fear, in turn, can have a number of painful consequences – provoking the target state to initiate costly foreign military interventions, allocate vast resources to homeland security over other national priorities, and undermine its own civil liberties and democratic accountability as it seeks to eradicate the threat.

Nowhere are these dynamics more apparent than in the U.S. over the past quarter century. With two overseas wars, vast intelligence operations, an extensive targeted killing campaign, and unprecedented domestic oversight mechanisms, the ongoing Global War on Terror is likely to go down as the most expensive defense campaign in American history, with a price tag of around \$5 trillion and rising (Crawford, 2016). One of the main drivers of this reaction is public perceptions: polls consistently reveal high levels of fear about terrorism among Americans, with roughly 40% of the country afraid that they or their families will be victimized and 70% viewing major attacks as likely in the near future (Mueller and Stewart, 2015, pp.80-88). Indeed, these fears have remained remarkably stable over time, with little if any decline since 9/11. Nevertheless, they belie the fact that the actual risk of terrorism in the U.S. is extremely low: under current conditions, the chance that an American is killed by a terror attack on U.S. soil is around 1 in 3.5 million, with under 90 deaths per year between 1970-2007. In contrast, the risks of death from other dangers like cancer (1 in 540), car accidents (1 in 8,000), drowning in a bathtub (1 in 950,000), and flying on a plane (1 in 2.9 million) are all far higher (Mueller, 2006). While these hazards kill far more people each year than terrorism, we expend far fewer resources trying to counter them.

This observation is not new – scholars of International Relations and U.S. foreign policy,

among other fields, have written at length about the American public's overreaction to terrorism. Scholars have likened the dynamic to falling into a trap set by the terrorists, or activating a series of powerful "antibodies" that kill the host as well as the virus (Lustick, 2006). Empirical research on the issue has studied how fundamental cognitive biases (Sunstein, 2003), religious stereotypes about Muslims (Sides and Gross, 2013), the nature of the contemporary mass media environment (Nacos et al., 2011), emotional responses about threatening stimuli (Huddy et al., 2005), and the prevailing political rhetoric around terrorism (Friedman et al., 2010) all contribute to this overreaction. Studies have also shown that this fearful response spans the ideological spectrum, afflicting liberals as much (if not more so than) than conservatives (Hetherington and Suhay, 2011). In sum, there has been substantial research on the fear-generating properties of terrorism and the misperceptions of Americans which have contributed to the overallocation of resources to this threat.

However, scholars have generally not explored a simple – yet profoundly important – question: can these misperceptions be mitigated? If so, how? Indeed, existing research on terrorism treats these reactions as relatively fixed, as something to be explained and understood rather than challenged. It is our aim to test, for the first time, the extent to which public perceptions surrounding the risk and threat of terrorism in the U.S. are in fact malleable to the facts and messages citizens receive about the threat. In so doing, we contribute to research on the consequences (and thus the utility) of terrorism by showing how they react to a different narrative about the threat. Additionally, we also make an important secondary contribution to research on belief correction. Indeed, from the perception of widespread voter fraud to beliefs that global warming is a hoax, there has in recent years been a surge of attention to the role of factual misperception in political life and the extent to which these misperceptions can be decreased. While some studies have found that corrections often "backfire" and deepen misperceptions among target ideological groups (Nyhan and Reifler, 2010), more recent studies have shown these beliefs to be surprisingly weak except in rare cases (Wood and Porter, 2018). The debate has increasingly matured from *whether* false beliefs can be corrected to

*which* strategies can most effectively do so, particularly on entrenched “hard cases” like Iraqi WMD which are especially resistant to facts. We contribute to this ongoing debate by testing the efficacy of efforts to correct the U.S. public’s inflated perception of the risk of death from terrorist attacks, a particularly sticky and costly belief in contemporary American politics. In this way, we aim to bring international security and conflict into this conversation around the correctability of inaccurate beliefs within political science and other related disciplines.

## **Pilot Research Design**

In order to study these issues, we fielded a survey experiment in which participants were randomly assigned varying forms of factual information about the risk of terrorism on U.S. soil. In doing so, we were able to test whether factual information can bring beliefs more in line with reality as well as how such factual information can be presented most effectively. Moreover, to test the persistence of any potential observed treatment effects, we fielded a follow-up survey two weeks later in which we asked participants the same post-treatment questions. In total, we recruited 800 participants in the first wave using Amazon’s Mechanical Turk platform and randomly assigned them to one of five conditions (one control and four treatment), yielding roughly 160 per group. In the second wave, we achieved a strong response rate of 84.25%, with 674 of the 800 responding to our invitations and completing the survey.

The survey experiment was designed as follows. First, all participants read a brief article about recent terrorist attacks that mirrored the general discussion of terrorism in the country and served to reinforce public concern. This follows the “belief-perseverance” approach employed in related studies, in which participants receive false information that is then corrected (Nyhan and Reifler, 2010). Participants who read just this initial priming article were in the control group. Then, participants assigned to the treatment groups also received one of four different vignettes with factual information about the actual risk of being the victim of a terrorist attack.

