Backing Out but Backing In Audience Costs? A Replication of Levy et al. (2015)

Makito Takei
University of North Texas
Dept. of Political Science
1155 Union Circle #305340
Denton, Texas 76203–5017
MakitoTakei@unt.edu

Philip Paolino
University of North Texas
Dept. of Political Science
1155 Union Circle #305340
Denton, Texas 76203–5017
philip.paolino@unt.edu

October 24, 2022

Abstract

Levy et al. (2015) propose that inconsistency audience costs are caused in two ways: backing out of military commitment or backing into foreign conflicts. We replicate their experiment in July-August, 2021. Like many other studies, we find evidence for audience costs caused by backing out. However, our findings indicate that, unlike Levy et al., citizens are no less supportive of a leader who backs into a military conflict despite an initial commitment to stay out than one who behaves consistently. This study has significant theoretical and methodological implications. Theoretically, the importance of inconsistency in audience cost literature may be overstated or inconsistency audience costs may be bounded by temporal domains or contexts. Methodologically, our study emphasizes the necessity of replications because major findings can become conventional wisdom without such additional analysis.

Introduction

Since Fearon's (1994) seminal article, many scholars in International Relations have contended that inconsistency between words and deeds generates domestic audience costs. These occur when a leader backs down after the escalation of international disputes (Backing Out Audience Costs hereafter). Domestic audiences perceive that a leader who fails to follow through on a threat hurts a country's reputation or prestige (Fearon, 1994; Guisinger and Smith, 2002; Tomz, 2007) or is incompetent (Smith, 1998). Levy et al. (2015) introduce a more recent addition whereby domestic audiences also punish a leader who initially commits to staying out of a conflict but sends its forces in the end because this behavior is also inconsistent between words and deeds (Backing In Audience Costs hereafter). Using a modification of Tomz's (2007) experimental design, Levy and his colleagues find that leaders who "back into" foreign crises lose support from domestic audiences. They argue that, as with backing down from a threat, domestic audiences perceive that backing in constitutes reneging on a commitment and leads the public to punish the leader for an inconsistency that can damage the US's credibility.

This result is now the basis for several subsequent studies. For instance, an experimental study by Quek (2017) finds a similar result to Levy et al. In fact, his finding suggests that the domestic political costs caused by backing in are larger than those by backing out, which means that he provides stronger evidence for the significance of Backing In Audience Costs (or what he called "Type II Audience Costs") than Levy et al. Also, in her study on China's maritime disputes and how the country manages crises in the South China Sea, Luo (2021, 2022) applies the idea of Backing In Audience Costs and argues that international audience costs for China get larger when it escalates a conflict because China has committed to peaceful development.

The existence of Backing In costs would have several implications for how we understand audience costs. First, it would mean that audience costs are rooted in the domestic public's "distaste for inconsistency" (Levy, et. al 2015, 990). Second, audience costs apply not just

to how leaders use threat, but more generally how promises to intervene or to stay out can bind leaders to a course of action if they hope to avoid punishment from domestic audiences. Third, Backing In Audience Costs can be a foundation of credible assurances (Quek, 2017). While scholars in International Relations have extensively studied the credibility of threats, scholars and policymakers are also interested in how states commit to assurances credibly (e.g. Cebul, Dafoe and Monteiro, 2021; Christensen, 1992; Haynes and Yoder, 2020; Jervis, 1976; Kydd, 2000; Kydd and McManus, 2017; Montgomery, 2006). Just like the case of backing out, if domestic audiences are expected to punish a leader of a country who breaks the promise not to use force, other states should believe that leader's commitment.

We replicate the core treatments in Levy, et al. (2015). Using a larger sample, we do not find evidence for Backing In costs. While our study confirms previous research regarding the existence of audience costs that leaders incur from Backing Out, we do not find that leaders suffer a loss of approval from Backing In. Given the larger sample size and support for Backing Out audience costs, these results raise questions about the robustness of Backing In audience costs.

