A Dispositional Theory of Reputation Costs
Supplementary Online Appendix

Ryan Brutger[*] and Joshua D. Kertzer[†]

Contents

[*]Assistant Professor, Department of Political Science, University of Pennsylvania. 208 Stiteler Hall, Philadelphia PA 19104. Email: brutger@sas.upenn.edu. Web: http://web.sas.upenn.edu/brutger/.

[†]Assistant Professor of Government, Harvard University. 1737 Cambridge St, Cambridge MA 02138. Email: jkertzer@gov.harvard.edu. Web: http://people.fas.harvard.edu/˜jkertzer/.

1   Dispositional measures

Militant assertiveness

- The best way to ensure world peace is through American military strength. [*Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree*]

- The use of military force only makes problems worse. [*Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree*]

- Going to war is unfortunate, but sometimes the only solution to international problems. [*Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree*]

International trust

- Generally speaking, would you say that the United States can trust other nations, or that the United States can't be too careful in dealing with other nations? [*The United States can trust other nations, The United States can't be too careful*]

2 Sampling methodology and sample characteristics

In the main text we present results from survey experiments fielded on two samples. The main analysis comes from a sample of 589 registered voters recruited as part of an omnibus survey experiment fielded in the spring of 2014 by Survey Sampling International (SSI). SSI panels employ an opt-in recruitment method, after which panel participants are randomly selected for survey invitations, using population targets rather than quotas to produce a nationally diverse sample of registered voters (see Table 1 for sample characteristics). The experiment was embedded in a larger, unrelated survey, and participants were unaware of the content of the survey when they chose to participate. Because of the recruitment technique, the sample is nationally diverse, although not a national probability sample.[3] The results with the free response analysis comes from a supplementary survey experiment fielded in the spring of 2013 on a sample of 1173 American adults recruited by Amazon Mechanical Turk (MTurk). As Table 2 shows, the MTurk sample skews considerably younger and more educated, consistent with other research on the demographic characteristics of MTurk survey participants (Huff and Tingley, 2015), although as we show in Appendix §4, we obtain the same results from both experiments.[4]

Table 1: Survey sample characteristics (SSI)

Gender	
Male	0.531
Female	0.469
Age	
18-29	0.129
30-44	0.270
45-64	0.392
65+	0.209
Education	
High school or below	0.239
Some college	0.401
College/university	0.244
Graduate/professional school	0.115
Race/Ethnicity	
Hispanic/Latino	0.112
Non-Hispanic White	0.716
African-American	0.100

---

[3]For other examples of recent political science articles using SSI samples, see Barker, Hurwitz, and Nelson (2008); Healy, Malhotra, and Mo (2010); Popp and Rudolph (2011); Kam (2012); Malhotra and Margalit (2010); Malhotra, Margalit, and Mo (2013); Berinsky, Margolis, and Sances (2014); Kertzer and Zeitzoff (2017).

[4]For other recent political science articles using MTurk samples, see Berinsky, Huber, and Lenz (2012); Chaudoin (2014); Huff and Kertzer (Forthcoming).

Table 2: Survey sample characteristics (MTurk)

Gender	
Male	0.568
Female	0.432
Age	
18-29	0.489
30-44	0.343
45-64	0.157
65+	0.011
Education	
High school or below	0.115
Some college	0.295
College/university	0.484
Graduate/professional school	0.107

3 Balance across treatments

Table 3 shows the composition of respondents within each treatment group for the main survey experiment. The randomization achieved a well-balanced sample across treatment conditions, with the sample proportions being quite balanced across each condition. Table 4 does the same for the MTurk supplementary study.

Table 3: Balance across treatments (SSI)

	Treatment Condition		
	Stay Out	Engage	Not Engage
Gender			
Male	0.545	0.507	0.533
Female	0.455	0.750	0.445
Education			
High school or below	0.237	0.221	0.253
Some college	0.397	0.412	0.397
College/university	0.263	0.243	0.227
Graduate/professional school	0.103	0.125	0.122
Age			
18-29	0.103	0.154	0.140
30-44	0.317	0.228	0.249
45-64	0.384	0.382	0.406
65+	0.196	0.235	0.205

Table 3 displays the distribution of population demographics across treatment groups in the SSI sample. Each value represents the proportion of the treatment group population composed by the corresponding demographic group.