These treatments tested both the *general* efficacy of providing factual information as well

as whether *specific* types of corrections were more potent than others. In particular, the vignettes differed in two key ways – whether there was an acknowledgment of fear as understandable (*Acknowledgment*), and whether the correction was endorsed by a co-partisan elite (*Source*). The acknowledgment treatment was a recognition of the prevalence of fear-inducing discourse and imagery about terrorism in the U.S., signaling that holding inaccurate risk assessments does not call one’s intellect into question. The purpose of this treatment was to buffer the self-esteem of participants; updating one’s beliefs includes recognizing that one’s prior beliefs were wrong, which participants may seek to avoid in order to maintain self-esteem. The partisan cue treatment was an endorsement of the facts by a fictitious congressman who is a member of the individual’s political party (using a pre-test question on partisan identification). The purpose of this treatment was to test whether co-partisan elite rhetoric can serve to improve the accuracy of public beliefs. The four treatment conditions can be seen in the table below:

**Table 1: Treatment Conditions**

Acknowledgment Source	Yes	No
Yes	Acknowledgment & Source	No Acknowledgment & Source
No	Acknowledgment & No	No Acknowledgment & No

The treatments themselves took the form of brief vignettes, plus a table (see Table 2) that presented statistics about the risks of death via terrorism vs. other types of hazards to Americans. The vignettes thus built on previous research showing that visual aids and comparative reference points enhance the efficacy of corrective information (Sandman et al., 1993). In order to make the information as easily interpretable as possible, all of the vignettes included a brief explanation of the risk statistics after the table. Meanwhile, the acknowledgment condition included an opening with the aforementioned self-esteem buffer and the source cue

treatment closed the vignette with the statement from a fictitious co-partisan congressman endorsing low risk of death by terrorism.

**Table 2: Risk Statistics Table**

Threat	Country	Period	Annual Deaths	Annual Risk
Cancers	U.S.	2009	560,000	1 in 540
Traffic accidents	U.S.	2008	34,017	1 in 8,000
Homicide	U.S.	2006	14,180	1 in 22,000
Natural disasters	U.S.	1999-2008	629	1 in 480,000
Home appliances	U.S.	Yearly average	200	1 in 1,500,000
Commercial aviation	U.S.	1989-2007	103	1 in 2,900,000
<b>Terrorism</b>	<b>U.S.</b>	<b>1970-2007</b>	<b>87</b>	<b>1 in 3,500,000</b>
Lightning	U.S.	1999-2008	42	1 in 7,000,000

In order to measure participants' beliefs about the threat posed by terrorism, we used two types of outcome questions asked after the treatment – those capturing beliefs about (1) threat to the individual participant and (2) threat to the country as a whole. We believe that these are both important 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. The questions used are as follows:

*Individual Threat:*

- 1) How worried are you that you or someone in your family will become a victim of terrorism? Please tell me whether you are 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. For each one, please tell me whether it is very serious, fairly serious, not too serious, or not a threat at all. (*Terrorist attacks*)

*National Threat:*

- 1) Here is a list of potential threats to our country's national security over the next 10 years. For each one, please tell me whether it is very serious, fairly serious, not too serious, or not a threat at all. (*Terrorism*)
- 2) Here is a list of potential goals that the U.S. can pursue in its foreign policy. For each one, please tell me whether it is very important, fairly important, not too important, or not important at all. (*Preventing future acts of terrorism on U.S. soil*)

Principal-factors factor analysis shows that – across both survey waves – responses to the questions in each category hang closely together, with  $\alpha$ -scores above 0.7 for both pairs of items. Thus, we averaged participants’ responses to the personal threat questions as well as the national threat questions, and use these as our two primary DVs in the study.

## Pilot Model and Results

In each wave of the experiment, we estimated three OLS regression models to capture the average effect of each treatment and check for the presence of theoretically motivated interaction effects. The first model estimates the effect of each correction category:

$$DV = \beta_0 + \beta_1(\text{correction}) + \beta_2(\text{acknowledgment}) + \beta_3(\text{source}) + \beta_4(\text{full}) \quad (1)$$

We also estimate regression models which separately test for interactive effects between the treatment and both (1) political ideology as well as (2) right-wing authoritarianism (RWA). We include political ideology as an interaction term to test for the presence of “backfire effects” among conservatives (e.g., Nyhan and Reifler 2010), given that we would expect them to be especially skeptical of reassuring information about the threat of terrorism. RWA is also included as a key interaction, as it is deeply connected to aggressive responses to threat in general and to terrorism in particular (Hetherington and Suhay, 2011). These two models are estimated separately due to problems associated with parameter estimates in the presence of multicollinearity – in this case the correlation between ideology and right-wing authoritarianism. For each interaction term, we interact ideology or RWA with the elite source cue treatment, as it is found throughout to be the most robust and effective. The two models are as follows:

$$DV = \beta_0 + \beta_1(\text{correction}) + \beta_2(\text{acknowledgment}) + \beta_3(\text{source}) + \beta_4(\text{full}) \\ + \beta_5(\text{ideology}) + \beta_6(\text{ideology} \times \text{source}) \quad (2)$$