Research Design

To investigate the robustness of their findings, we replicated Levy et al. with a sample of U.S. adults recruited from July 31 to August 1, 2021, via Prolific, a participant recruitment, payment, and management crowdsourcing platform which has been increasingly utilized by studies in political science (e.g. Diamond, 2020; Tappin and Hewitt, 2021) because of its comparative advantage over other platforms such as Amazon Mechanical Turk (MTurk) (Palan and Schitter, 2018).[1] Following Levy et al., the United States is a potential intervener in our survey experiments. The use of the template allows us to compare the results by Levy et al. as well as the original experiment by Tomz (2007). The first treatment is whether the U.S. leader issues a threat to use force or not. Some respondents were told that the

---

[1]Prolific respondents earned $1.59 per completion as economic compensation.

Table 1: Experimental Design

Group	Threat	Use of Force	N (Ours)	N (Levy et al.)
Stay Out	No	No	442	369
Back In	No	Yes	525	372
Back Out	Yes	No	485	372
Go In	Yes	Yes	464	365

U.S. president "said that the United States would stay out of the conflict," while other respondents were told that the U.S. president announced that "if the attack continued, the United States would send military forces to help to push out the attacking country." The second treatment concerns the actual use of U.S. military force. Respondents were told either that the president did not send the military or that U.S. forces were deployed. In the end, there are four conditions: Stay Out, Back In, Back Out, and Go In. We compare the group of Stay Out and Back Out for a test for Backing Out Audience Costs. It is because the end result (i.e. non-use of force) is the same but the only difference is whether statements and actions are inconsistent or not. For the same reason, the comparison between Back In and Go In consists of a test for Backing In Audience Costs. Table 1 summarizes the experimental design of this study as well as the sample sizes of ours and Levy et al. To note, our sample sizes of all the treatment groups are larger than Levy et al.

One minor difference in research designs between Levy et al. and ours is that Levy et al. included New Information conditions where a U.S. leader received information from experts which can justify their decision, so theirs is 2 by 3 factorial design. We excluded this treatment from our study not only because this is not our interest but also in order to increase statistical power. We drop the observations of respondents who fail to pass manipulation checks.[2] As a result, the total sample size is 1,916. Descriptive statistics are

---

[2] For a manipulation check, we asked (1) whether the U.S. president initially said that the United States would stay out of the conflict, and (2) whether the U.S. president eventually sent U.S. military forces to help to push out the attacking country. We used the same manipulation check questions of Levy et al. While it appears that their analysis includes those who did not pass manipulation checks, we believe that they should be excluded because these respondents are inattentive. In Appendix, we show that the results are similar even if those who fail to pass manipulation checks are included in the sample.

Table 2: Domestic Political Consequences of Being Inconsistent (Outcome = No Force)

Study	% Who Approve "Stay Out"	% Who Approve "Back Out"	Difference in Approval
Levy et al.	51 [46, 57]	29 [24, 33]	-22 [-30, -16]
This Study	28 [24, 32]	20 [16, 24]	-8 [-13, -2]

Note: 95% confidence intervals in parentheses. Reported first differences may differ slightly due to rounding.

Table 3: Domestic Political Consequences of Being Inconsistent (Outcome = Use Force)

Study	% Who Approve "Go In"	% Who Approve "Back In"	Difference in Approval
Levy et al.	53 [47, 58]	41 [36, 46]	-12 [-19, -4]
This Study	58 [53, 62]	60 [56, 64]	2 [-4, 8]

Note: 95% confidence intervals in parentheses. Reported first differences may differ slightly due to rounding.

shown in Appendix.