Table 4: Balance across treatments (MTurk)

	Treatment Condition		
	Stay Out	Engage	Not Engage
Gender			
Male	0.562	0.557	0.586
Female	0.438	0.443	0.441
Education			
High school or below	0.102	0.098	0.143
Some college	0.313	0.295	0.275
College/university	0.480	0.492	0.479
Graduate/professional school	0.104	0.105	0.105
Age			
18-29	0.451	0.506	0.510
30-44	0.362	0.324	0.342
45-64	0.178	0.157	0.136
65+	0.009	0.013	0.011

Table 4 displays the distribution of population demographics across treatment groups in the MTurk sample. Each

value represents the proportion of the treatment group population composed by the corresponding demographic group.

4 Supplementary Experiment

Our supplementary study (fielded on Amazon Mechanical Turk), which is used primarily for the free response analysis, followed the same basic structure as our main experiment and began with the same introductory text that read:

> The following questions are about U.S. relations with other countries around the world. You will read about a situation our country has faced many times in the past and will probably face again. Different leaders have handled the situation in different ways. We will describe one approach U.S. leaders have taken, and ask whether you approve or disapprove.

In addition to asking respondents to complete free response questions in order to analyze how respondents talk and think about reputation, the supplementary study also differed from the main study by providing additional contextual information about the target of the scenario. The experiment informed respondents about the target's regime type and military capabilities (and, by extension, the expected costs of the intervention for the United States).

> A country sent its military to take over a territorial region in a neighboring country. The attacking country was led by a [democratically elected leader *or* unelected dictator]. The attacking country had a [strong military, so it would have taken a major effort, likely resulting in significant casualties *or* weak military, so it would have taken little effort, likely resulting in few casualties], for the United States to prevent the attacking country from gaining the contested territory.

To facilitate a direct comparison between the two studies, in our analysis of the free response questions in the main paper we excluded all respondents who were told casualties were present, which is consistent with the main experiment that holds outcomes and casualties constant.

Nonetheless, this additional context serves two important purposes. First, since our theoretical interest in the manuscript is how individuals turn to core dispositions to assess reputational considerations, it allows us to address a potential concern that the important role of dispositions in our main findings are simply due to an artificially spare vignette. Second, the additional context allows us to explore questions about experimental confounding (Dafoe, Zhang, and Caughey, 2014).

In regards to the first question, there are two countervailing considerations. On one hand, individuals who evaluate a foreign policy situation with very little information, such as introduced

by a relatively spare vignette, could be more likely to rely on their pre-existing beliefs, since there is relatively little information on which to base further analysis. This would potentially inflate the importance of dispositional characteristics compared to situational ones. On the other hand, individuals who receive more information may rely more heavily on preexisting beliefs to help simplify the situation, in which case increased information could make the effects of dispositional traits more pronounced. By comparing the effects of hawkishness between the supplementary and main experiments we can test whether the inclusion of additional contextual information inflates, deflates, or has no effect on the importance of dispositional traits.

A second concern that has been raised by Dafoe, Zhang, and Caughey (2014) and Tomz and Weeks (2013) is that failing to provide contextual information in vignettes may lead to unintended confounding. The main experiment, for example, doesn't mention the regime type or military capabilities of the target; if manipulating the president's policy choice changes beliefs about either unmentioned characteristic (perhaps, for example, hearing that the president decided to use force causes participants to assume the target is nondemocratic), our estimate of our quantities of interest may be biased. We can therefore leverage the contextual variation of our main and supplementary studies to evaluate whether the effects are significantly different when the amount of contextual information respondents received varies across studies.
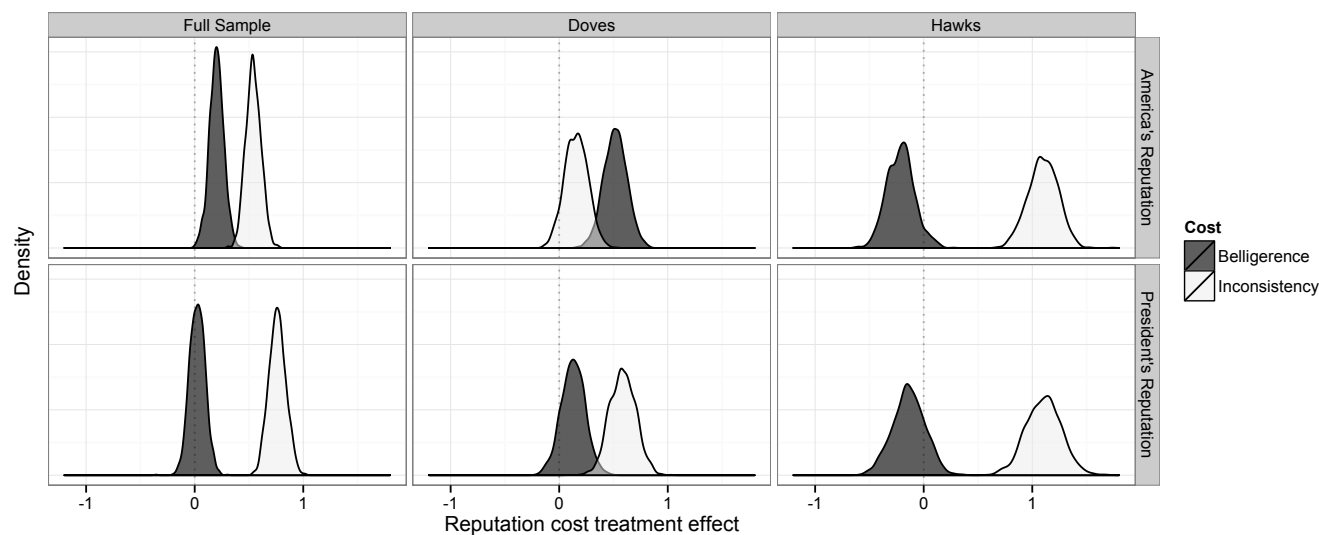
The results from our supplementary experiment are displayed in Figure 1, which can be compared to Figure 1 of the paper. Our primary substantive findings are all replicated in the supplementary experiment with more contextual detail. The density plots show that the results are remarkably stable across studies. In the supplementary study, we find that doves continue to believe that both belligerence and inconsistency generate reputation costs. We also find that hawks continue to believe that inconsistency generates reputation costs and that belligerence strengthens reputation.