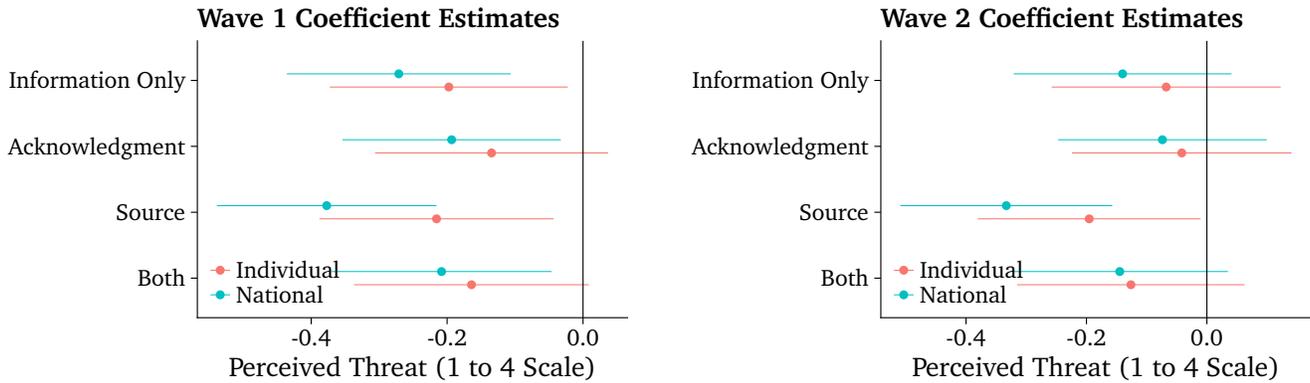
$$\begin{aligned}
DV = & \beta_0 + \beta_1(\text{correction}) + \beta_2(\text{acknowledgment}) + \beta_3(\text{source}) + \beta_4(\text{full}) \\
& + \beta_5(\text{RWA}) + \beta_6(\text{RWA} \times \text{source})
\end{aligned}
\tag{3}$$

Results from the baseline models for both the initial and follow-up surveys can be found in Figure 1. As can be seen, the treatment is often effective, meaning that the overblown fear of terrorism among the American public are indeed correctible after exposure to the facts. Yet, we can also see that the efficacy of the correction varies widely by treatment type. Specifically, the acknowledgement is the weakest approach, dampening the treatment effect to the point that it is no longer significant in most of the models. In contrast, the partisan elite cue is the most robust condition – it has the largest effect in wave one, and is the only significant effect that persists in the follow-up survey two weeks later. This provides evidence that politicians have the ability to reduce public fears of terrorism, and that endorsement by co-partisan elites is likely a necessary condition for the information to have a lasting effect.

Turning to the interaction effects, we find surprising conditioning effects with respect to national threat perceptions, as seen in Figure 2 (though we do not find evidence of these effects with individual threat perceptions). This is important in light of the debate over backfire and pushback against the facts by target ideological groups with strong priors in the other direction. Indeed, terrorism poses one of the stickiest cases of factual misperception in contemporary U.S. politics, and we would expect conservatives to cling to this misperception more strongly than liberals given the focus on Islamic extremism and terrorism in conservative media and ideology today (Bail, 2014; Isani and Silverman, 2016). Yet not only do we find no evidence for backfire effects among conservatives, we find that the treatment effects are actually stronger among conservatives than liberals. Moreover, we find the same trends around authoritarianism, with authoritarians updating their national threat perceptions more in response to the correction than non-authoritarians.

However, closer inspection reveals that this is most likely a function of a “floor effect”