After the two experimental manipulations, we ask, "Do you approve, disapprove, or neither approve nor disapprove of the way the U.S. president responded to the situation?" If a subject chooses Approve or Disapprove, the follow-up question asks whether they (dis)approve "Very strongly" or "Somewhat." If they neither approve nor disapprove, we ask, "Do you lean toward approving of the way the U.S. president handled the situation, lean toward disapproving, or don't lean either way?" Our dependent variable of approval is dichotomous taking a value of 1 if a respondent answers "Very strongly approve" or "Somewhat approve," otherwise 0.[3]

Results

Tables 2 and 3 compare the results of Levy et al. and ours. Following Levy et al., we report simple cross-tabulations. Regarding Backing Out Audience Costs, we find comparable

---

[3]As a robustness check, we also run ordinary least squares (OLS) regressions where the dependent variable is a 7-point Likert scale of approval. The results shown in Appendix are similar to those in the main text.

results to Levy et al. Comparing Back Out to Stay Out, there is an 8-percentage point drop in presidential approval with our sample (Chi^2= 7.829, p = 0.005). While these substantive significances are a little weaker than in Levy et al. (i.e. a 22-percentage-point drop), our results corroborate those from previous experiments that suggest that domestic audiences punish leaders who fail to follow through on an initial threat. Additionally, our result is closer to a 12-percentage point drop in Tomz (2007) and an 11-percentage point drop in Quek (2017).

On the other hand, the results on Backing In Audience Costs provide a significantly different picture. In contrast to Levy et al., our results indicate that citizens no more approve of a leader being consistent than one being inconsistent. In fact, there is a 2-percentage point *increase* in presidential approval from Go In to Back In in our experiment (Chi^2=0.428, p = 0.513), though this difference does not reach statistical significance.

Another interesting finding is that, compared to Levy et al., more respondents show approval of use of force than no use, regardless of (in)consistency. 59% of those who were told the U.S. used force (i.e. Back In and Go In combined) approved of the president's handling of the situation, while only 24% did not approve (Chi^2 = 242.434, p = 0.000). This 35-percentage point difference indicates that current U.S. citizens care more about whether the president (or government, etc.) uses force than inconsistency between words and actions.

Discussion

Since our experimental results suggest that the existence of Backing In Audience Costs is questionable, one might argue that the difference between Levy et al. and us stems from the samples. Notably, Levy et al. conducted their experiment on MTurk, while ours relies on Prolific participants. While a difference in samples may explain the different results, we address this possibility in several ways. First, a recent study finds that the data quality of Prolific is as high as MTurk (Peer et al., 2017). Therefore, it is unlikely that our null

results stem from the low quality of the data. Second, our results hold even if we control for many pre-treatment variables. Table 4 shows the results of logistic regressions including the control variables of age, male, white, Democrat, ideology, income, and veteran status. In the first model, compared to Stay Out, Back Out is negative and statistically significant. On the other hand, the coefficients of Back In are positive, but statistically insignificant, which is inconsistent with the Backing In Audience Cost argument. Third, it is important to note that our sample size in the treatment groups we test is larger than the sample size in the treatment groups in Levy et al. Therefore, it is unlikely that this null result stems from a lack of statistical power. Fourth, our results on backing down are consistent with Levy et al., and the chance that supporting some of Levy's findings but not others as a result of sampling variability is low. For these reasons, we believe that the difference in the samples is responsible for the different results.

Conclusion

In this study, we try to replicate one of the most important previous studies on two types of inconsistency audience costs, and we find evidence for only one type. Our results suggest that, while citizens are more likely to disapprove of a leader who backs down from a threat, they are no more disapproving of a leader's inconsistency when the leader uses force.

Our study has important theoretical and methodological implications for future debates. One of the theoretical implications of this null result is that the importance of inconsistency may be overstated in the literature on audience costs. Previous literature maintains that a leader's inconsistency between words and deeds can hurt a country's reputation (Fearon, 1994; Guisinger and Smith, 2002; Tomz, 2007) or signals his/her incompetence (Smith, 1998; Levendusky and Horowitz, 2012) and thus generate audience costs. Many people believe saying one thing and doing another is meaningful, but our study suggests that inconsistency costs may be not as strong as we think. Our results suggest that U.S. citizens may be more