Table 5 compares the effects of military assertiveness between the two experiments more formally. Models 1 and 4 estimate a set of regression models interacting the study treatments with an indicator variable referring to the supplementary experiment (i.e., the supplementary experiment versus the main experiment) and military assertiveness, finding no evidence of significant three-way interactions. Since three-way interactions are often hard to substantively interpret, however, Models 2-3 and 5-6 estimate separate regressions for doves and hawks, respectively, showing that the treatment effects in each subgroup of military assertiveness do not systematically differ between the two experiments, despite the contextual differences between the two vignettes.

Finally, although orthogonal to the questions above, in replicating the results we did find one subtle difference that may be of interest to some scholars of reputation. As displayed in Figure 1, we find that doves appear to differentiate slightly more between the country and the President's

reputation than hawks. Although we remain agnostic about the relative importance of leader versus country reputation (as discussed below), future research may wish to examine how different types of audiences perceive country and leader reputations.

Figure 1: Supplementary Experiment ATEs

Note: Figure 1 shows the reputation cost, where higher values indicate more damage to the reputation, for both the inconsistency and belligerence effects. These density plots show that hawks and doves believe reputations are damaged for different reasons, showing a strong connection between foreign policy orientation and beliefs about reputation.

8

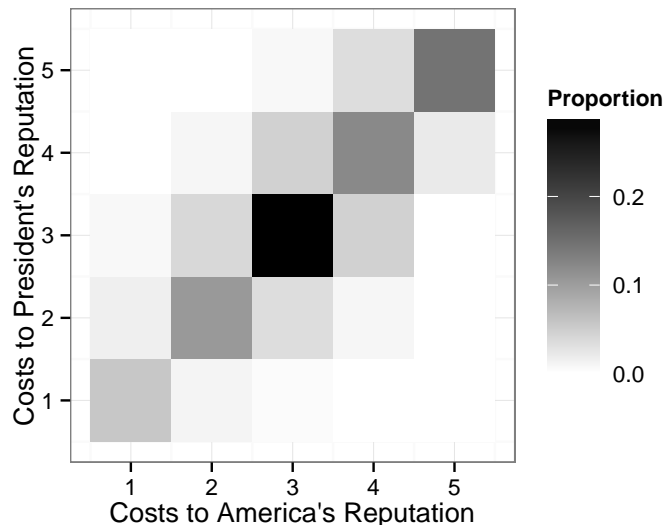Table 5: Comparing the role of military assertiveness across the two experiments