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



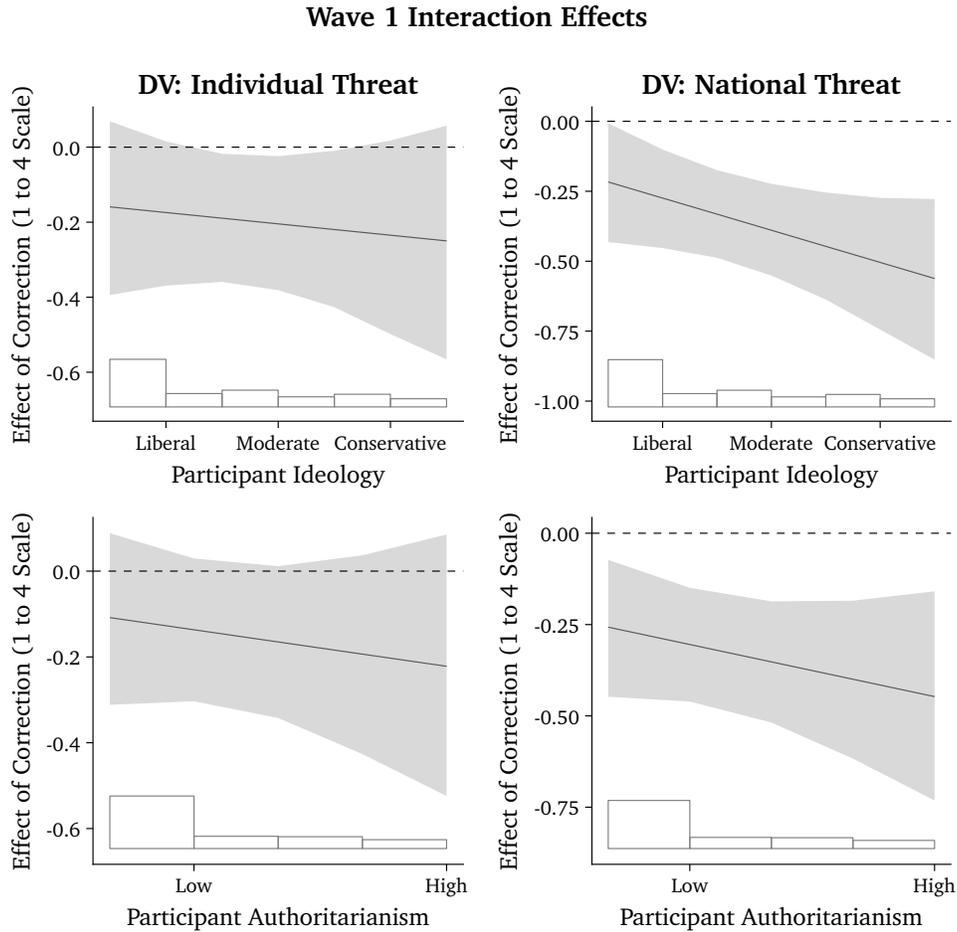
Treatment effects for the initial and follow-up surveys. The x-axis represents the magnitude of a treatment effect: the difference between the average response in the control and treatment groups. Responses are on a four-point scale where 1 corresponds to assigning terrorism the lowest threat level and 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.

where conservatives (and authoritarians) in the control group view terrorism as more of a threat than do their liberal counterparts. This produces a situation in which conservatives and authoritarians are more responsive to the facts because there is more room for them to move – that is, there is more space between their prior beliefs and the floor level of perceived threat below which most people will not drop (in this case, that terrorism is “not much of a threat”).<sup>8</sup> This highlights the need for analysts to pay attention to ceiling and floor effects when comparing the correctability of beliefs across groups, and fits with recent evidence that the backfire effect is rare (if not non-existent) in people’s reactions to factual information (Wood and Porter 2018).

Moreover, these effects hold across survey waves. When we run the interaction effect models on the follow-up survey, all interactive effects hold with only minor attenuation. This is not surprising in that the source treatment effect only dropped by a small amount across

<sup>8</sup>The lowest category (“not a threat at all”) was rarely chosen by respondents, making the next lowest (“not much of a threat”) the effective floor.

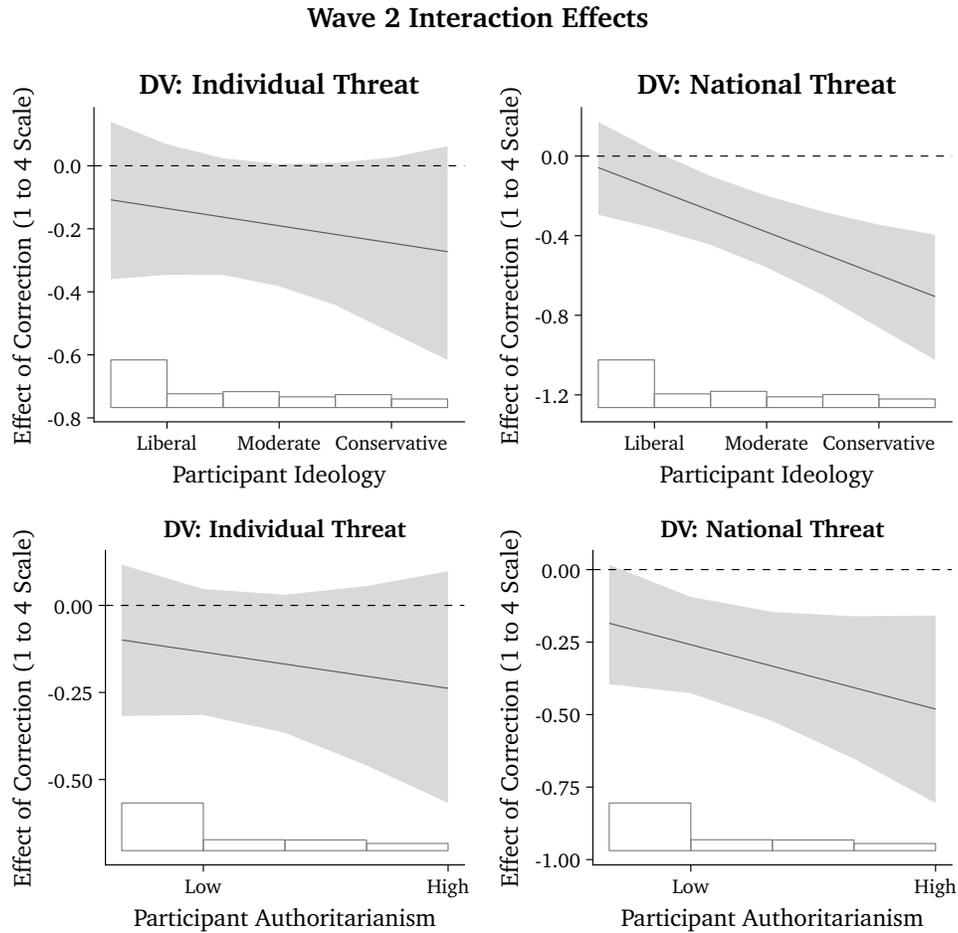
**Figure 2: Initial survey interaction effects**