Table 4: Logitic Regression with Control Variables

	Model 1	Model 2
	Stay Out vs Back Out	Go In vs Back In
Back Out	-0.505***	
	(0.167)	
Back In		0.00882
		(0.141)
Age	0.0694	-0.00540
	(0.0869)	(0.0825)
Male	0.145	0.401**
	(0.208)	(0.195)
White	0.163	-0.250
	(0.191)	(0.163)
Democrat	-0.219	0.306*
	(0.198)	(0.174)
Ideology	0.0693	-0.283***
	(0.0644)	(0.0605)
Income	-0.0147	0.0637**
	(0.0383)	(0.0320)
Veteran	-0.208	0.670
	(0.674)	(0.686)
Constant	-1.354***	1.564***
	(0.404)	(0.373)
Observations	830	886
Log-Likelihood	-448.9	-577.8

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

concerned about whether the president uses force than whether his behavior is consistent.[4] So as some scholars argue (Chaudoin, 2014; Snyder and Borghard, 2011), policy substance may be more consequential than policy (in)consistency.

Also, the results of our experiments suggest that it is worthwhile investigating when and in what context audience costs can emerge. They indicate that Backing In Audience Costs do not always exist across time and context and our ability to find evidence for it in survey experiments depends on what is happening in the world. International contexts

---

[4]In our sample, 59% of those who were told the U.S. used force (i.e. Back In and Go In combined) approved of the president's handling of the situation, while only 24% of those not (i.e. Stay Out and Back Out combined) approve (Chi^2 = 242.434, p = 0.000). This 35-percentage point difference indicates that current U.S. citizens care more about whether the president (or government, etc.) uses force than inconsistency between words and actions.

are significantly different in the timing of the survey experiments between Levy et al. and ours. Levy et al. started to conduct their experiment one month after the initiation of the Second Libyan Civil War (i.e. June-August 2014). On the other hand, we fielded our survey experiment in the midst of the U.S. withdrawal from Afghanistan in 2021. This difference in international contexts may have affected how U.S. citizens reacted to backing in. For instance, while our results on Backing Out Audience Costs are consistent with Levy et al., the effect size is smaller. In addition, (Quek, 2017, p.1442) finds a nonbelligerent dividend for staying out compared to going in, which is consistent with the belligerent audience cost argument by Kertzer and Brutger (2016). On the contrary, we do not find evidence for a (non)belligerent dividend. Compared to "Stay Out," approvals in the "Go In" condition *increase* by a 30-percentage point (Chi^2 = 82.663, p= 0.000). It is unlikely that the difference between Quek and ours may stem from those in the samples because our sample has a large component of liberals who should usually prefer peace to war–and remains even after we control for ideology and party identification. This result may reflect the Afghanistan situation where not fighting was seen as the poorer choice; although, we do not have any evidence to substantiate this possibility. Rather, our study opens the door for yet more research on the conditions under which both backing in and backing out audience costs can be incurred.

Methodologically, our study demonstrates why replication is necessary for the accumulation of scientific knowledge. Political scientists and scholars in International Relations usually put more importance on the introduction of new ideas or hypotheses based on pre-existing studies. On the other hand, a mere replication of previous studies tends to be underappreciated in the field. However, an experimental finding could become conventional wisdom quickly without replication. This study suggests that, as with some journals such as *Journal of Experimental Political Science* and *Research and Politics*, more and more journals in Political Science and International Relations should publish a category of replication manuscripts.

References

Cebul, Matthew D., Allan Dafoe and Nuno P. Monteiro. 2021. "Coercion and the Credibility of Assurances." *The Journal of Politics* 83(3):975–991.

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

Christensen, Thomas J. 1992. "Threats, Assurances and the Last Chance for Peace." *International Security* 17(1):122–154.

Diamond, Emily P. 2020. "The Influence of Identity Salience on Framing Effectiveness: An Experiment." *Political Psychology* 41(6):1133–1150.

Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88(3):577–592.

Guisinger, Alexandra and Alastair Smith. 2002. "Honest threats: The interaction of reputation and political institutions in international crises." *Journal of Conflict Resolution* 46(2):175–200.