Reputational unit	America's reputation			President's reputation		
Sample	Full	Doves	Hawks	Full	Doves	Hawks
	(1)	(2)	(3)	(4)	(5)	(6)
Stay Out	−0.592***	−0.592***	0.413**	−0.328	−0.328	0.413*
	(0.222)	(0.220)	(0.203)	(0.240)	(0.235)	(0.224)
Not Engage	0.319	0.319	1.012***	0.481**	0.481**	1.095***
	(0.211)	(0.209)	(0.202)	(0.229)	(0.223)	(0.222)
Supplementary Experiment	−0.441**	−0.441**	−0.676***	−0.346*	−0.346*	−0.518**
	(0.182)	(0.180)	(0.193)	(0.197)	(0.193)	(0.214)
Mil. Assert	−0.267			−0.145		
	(0.232)			(0.252)		
Stay Out × Supplementary Experiment	0.075	0.075	−0.204	0.201	0.201	−0.275
	(0.246)	(0.244)	(0.248)	(0.267)	(0.261)	(0.274)
Not Engage × Supplementary Experiment	−0.159	−0.159	0.085	0.107	0.107	0.009
	(0.236)	(0.234)	(0.248)	(0.257)	(0.251)	(0.274)
Stay Out × Not Engage	1.006***			0.741**		
	(0.299)			(0.324)		
Not Engage × Mil. Assert	0.693**			0.614*		
	(0.290)			(0.314)		
Mil. Assert × Supplementary Experiment	−0.234			−0.171		
	(0.264)			(0.286)		
Stay Out × Supplementary Experiment × Mil. Assert	−0.279			−0.476		
	(0.347)			(0.377)		
Not Engage × Supplementary Experiment × Mil. Assert	0.244			−0.098		
	(0.341)			(0.369)		
Constant	3.220***	3.220***	2.952***	3.073***	3.073***	2.929***
	(0.165)	(0.164)	(0.165)	(0.179)	(0.175)	(0.182)
N	1,281	740	541	1,282	740	542
Adjusted R²	0.176	0.108	0.233	0.136	0.079	0.191

*p < .1; **p < .05; ***p < .01

The results show the effects of military assertiveness do not vary between the two experiments.

9

5 Leader and Country Reputations

Figure 2: Joint Distribution of the President and the Country's Reputations

Each cell of Figure 2 plots the proportion of respondents with the corresponding reputation costs for the President and America. Reputation costs are measured from 1 to 5, with higher values indicating greater damage to reputation. The plot shows that the vast majority of respondents, 72.1 percent, ascribe the same reputation cost to both the President and America, illustrated by the density of respondents along the diagonal.

Although our experiment was not designed to specifically test the relationship between leader- and country-specific reputations, we asked respondents to identify reputation costs for both. While the theory we present is agnostic about the relative important of one versus the other, we employ the data to examine how the two relate, which may be of interest to scholars of reputation. This analysis sheds light on the debates in the IR literature about the relevant unit of analysis for the study of reputation costs: when the public evaluates the President's handling of foreign policy crises, does it believe reputation costs are incurred by the President, but not the country, such that the national honor is relatively insulated from Presidential deeds? Or does the public see reputational costs borne by the President as spilling over and tarnishing the country's reputation as well?

To address these questions, we examine the correlation between respondents' beliefs about damage caused to the President's reputation and the country's reputation. The correlation between these two measures is extremely high ($r = 0.86$), suggesting that even if foreign observers attribute different reputations to countries and their leaders, domestic publics view them as intertwined. To illustrate the point visually, we plot the joint distribution of the President and America's reputation cost variables in Figure 2. Our data show that a preponderance of participants ascribe the same reputation cost to both the President and America, illustrated by the proportion of respondents

along the main diagonal in Figure 2. In total, 72.1% of respondents believe the President and the country face the same reputation cost and only 3.4% of participants believe the reputation costs of the President and country differ by more than 1 on a 5 point scale.

Our results thus suggest that when the President acts on the international stage, her domestic audience believes her actions not only affect her own reputation, but also that of the country at large. Thus, concerns about the President's reputation appear to spill over to its national counterpart, suggesting that domestic audiences do not view reputations as being entirely leader-specific, as the two types of reputation costs appear to be tethered together.

6 Text Analysis of Open-Ended Responses

In our experiment, each respondent was asked a separate question about their perception of America's reputation costs and the President's reputation costs. In the main STM analysis of our paper, we pooled these responses to allow the models to analyze as much text as possible. We also believe the decision is justified given the high correlation of perceived reputation costs for America's and the President's reputation costs. In Figures 3 and 4 we also show the significant differences in topic proportions when the data are not pooled across responses for America's and the President's reputation. Consistent with the main results of the paper, we find significant differences in the language doves and hawks use to talk about reputation costs. In both cases hawks are less likely to use language associated with anti-interventionism. In the Not Engage treatment we find that hawks are more likely to express concern about inconsistency, regardless of whether they are talking about the President's or America's reputation. These results give us greater confidence in the decision to pool responses and allow the models to analyze the the complete reputation responses of each respondent.