Initial wave interaction effects of ideology and authoritarianism upon treatment effect. Each curves represent 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.

surveys. But it is further evidence that the underlying treatment effects across participants held over time.

**Figure 3: Follow-up survey interaction effects**



Follow-up survey interaction effects of ideology and authoritarianism upon treatment effect. Each curves represent 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.

## **A Revised Study**

Our pilot study results strongly suggest that exaggerated public fears of terrorism can be corrected, and that political elites can play a powerful role in helping the public form more realistic beliefs about the risks of terrorist activity. Specifically, the robustness of our co-partisan endorsed factual corrections indicates a potentially powerful impact for even modest changes in elite rhetoric on this issue. However, these results also point to several areas that could benefit from additional investigation in a follow up study. Most prominently, such a revised study could build on our finding that elite endorsements can enhance the impact of belief correction efforts – even around entrenched issues like the risks of terrorism to Americans – by exploring *which types of elite cues* are most effective in doing so.

## **Expansion of Treatments**

In order to develop a theoretically grounded set of expectations for our new study, 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 regard to 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 will embed co-partisan, cross-partisan, and non-partisan expert cues in our treatments. In particular, rather than always matching subjects with rhetoric from a fictional co-partisan politician as in our pilot study, we will randomly expose 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, the variation in the partisan identity of the cue-sender – with both co-partisan as well as cross-partisan messages – allows us to capture the effect of shared

interest with respondents, following the use of party labels to do so in a number of other studies. Moreover, the messages will include costly signals which contradict partisan stereotypes, such as Republican politicians advocating for less concern about terrorist attacks. And finally, the revised design will allow us to distinguish the persuasiveness of policy expertise from partisanship, by exploring the impact of cues sent from military generals with special expertise in national security issues as opposed to generic civilians politicians.

### **Additional Benefits of New Study**

In addition to a richer set of treatments, the revised study will also improve on our pilot effort in several other ways. First, as is often the case, our MTurk sample differed from a representative sample of the American public in several key ways. In particular, our sample skewed more Democratic, educated, and comfortable with numerical information 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 many rank-and-file Republican's from the effect of the treatment. We instead find some suggestive evidence that conservatives (and authoritarians, who follow a similar pattern) may be especially responsive to these facts when endorsed by trusted elites. However, these intriguing findings require further exploration with nationally representative sample. Consequently, we have obtained support for fielding this study on a national probability sample through Time-Sharing Experiments for the Social Sciences (TESS) and the National Opinion Research Center (NORC) at the University of Chicago.

Second, we plan to conduct our follow-up study simultaneously on the NORC and Amazon Mechanical Turk platforms in order to provide a more direct comparison of the treatment effects estimated on these two samples. Survey experiments 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 significantly limit the number and variety of experimental

studies of public opinion. By conducting 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 Amazon Mechanical Turk or through market research panels such as those maintained by Qualtrics.

Third, the design of the follow-up study also allows us to speak more clearly to the growing literature on belief correction and “backfire” in political psychology and behavior. As suggested earlier, this literature initially focused on the fairly general question of whether efforts to correct false beliefs “work” or are likely to backfire and entrench public misperceptions (e.g., Nyhan and Reifler 2010). However, the scholarly discussion has since evolved to a more nuanced effort to understand when, where, and why effective belief correction – or failure and even backfire – is most likely to occur (e.g., Flynn et al., 2017; Wood and Porter, 2018). In our pilot study, we did not find evidence of backfire effects, as participants successfully updated their beliefs in response to the facts, especially when those facts were endorsed by trusted elites. However, this may have been due both to the fact that our MTurk sample was not fully representative of the American public (in particular, its conservative elements), and that we only tested the impact of co-partisan cues – which are unlikely to provoke backlash. By (1) fielding a revised experiment on a high-quality, nationally representative sample of the American public, and (2) expanding our treatments to include cues from both co- and cross-partisan (as well as expert) elites, we are offering a much richer and more realistic test of the potential for backfire surrounding efforts to reduce entrenched misperceptions like the public’s inflated fears of terrorism in the United States.