Haynes, Kyle and Brandon K. Yoder. 2020. "Offsetting Uncertainty: Reassurance with Two-Sided Incomplete Information." *American Journal of Political Science* 64(1):38–51.

Jervis, Robert. 1976. *Perception and Misperception in International Politics.* Princeton University Press.

Kertzer, Joshua D. and Ryan Brutger. 2016. "Decomposing Audience Costs: Bringing the Audience Back into Audience Cost Theory." *American Journal of Political Science* 60(1):234–249.

Kydd, Andrew H. 2000. "Trust, Reassurance and Cooperation." *International Organization* 54(2):325–357.

Kydd, Andrew H. and Roseanne W. McManus. 2017. "Threats and Assurances in Crisis Bargaining." *Journal of Conflict Resolution* 61(2):325–348.

Levendusky, Matthew S. and Michael C. Horowitz. 2012. "When Backing Down Is the Right Decision: Partisanship, New Information, and Audience Costs." *Journal of Politics* 74(2):323–338.

Levy, Jack S., Michael K. McKoy, Paul Poast and Geoffrey P.R. Wallace. 2015. "Backing Out or Backing In? Commitment and Consistency in Audience Costs Theory." *American Journal of Political Science* 59(4):988–1001.

Luo, Shuxian. 2021. Take It to the Sea: Nonmilitary Actors in China's Maritime Disputes and Crisis Management. PhD dissertation. Johns Hopkins University.

Luo, Shuxian. 2022. "The Rising Power's Audiences and Cost Trade-offs: Explaining China's Escalation and Deescalation in Maritime Disputes." *Asian Security* 18(2):172–199.

Montgomery, Evan Braden. 2006. "Breaking Out of the Security Dilemma: Realism, Reassurance and the Problem of Uncertainty." *International Security* 31(2):151–185.

Palan, Stefan and Christian Schitter. 2018. "Prolific.ac—A Subject Pool for Online Experiments." *Journal of Behavioral and Experimental Finance* 17:22–27.

Peer, Eyal, Laura Brandimarte, Sonam Samat and Alessandro Acquisti. 2017. "Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research." *Journal of Experimental Social Psychology* 70:153–163.

Quek, Kai. 2017. "Type II Audience Costs." *The Journal of Politics* 79(4):1438–1443.

Smith, Alastair. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92(3):623–638.

Snyder, Jack and Erica D. Borghard. 2011. "The Cost of Empty Threats: A Penny, not a Pound." *American Political Science Review* 105(3):437–456.

Tappin, Ben M. and Luke B. Hewitt. 2021. "Estimating the Persistence of Party Cue Influence in a Panel Survey Experiment." *Journal of Experimental Political Science* pp. 1–12.

Tomz, Michael. 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61(4):821–840.

Appendix

Instrument for the Survey Experiment

GENERAL OPENING PROMPT PROVIDED TO ALL RESPONDENTS
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.

GENERAL INVASION SCENARIO PROVIDED TO ALL RESPONDENTS
A foreign country sent its military to take over a neighboring country.

TREATMENT #1 INITIAL THREAT

No Threat
The U.S. president said the United States would stay out of the conflict.

Threat
The U.S. president said that if the attack continued, the United States would send military forces to help to push out the attacking country.

QUESTION ON INITIAL POLICY
Do you approve, disapprove, or neither approve nor disapprove of the way the U.S. president responded to the situation?
• Approve
• Disapprove
• Neither approve nor disapprove

[Ask if answer "approve"] Do you approve very strongly, or only somewhat?
• Very strongly
• Somewhat

[Ask if answer "disapprove"] Do you disapprove very strongly, or only somewhat?
• Very strongly
• Somewhat

[Ask if answer "neither approve nor disapprove"] Do you lean toward approving of the way the U.S. president handled the situation, lean toward disapproving, or don't lean either way?
• Lean toward approving
• Lean toward disapproving
• Don't lean either way

TREATMENT #2 SUBSEQUENT USE OF FORCE

11

No Use of Force
The attacking country continued to invade. In the end, the president did not send U.S. military forces to help to push out the attacking country.