Figure 3: America's Reputation - Difference in Topic Proportion Between Doves and Hawks
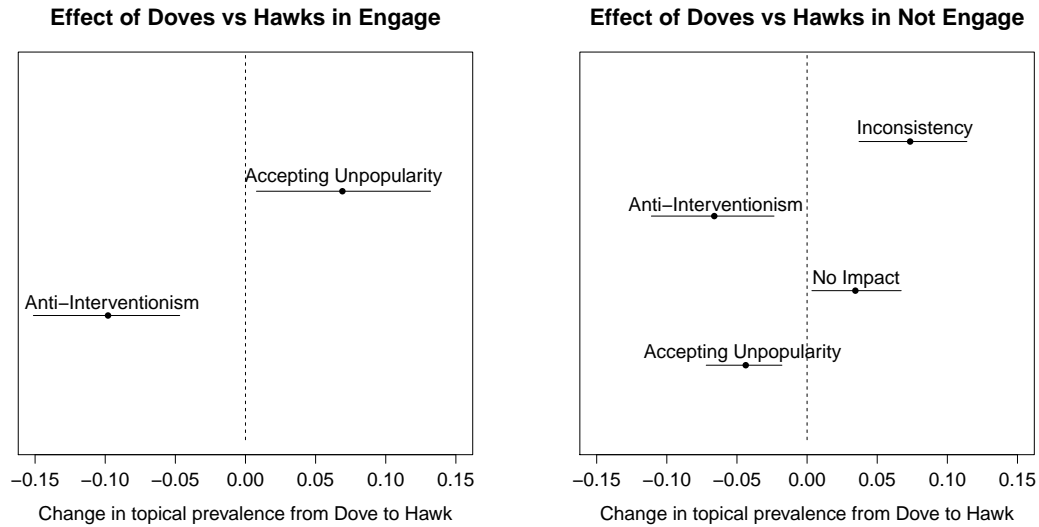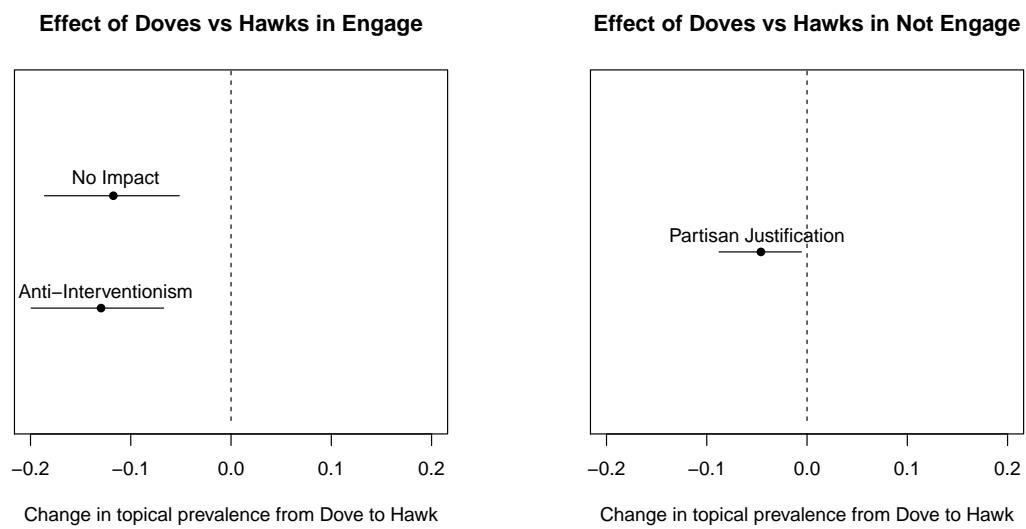

12

Figure 4: President's Reputation - Difference in Topic Proportion Between Doves and Hawks

Figures 3 and 4 show the change in the topic proportion when shifting from doves to hawks when responses about America's and the President's reputation are the respective dependent variables. Estimates are made by incorporating estimation uncertainty for the topic proportions into the uncertainty estimates (Roberts, Stewart, and Tingley, 2014). Only topics with a statistically significant difference between groups are displayed. An additional topic differed in prevalence between hawks and doves, but it lacked a clear substantive interpretation and is thus omitted from the main analysis, consistent with norms in the text analysis literature.

Figure 5 shows representative responses for each topic in the Engage treatment for both the America's reputation and President's reputation dependent variables. Comparing the text in the Anti-Interventionism topics, both topics reflect a clear sentiment that the United States and the President are too eager to become involved in international conflicts. As expected, the language in the America's Reputation response tends to specifically mention the US or America, whereas responses in the President's Reputation response tend not to include such references.

Figure 5: Representative Responses for Each Topic in the Engage Condition

Engage (America's Reputation)

Anti-Interventionism

It would make the US look like a bully, being one of the top superpowers we can't be big brother to everyone.