## **Revised Research Design**

Our revised study employs a fully randomized design with four treatment groups and one control group. The control group only receives the news story vignette about terrorism (*Control*). All four of the treatment receive the control vignette and are subsequently provided with information on the actual risk that an individual American will be killed in a terrorism incidence

each year. The first of these four treatment groups receive only the vignette and the corrective information (*Correct\_Only*). The remaining three treatment groups receive the vignette, the corrective information, and an endorsement of the legitimacy of that corrective information by one of three fictitious elite actors: 1) A Republican representative (*Correct\_Republican*), 2) a Democratic representative (*Correct\_Democrat*), or 3) a high-ranking military officer (*Correct\_Military*).

This design allows us directly to compare the impact of co-partisan and cross-partisan cues. Moreover, it allows us to compare how the impact of these different types of cues may vary across Democratic, Republican, and Independent respondents.

### **Empirical Hypotheses in Revised Study**

At one level, the focus of our study is an empirical description of the influence of co-partisan, cross-partisan, and expert cues in supplementing factual corrections. The research on elite cues emphasizes shared interests, expertise, and costly signals as possible sources of persuasiveness (Lupia and McCubbins, 1998), and we do not have strong *a priori* expectations about which one of these effects will be the strongest. Moreover, we see no theoretical incompatibility between them such that they cannot jointly contribute to our results.

At the same time, we also aim to parse out the relative impact of these different mechanisms. In this endeavor, we are aided by the fact that there is some natural alignment between the mechanisms and the cues. 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 likely 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 teasing apart the mechanisms is thus difficult, we believe that examining carefully chosen comparisons and patterns of our treatment effects will allow us to do so. 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  $R \rightarrow R$ ) mixes surprise with shared interest. In contrast, a Democratic cue to a Democratic citizen (denoted  $D \rightarrow D$ ) simply taps into shared interest. Thus, the difference between the former and latter ( $R \rightarrow R - D \rightarrow D$ ) gives us a measure of any “surprise effect” net of the impact of shared interest between sender and receiver.

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 ( $R \rightarrow R$ )

represents the effect of shared interests – or trust – and surprise. Meanwhile, a military cue sent to a Republican citizen ( $M \rightarrow R$ ), we would argue, taps into all three factors: trust, surprise, and expertise. Indeed, given the overwhelming support for military leaders among Republican citizens and the fact that such leaders generally tend to affiliate with the Republican party (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 citizens. 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,  $M \rightarrow R - R \rightarrow R$ ) allows us to gauge the effect of policy expertise net of the effects of trust and of surprise.

#### *Model for Revised Study*

We formalize the above discussion of estimating these precise mechanism by drawing upon Angrist and Pischke’s (2009) 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.<sup>9</sup> 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:

---

<sup>9</sup>In the Mostly Harmless example, Angrist and Pischke look at employment in the bordering areas of New Jersey and Pennsylvania at two time periods, February and November. Because the minimum wage was raised in New Jersey during this time period, but not Pennsylvania they can compare the two like-units trends. So they compare how the two state employments varied over time, where Pennsylvania’s trends serve as the control group and New Jersey’s serve as the treatment group. The difference of these two trends serves as the difference-in-differences; the effect of interest.

$$Y_{irs}|r,s = \gamma_r + \beta_j D_s + \varepsilon_{irs} \quad (4)$$

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).  $\beta_j$  represents the treatment effect – the average change in fear – given a certain combination of  $s$  and  $r$ . And  $\gamma_r$  represents the baseline average fear in the control group for each partisan group. Relevant values for  $r$ ,  $s$ , and  $j$ <sup>10</sup> are:

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

Turning to our treatment effects, though there are more possible sender-receiver combinations for  $\beta_j$  than denoted, we only focus on the following three combinations due to the aforementioned theoretical concerns.  $\beta_1$  captures the treatment effect for a democratic recipient and sender.  $\beta_2$  captures the treatment effect for a republican recipient and sender.  $\beta_3$  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:

$$E(Y_{0irs}|r, s) = \gamma_r + \beta_j D_s \quad (5)$$

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

$$E(Y_{0irs}|r, s) = \gamma_r \quad (6)$$

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

---

<sup>10</sup>There are additional possible combinations for  $\beta_j$  which we do not consider and leave out to avoid confusion.

$$E(Y_{irs}|r = D, s = D) - E(Y_{irs}|r = D, s = 0) = \beta_1 \quad (7)$$

Next, we consider the ‘costly signal’ model, in which we look at the difference between a  $R \rightarrow R$  and  $D \rightarrow D$  model:

$$E(Y_{irs}|r = R, s = R) - E(Y_{irs}|r = R, s = 0) = \beta_2 \quad (8)$$

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

$$[E(Y_{irs}|r = R, s = R) - E(Y_{irs}|r = R, s = 0)] - \quad (9)$$

$$[E(Y_{irs}|r = D, s = D) - E(Y_{irs}|r = D, s = 0)] = \quad (10)$$

$$\beta_2 - \beta_1 \quad (11)$$

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  $R \rightarrow R$  from the effect of  $M \rightarrow R$ . Here we estimate the effect of a military endorsement upon republicans,  $\beta_3$  as:

$$E(Y_{irs}|r = R, s = M) - E(Y_{irs}|r = R, s = 0) = \beta_3 \quad (12)$$

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

$$[E(Y_{irs}|r = R, s = M) - E(Y_{irs}|r = R, s = 0)] - \quad (13)$$

$$[E(Y_{irs}|r = R, s = R) - E(Y_{irs}|r = R, s = 0)] = \quad (14)$$

$$\beta_3 - \beta_2 \quad (15)$$

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.