Use of Force
The attacking country continued to invade. The attacking country continued to invade. In the end, the president sent U.S. military forces to help to push out the attacking country.

QUESTION ON SUBSEQUENT POLICY
Do you approve, disapprove, or neither approve nor disapprove of the way the U.S. president handled the situation?
• Approve
• Disapprove
• Neither approve nor disapprove

[Ask if answer "approve"] Do you approve very strongly, or only somewhat?
• Very strongly
• Somewhat

[Ask if answer "disapprove"] Do you disapprove very strongly, or only somewhat?
• Very strongly
• Somewhat

[Ask if answer "neither approve nor disapprove"] Do you lean toward approving of the way the U.S. president handled the situation, lean toward disapproving, or don't lean either way?
• Lean toward approving
• Lean toward disapproving
• Don't lean either way

Variable Codings

Condition: Indicator for experimental group, where 1 = Stay out; 2 = Back in; 3 = Back out; 4 = Go in.

Male 1 = Male; 0 = Female.

Age: 1 = 18-24 years; 2 = 25-34 years; 3 = 35-44 years; 4 = 45-54 years; 5 = 55-64 years; 6 = 65 years or older.

White: 1 = White, 0 = Otherwise.

Democrat: 1 = Democrat, 0 = Otherwise.

Ideology: 1 = Extremely Conservative, 2 = Conservative, 3 = Slightly Conservative, 4 = Moderate, Middle of the Road or Don't Know, 5 =Slightly Liberal, 6 =Liberal, 7 =Extremely Liberal.

Income: 1 =Less than $25,000, 2 =$25,000 - $49,999, 3 =$50,000 -$74,999, 4 =$75,000 - $99,999, 5 =$100,000 - $124,999, 6 =$125,000 - $149,999, 7 =$150,000 - $174,999, 8 =$175,000 - $199,999, 9 =More than $200,000

Veteran: Respondent's military background, where 1 = Some form of military experience (active duty in armed forces, previous active duty, or trained for national reserves or guard); 0 = Otherwise.

Descriptive Statistics

Table 5: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Approval	1,916	.4185804	.4934551	0	1
Stay Out	1,916	.2306889	.4213837	0	1
Back In	1,916	.2740084	.4461296	0	1
Back Out	1,916	.2531315	.4349192	0	1
Go In	1,916	.2421712	.4285092	0	1
Age	1,915	1.559791	.9016135	1	6
Male	1,862	.1954887	.396683	0	1
White	1,904	.7179622	.4501099	0	1
Democrat	1,902	.615142	.4866897	0	1
Ideology	1,906	5.304302	1.490038	1	7
Income	1,789	3.717719	2.230865	1	9
Veteran	1,915	.0146214	.1200631	0	1

Balance Checks

Through a series of t-test, we find 2 variables (i.e. Age and Male) where the difference in means show statistical significance in our sample. To address this problem, we also run the logistic regression models with these sets of control variables shown in the main text. Our main results hold even accounting for the control variables.

Table 6: Difference in Means (with T-test P-values) between Covariate Mean for Respondents in Stay Out and....

Covariants	Back In	Back Out	Go In
Age	0.215***	0.223***	0.130**
Male	0.052*	0.058**	0.040
White	-0.001	0.006	-0.026
Democrat	0.005	0.013	0.049
Ideology	-0.037	-0.034	0.016
Income	-0.064	0.050	0.091
Veteran	0.007	0.012	0.012

OLS Regressions

Table 7 shows the results of ordinary least squares (OLS) regressions without and with the control variables where the dependent variable is a 7-point Likert scale of approval. The control variables are age, male, white, democrat, ideology, income, and veteran status. Across all the models, compared to Stay Out, Back Out is negative and statistically significant. On the other hand, the coefficients of Back In are positive and statistically significant in one out of the two models, but this effect is no longer significant once we include controls.