It reaffirms other countries' idea that the U.S. is an aggressor nation.

Americans are looked at the police of the world and we often get into conflicts that we should let the countries involved sort out.

Accepting Unpopularity

When you're the parent, there is always someone who won't like you.

Our reputation is what it is. There are countries that will not care what we have done and will not be swayed by any action. This is not to say we should bully everyone and not care, however we should take care of things without worrying so much what people think.

The[re] will always be people who don't like America.

Engage (President's Reputation)

Anti-Interventionism

I would assume I'm not the only person who feels the way I do about war and military use, so I assume other people would not be happy with how the situation was handled.

He will be looked at as a jerk. He will seem violent.

It makes him look like a war hawk who goes to war over everything.

No Impact

It just wouldn't cause much damage. People who voted and support him would be pleased and those that don't wouldn't be surprised.

It was just so [it] shouldn't hurt his or her reputation.

It depends what party he is a part of, but American presidents sending in US troops to foreign countries is nothing new, and maybe even expected.

Figure 6 shows representative responses for each topic in the Not Engage treatment for both the America's reputation and President's reputation dependent variables. Comparing the text in the Inconsistency topics demonstrates that the same concerns emerge, regardless of whether respondents are talking about America's or the President's reputation. In both cases respondents are primarily concerned about whether other countries will trust or believe the US or President in the future.

While figures 3, 4, 5, and 6 show many similarities across topics in the America's reputation and President's reputation responses, we also see that not all topics are consistently identified for each type of reputation. We also observe in 3 and 5 that a new topic "Strength and Justice" emerges for hawks in the engage condition. In part, these differences are a function of how structural topic models compile topics, where different "runs" of the analysis may yield slightly different results. That said, even where the new topic emerges or others lose significance, all the results show reasonable differences between doves and hawks and the differences are in line with expectations. The fact that hawks are more likely to use language that emphasizes America showing strength and enforcing justice reinforces our understanding of how hawks view foreign policy. Given that key topics are consistently identified across America's and the President's reputation responses, and that the different topics that emerge are in line with expectations, we are confident our pooled responses are not misrepresenting the data or overlooking key insights.

Figure 6: Representative Responses for Each Topic in the Not Engage Condition

Not Engage (America's Reputation)

Accepting unpopularity

America has a pretty bad reputation, I can't see this doing much harm.

The reputation is already pretty bad to a lot of people, it probably wouldn't make much of a difference.

The United States already has a pretty bad reputation and the conflict in this study is not serious enough to change it.

Inconsistency

No one will believe us next time we say we are going to do something

We need to do what we say we are going to do.

Again you have to back up what you say other wise no one will believe anything you say from that point on.

Anti-interventionism

I think the US is seen as a war bully, not attacking would not harm that image.

The US already is seen as a bully, the President [meddling] did nothing more but to pour gasoline on top of an out of control fire.

It's the [meddling] that's pissing off some countries. The way it was handled in this situation wasn't quite as severe, but it would still reaffirm the distrust and anger towards the us.

No impact

I don't think this one incident would greatly affect things.

I think this action will have a negligible effect on America's reputation because it is an isolated incident.

I don't think it would necessarily effect it

Not Engage (President's Reputation)

Partisan Justification

The opposing party would try to use this incident to portray the president in a negative light and would probably succeed.

The opposing party very vocally critical of whatever he did.

The other party would use this to hammer the POTUS again and again.

Figure 6 shows representative open-ended responses for each topic, selected for coherence among the ten most highly associated responses for each topic estimated using findThoughts from the STM package (Roberts, Stewart, Tingley et al., 2014)

17

7 Mediation Analysis

Figure 7: Full sample - Reputational Consequences of Belligerence and Inconsistency on Presidential Approval