## **Conclusion**

This study explores whether the inflated perceptions about the threat of terrorism among the American public can be mitigated by exposure to the facts. Indeed, scholars have written at length about the reasons for this costly misperception in U.S. foreign policy – treating the public’s overreaction to terrorism as a given – but they have not examined the degree to which it can be corrected with factual information about the threat. In order to do so, we fielded a pilot survey experiment in which participants were exposed to data about the actual risks of terrorism on U.S. soil relative to other hazards, varying whether that data was accompanied by a self-esteem buffer and whether it was endorsed by a co-partisan political elite. Overall, we found that the correction was often effective – that is, citizens’ misperceptions about the threat were significantly reduced by exposure to the facts. Moreover, we found that while the acknowledgement of public fear did not work very well, the co-partisan elite endorsement proved very effective. Indeed, the elite cue had the strongest initial effect, and remained significant two weeks later in the follow-up survey. Finally, we found the elite treatment was effective even on conservatives and authoritarians, two groups expected to be strongly resistant to the corrective information. This study thus shows that exaggerated perceptions of the terror threat are surprisingly correctable with factual information, provided those facts are reinforced by statements from co-partisan elites.

These preliminary findings offer several key implications for our understanding of (counter) terrorism and contemporary U.S. foreign and national security policy. First, they suggest that much of the vast overreaction to terrorism in the U.S. over the past quarter century might have been avoided if Americans were given a more accurate picture of the threat and the risks that it poses to them. As noted earlier, the U.S. has lost some \$5 trillion dollars since 9/11 on a threat that claims very few lives per year relative to other dangers facing Americans. Additionally, the fearful response to terrorism in the U.S. has also cost the country thousands of lives, ballooned

the national debt, undermined U.S. civil liberties, encouraged risk-averse individual behavior (such as not flying), distracted the country from other more pressing issues, and possibly generated more malice and violence toward Americans than it has eliminated (Friedman et al., 2010; Mueller and Stewart, 2015) While these responses to terrorism stem from a number of different sources, this study reveals that one important source – the American public’s exaggerated perceptions of the threat – is not a fixed property of how people react to dreaded, rare, or unpredictable events, but something that can be substantially ameliorated by providing them with better and more accurate information about the phenomena they face.

At the same time, our research highlights the essential role of political elites in producing these effects. Indeed, affirmation by a co-partisan elite not only strengthened the treatment effect, but was the only way to make it “stick” two weeks later. It is thus trusted elites who hold the key to generating meaningful and lasting change in citizens’ perceptions of the threat from terrorism. This calls further attention to the issue of why they have not done so, and the political incentives for leaders to exaggerate or manipulate the terror threat. In fact, isolated statements can be found in which political elites urge Americans not to be paralyzed by fears of terrorism.<sup>11</sup> But this study raises the question: what conditions would have to take hold for there to be wider adoption of more sober discourse about the threat at the elite level? Additionally, focusing on elites also pushes us to reinterpret existing research on public reactions to terrorism (Huddy et al., 2005; Hetherington and Suhay, 2011; Kaltenthaler and Miller, 2017). Because these findings occur in an environment in which elites reinforce rather than undermine public fear, they may be less a function of innate psychological dynamics than the political context in which they take place. For example, the fact that conservatives and authoritarians are seen as reacting more strongly to the threat of terrorism may reflect the fact that conservative politicians have touted this threat most since 9/11, and that authoritarians are more likely to take cues from figures of authority – instead of representing the “intrinsic” preferences and

---

<sup>11</sup>For example, in his book *Why Courage Matters*, Sen. John McCain (R-AZ) proclaims: “Get on the damn elevator! Fly on the damn plane! Calculate the odds of being harmed by a terrorist! It’s still about as likely as being swept out to sea by a tidal wave. Suck it up, for crying out loud. You’re almost certainly going to be okay” (2004: 35-36).

predispositions of these two constituencies.

Despite these insights, our pilot study left many crucial questions still on the table, including which types of elite cues would enhance these corrections the most, and whether the results would generalize to a more representative sample (particularly among the target ideological group of conservatives).