Table 7: Regression Analysis

	(Model 1) Stay Out vs Back Out	(Model 2) Stay Out vs Back Out	(Model 3) Go In vs Back In	(Model 4) Go In vs Back In
Back Out	-0.631***	-0.687***		
	(0.131)	(0.138)		
Back In			0.252**	0.168
			(0.121)	(0.125)
Age		0.0943		0.0285
		(0.0754)		(0.0729)
Male		0.0591		0.207
		(0.178)		(0.167)
White		0.345**		-0.152
		(0.155)		(0.144)
Democrat		-0.374**		0.402***
		(0.167)		(0.152)
Ideology		0.169***		-0.282***
		(0.0540)		(0.0504)
Income		-0.0251		0.0549*
		(0.0315)		(0.0281)
Veteran		-0.516		0.0891
		(0.555)		(0.546)
Constant	3.699***	2.753***	4.845***	5.951***
	(0.0945)	(0.339)	(0.0878)	(0.319)
Observations	927	830	989	886
R^2	0.025	0.053	0.004	0.051

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Weighted Analyses

We recognize that our sample is not representative of the US population, so we also did weighted analysis. For the weighted analysis, we used the anesrake package in R to do iterative proportional fitting of the data on variables for sex, age, race, and income. The raw data are fairly highly skewed from the population, so the raking procedure had difficulty generating weights, but ultimately did converge. The weights brought the sample into reasonably close population levels for sex and income, but less well for age and race. The weighted results, however, do confirm the results from the unweighted data, producing a similar effect for Back Out and even indicating a larger and statistically significant positive effect upon approval for Back In. But given the difficulty in generating the weights, we are not confident in saying anything more than that the effect of Back In is no different than Go In.

Table 8: Weighted Regression Analysis

Variable	Stay Out vs. Back Out		Go In vs. Back In	
	Coefficient	Std. Error	Coefficient	Std. Error
Intercept	2.373	0.368	4.143	0.318
Back Out	-0.490	0.144		
Back In			0.376	0.132
Male	-0.087	0.146	0.470	0.142
White	-0.529	0.222	-0.281	0.211
Age	0.303	0.091	0.190	0.080
Income	-0.048	0.032	0.097	0.028
Democrat	-0.395	0.160	0.708	0.145
Ideology	0.224	0.044	-0.122	0.040
N	830		886	
R^2	.06		.06	

Analysis Including Subjects Who Fail to Pass Manipulation Checks

Table 9 shows the number of observations of each group including respondents who fail to pass manipulation checks. As you can see, the sample is more balanced across the groups before dropping inattentive respondents. In our sample,1916 out of 2517 (i.e. 76%) respondents answered the manipulation checks correctly. The results of cross-tabulations are similar to what we report in the main text, even if we include these observations.

Table 9: Sample Size

Group	Threat	Use of Force	N
Stay Out	No	No	624
Back In	No	Yes	621
Back Out	Yes	No	653
Go In	Yes	Yes	619

Table 10: Domestic Political Consequences of Being Inconsistent (Outcome = No Force)

Study	% Who Approve "Stay Out"	% Who Approve "Back Out"	Difference in Approval
Levy et al.	51 [46, 57]	29 [24, 33]	-22 [-30, -16]
This Study	28 [25, 32]	22 [19, 25]	-6 [-11, -2]

Note: 95% confidence intervals in parentheses. Reported first differences may differ slightly due to rounding.

Table 11: Domestic Political Consequences of Being Inconsistent (Outcome = Use Force)

Study	% Who Approve "Go In"	% Who Approve "Back In"	Difference in Approval
Levy et al.	53 [47, 58]	41 [36, 46]	-12 [-19, -4]
This Study	55 [51, 59]	59 [56, 63]	4 [-1, 9]

Note: 95% confidence intervals in parentheses. Reported first differences may differ slightly due to rounding.