(a) The Effect of the President's Reputation on Audience Costs

(b) The Effects of the Country's Reputation on Audience Costs
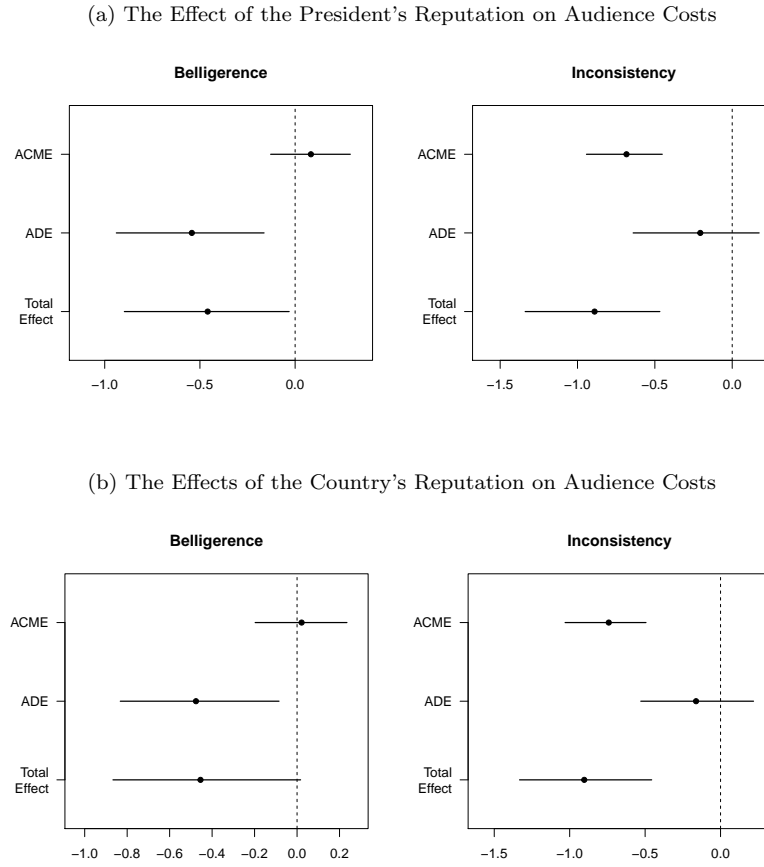

Figure 7 plot the Average Causal Mediation Effect (ACME) of the President's reputation and America's reputation respectively, the Direct Effect of Belligerence and Inconsistency treatments, and the Total Effect of the treatment and mediator on approval of the President's handling of the crisis. The President's approval is measured on a seven-point scale. The models control for pre-treatment convaraites of the respondents' education, income, gender, and ideology. The results are robust to including military assertiveness as a covariate. Mediation analysis conducted using the mediation package by Imai, Keele, Tingley et al. (2010)

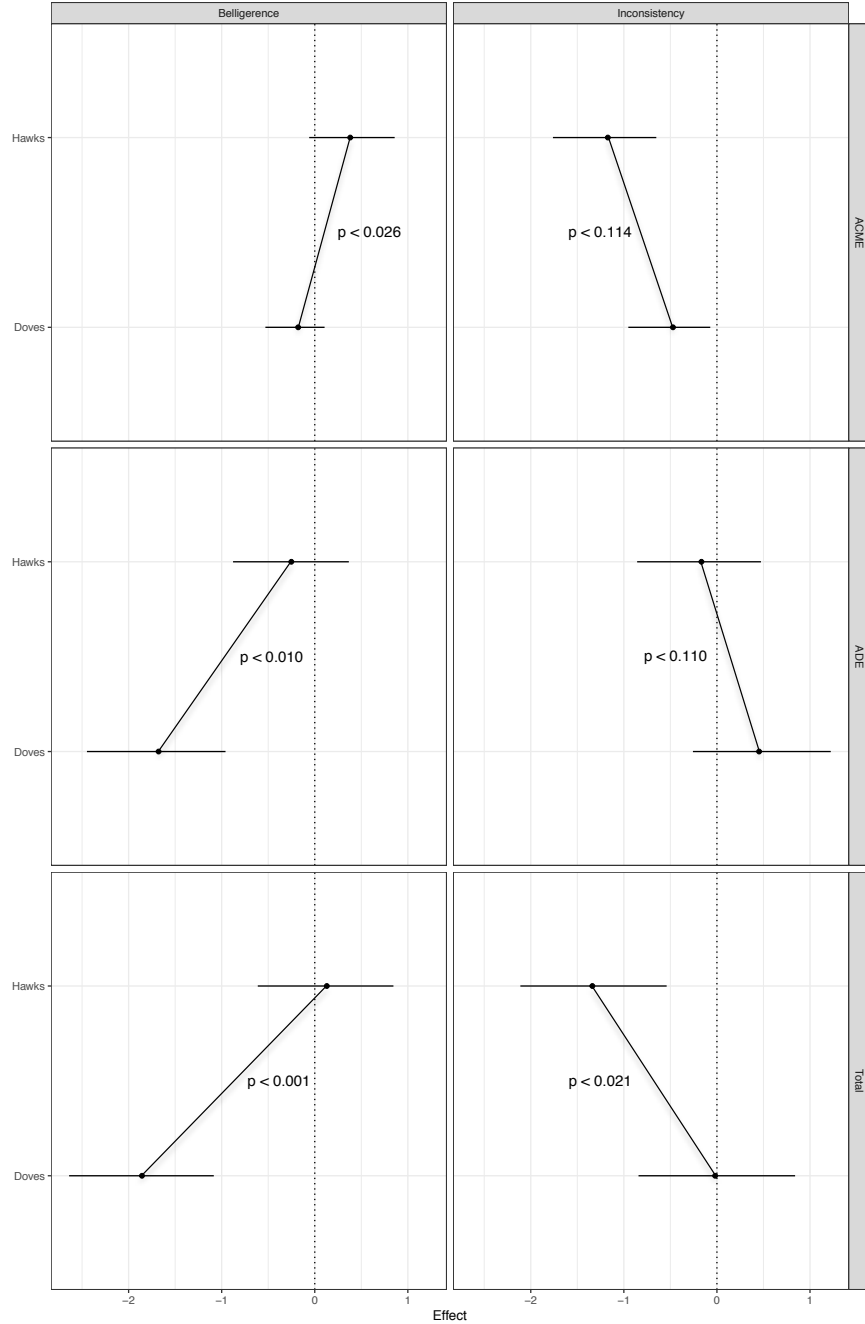Figure 8: President's Reputation: Moderated Mediation Results