We thus outlined the design of a new, revised study that would substantially build on these preliminary findings. In particular, this revised experiment will: (1) be run on a high-quality, nationally representative sample of the American public instead of a diverse convenience sample on MTurk, and (2) test the impact of a wider range of elite cues including co-partisan, cross-partisan, and expert non-partisan endorsements. This revised study will help illuminate which types of elite cues are most effective in mitigating terrorism misperceptions (and why), and will do so on a representative sample of the American public. This promises to substantially build on our pilot research by exploring not just whether America's overblown fears of terrorist attacks can be effectively mitigated, but what are the political messages and messengers best positioned to do so.

## References

- M. Abrahms. Why terrorism does not work. *International Security*, 31(2):42–78, 2006.
- 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.
- J. Angrist and J.-S. Pischke. *Mostly harmless econometrics: an empiricists guide*, 2009.
- C. A. Bail. *Terrified: How anti-Muslim fringe organizations became mainstream*. Princeton University Press, 2014.
- 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.
- 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.
- D. Flynn, B. Nyhan, and J. Reifler. The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, 38:127–150, 2017.
- 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.
- C. Gelpi, P. D. Feaver, and J. Reifler. *Paying the human costs of war: American public opinion and casualties in military conflicts*. Princeton University Press, 2009.
- 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. *Political Research Quarterly*, 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.
- M. Isani and D. Silverman. Foreign policy attitudes toward islamic actors: An experimental approach. *Political Research Quarterly*, 69(3):571–582, 2016.
- K. Kaltenthaler and W. Miller. Social psychology and fear of terrorism. *Working Paper*, 2017.
- A. H. Kydd and B. F. Walter. The strategies of terrorism. *International Security*, 31(1):49–80, 2006.
- A. Lupia and M. D. McCubbins. *The democratic dilemma: Can citizens learn what they need to know?* Cambridge University Press, 1998.
- I. Lustick. *Trapped in the War on Terror*. University of Pennsylvania Press, 2006.
- 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.
- 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.
- P. M. Sandman, P. M. Miller, B. B. Johnson, and N. D. Weinstein. Agency Communication, Community Outrage, and Perception of Risk: Three Simulation Experiments. *Risk Analysis*, 13(6):585–598, 1993.
- 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. Slovic. The perception of risk. risk, society, and policy series. London: Earthscan, 2000.
- 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. R. Sunstein. Terrorism and probability neglect. *Journal of Risk and Uncertainty*, 26(2-3): 121–136, 2003.
- 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. *Forthcoming at Political Behavior*, 2018.
- J. R. Zaller et al. *The nature and origins of mass opinion*. Cambridge university press, 1992.

## Appendix

An alternative means for estimating our quantities of interest which we have considered, and should provide the same results is to fit a single model:

$$Y_i = \beta_0 + \beta_1 r_D + \beta_2 D_D + \beta_3 r_D D_D + \beta_4 r_D + \beta_5 D_D + \beta_6 r_r D_D + \beta_7 D_M + \beta_8 r_R D_M + \varepsilon_i \quad (16)$$

Our hesitance in applying this model is simple. Though the model appears to fit a more traditional framework, correctly identifying our mechanisms of interest within it ends up being more nuanced and complicated than the aforementioned equations, where the data is subsetted by the recipient's party and then a series of regressions without interaction terms are estimated. However, the end result should be the same regardless of approach.

More specifically, if we fit a single regression, then we will isolate the mechanisms as:

- Trust:  $D \rightarrow D$

$$E(Y_i | r = D, s = 0) = \beta_0 + \beta_1 \quad (17)$$

$$E(Y_i | r = D, s = D) = \beta_0 + \beta_1 + \beta_2 + \beta_3 \quad (18)$$

$$E(Y_i | r = D, s = D) - E(Y_i | r = D, s = 0) = \beta_2 + \beta_3 \quad (19)$$

- Costly signal:  $R \rightarrow R - D \rightarrow D$

$$E(Y_i | r = R, s = 0) = \beta_0 + \beta_4 \quad (20)$$

$$E(Y_i | r = R, s = R) = \beta_0 + \beta_4 + \beta_5 + \beta_6 \quad (21)$$

$$E(Y_i | r = R, s = R) - E(Y_i | r = R, s = 0) = \beta_5 + \beta_6 \quad (22)$$

$$(R \rightarrow R) - (D \rightarrow D) = \quad (23)$$

$$(\beta_5 + \beta_6) - (\beta_2 + \beta_3) \quad (24)$$

- Expertise:  $M \rightarrow R - R \rightarrow R$

$$E(Y_i | r = R, s = R) = \beta_0 + \beta_4 + \beta_7 + \beta_8 \quad (25)$$

$$E(Y_i | r = R, s = M) - E(Y_i | r = R, s = R) = \quad (26)$$

$$(\beta_0 + \beta_4 + \beta_7 + \beta_8) - (\beta_0 + \beta_4 + \beta_5 + \beta_6) = \beta_7 + \beta_8 \quad (27)$$

$$(M \rightarrow R) - (R \rightarrow R) = \quad (28)$$

$$(\beta_7 + \beta_8) - (\beta_5 + \beta_6) \quad (29)$$