Figure 8 plots the Average Causal Mediation Effects (ACME), the Average Direct Effects (ADE) and Total Effects from a series of nonparametric mediation models in which the effect of each treatment on approval of the President's handling of the crisis is mediated through reputation costs to the President, for hawks, and doves, respectively. Presidential approval is measured on a seven-point scale; hawks and doves are defined as those scoring in the bottom and top quartile of militant assertiveness respectively. The models control for pre-treatment measures of respondents' education, income, gender, and political ideology. The p-values for the ACME and ADEs come from formal tests of a moderated mediation model where the effect of the treatment on the reputational mediator varies between hawks and doves; the p-values for the total effects simply come from the interaction term between the treatment and the hawkishness dummy variable in a regression model of Presidential approval on the treatment, military assertiveness, their interaction, and the pretreatment covariates described above. Analysis conducted using the `mediation` package by (Imai, Keele, Tingley et al., 2010).

References

Barker, David C., Jon Hurwitz, and Traci L. Nelson. 2008. Of Crusades and Culture Wars: "Messianic" Militarism and Political Conflict in the United States. *Journal of Politics* 70 (2):307–322.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis* 20 (3):351–368.

Berinsky, Adam J., Michele F. Margolis, and Michael W. Sances. 2014. Separating the Shirkers from the Workers? *American Journal of Political Science* 58 (3):739–753.

Chaudoin, Stephen. 2014. Promises or Policies? An Experimental Analysis of International Agreements and Audience Reactions. *International Organization* 68 (1):235–256.

Dafoe, Allan, Baobao Zhang, and Devin Caughey. 2014. Confounding in Survey Experiments. Unpublished manuscript.

Healy, Andrew J., Neil Malhotra, and Cecilia Hyunjung Mo. 2010. Irrelevant events affect voters' evaluations of government performance. *Proceedings of the National Academy of Sciences of the United States of America* 107 (29):12,804–12,809.

Huff, Connor, and Joshua D. Kertzer. Forthcoming. How the Public Defines Terrorism. *American Journal of Political Science* .

Huff, Connor, and Dustin Tingley. 2015. "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics* 2 (3):1–12.

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2010. Causal Mediation Analysis Using R. In *Advances in Social Science Research Using R*, edited by H. D. Vinod. New York: Springer-Verlag.

Kam, Cindy D. 2012. Risk Attitudes and Political Participation. *American Journal of Political Science* 56 (4):817–836.

Kertzer, Joshua D., and Thomas Zeitzoff. 2017. A Bottom-Up Theory of Public Opinion about Foreign Policy. *American Journal of Political Science* 61 (3):543–558.

Malhotra, Neil, and Yotam Margalit. 2010. Short-Term Communication Effects or Longstanding Dispositions? The Public's Response to the Financial Crisis of 2008. *Journal of Politics* 72 (3):852–867.

Malhotra, Neil, Yotam Margalit, and Cecilia Mo. 2013. Economic Explanations for Opposition to Immigration: Distinguishing between Prevalence and Conditional Impact. *American Journal of Political Science* 57 (2):391–410.

Popp, Elizabeth, and Thomas J. Rudolph. 2011. A Tale of Two Ideologies: Explaining Public Support for Economic Interventions. *Journal of Politics* 73 (3):808–820.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2014. *stm: R Package for Structural Topic Models*. R package version 1.0.8.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science* 58 (4):1064–1082.

Tomz, Michael R., and Jessica L. P. Weeks. 2013. Public Opinion and the Democratic Peace. *American Political Science Review* 107 (4):849–865.