

Supplementary Materials

**Rethinking Reputation:
When Fighting to Demonstrate
Resolve Backfires**

Joshua A. Schwartz*

Contents

Whether the Theory Can Hold In Autocracies	1
War-Weariness Experiment on the U.S. Public	2
Research Design	2
Results	5
Robustness	7
Heterogeneous Effects	9
Mechanisms	11
Elite Experiment on UK MPs	13
Robustness Tests	13
Open-Ended Responses	16
Survey Text	19
References	22

*Assistant Professor of International Relations, Carnegie Mellon University, joshschwartz@cmu.edu.

Whether the Theory Can Hold In Autocracies

One relevant question is whether my theory can operate only in democracies or also in autocracies. Because non-democracies typically do not hold elections, hold only sham elections, or hold elections so tightly regulated they do not give voters much choice at all, critics might wonder whether public opinion is actually a component of a state's resolve in autocracies and thus whether it can affect their reputation for resolve. Although it is true that public opinion plays a much more significant role in democracies than autocracies, meaning this mechanism of my theory is more likely to hold in the former than the latter, public opinion does play a role in autocracies. After all, if autocrats are not responsive at all to public opinion, then they risk losing support and triggering violent or non-violent actions against their regimes. In accordance with this view, a growing body of literature demonstrates that autocratic regimes do respond to public opinion, even in the realm of foreign policy (e.g., Weeks 2008; Reilly 2011; Chen, Pan, and Xu 2016; Truex 2016; Meng, Pan, and Yang 2017; Li and Chen 2021; Lueders 2022). Of course, it should be noted that misinformation and disinformation presented to the public, especially in autocracies where the government controls the media, could reduce the chances of war-weariness if the state minimizes negative information presented to the public. Nevertheless, some costs of war—like dead, maimed, or psychologically traumatized soldiers—are impossible to fully hide from family and friends. Finally, it is also important to note that in autocracies, war-weariness among *leaders* can still reduce a country's resolve. In fact, while the public opinion pathway of my theory may be weaker in autocracies than democracies, the leader pathway may be stronger given the increased power decision-making elites are conferred in non-democracies.

Take Russia for example. Following the shock of the disastrous Crimean War in the 1850s, Russia adopted a policy of retrenchment that involved a reduction in military spending and military commitments, a greater focus on domestic development, and a greater reliance on allies (MacDonald and Parent 2018). For example, Russian Foreign Minister Alexander Gorchakov told the Tsar Russia would “have to focus persistently on the realization of [its] internal development and the entire foreign policy [would] have to be subordinated to this main task” (Splidsboel-Hansen 2002, 380). One of Gorchakov's principal advisors similarly said, “The development of [Russia's] internal life, her productive resources, her prosperity, her culture, her commerce, her industry [are] all things which require peace. Her foreign policy should thus be purely preventative and defensive” (Tuminez 2000, 110). Likewise, scholars have argued and provided evidence for the contention that the Soviet Union's War in Afghanistan in the 1980s undermined its resolve and willingness to intervene in Eastern Europe in 1989, the events of which ultimately led to the USSR's collapse (e.g., Arnold 1993; Chafetz 1993; Bennett 2005). While the Soviet government tried to hide the high casualty numbers as a result of the Afghan War in order to minimize backlash, you cannot hide tens of thousands of casualties and, ultimately, public opinion turned against the war (Arnold 1993, 196). The Afghanistan fiasco also changed how Soviet policymakers thought about the utility of military force. In discussions about whether to invade Poland or not in 1989 in response to revolutionary activity, comments such as “one Afghanistan is enough for us” were made by high-level Soviet officials (Bloom 2004, 291). None of these examples are dispositive, but they provide suggestive evidence that the mechanisms of my argument can operate in autocracies as well as democracies.

War-Weariness Experiment on the U.S. Public

Research Design

In order to test whether war-weariness—which is the key mechanism of my theory—can occur and when it is relatively more likely, I designed and administered a 3 x 3 between-subjects experiment on a representative sample of 716 Americans recruited via Lucid in April 2020. Lucid generates representative samples based on age, gender, ethnicity, and region, and has been shown to perform well replicating previous studies (Coppock and McClellan 2019; Peyton, Huber, and Coppock 2021). I also block on respondent party identification to ensure approximately equal numbers of Democrats and Republicans in each experimental condition. I conducted this experiment on members of the American public given the United States’ outsized role in foreign affairs, which makes the views of American citizens substantively meaningful. As demonstrated by Kertzer (2016), there is also an important connection between public opinion and a state’s aggregate level of resolve, especially in democracies.

In accordance with prominent experimental studies in international relations conducted by Press, Sagan, and Valentino (2013), Jones (2019), and others, respondents are presented with a mock newspaper article written by the Associated Press in order to make the experimental scenario more realistic. The opening paragraphs of the article are the same for all respondents:

Three years ago today—on October 7, 2026—the Islamic State attempted to set off a bomb in the Wilshire Grand Center building in downtown Los Angeles. Luckily, the bomb failed to detonate, and no one was hurt. After the attack, President John Richards (R) announced that, according to U.S. intelligence, the dictatorial government of Yemen was providing funding to the Islamic State.

The survey scenario takes place several years in the future in order to distance the experiment from the political moment at which it was fielded.¹ I include the name and the partisan identification of the US president in order to control for leader gender and political party across experimental conditions (Schwartz and Blair 2020). The Wilshire Grand Center is an actual building in Los Angeles, increasing the realism of the survey scenario. The Islamic State conducts the attack because it is widely known as a terrorist group and could plausibly carry out such an operation. Yemen is chosen as the source of the attack because the Islamic State has a presence there and because the country is relatively unstable, making it credible that a future government could indeed support the Islamic State.

In the remainder of the article, I experimentally vary whether the United States government (1) declines to take military action in response to the attack (i.e., backs down from using force); (2) launches a war in response to this attack and faces a least-likely case for war-weariness; or (3) launches a war in response to this attack and faces a most-likely case for war-weariness. In the first condition, survey subjects are told that President Richards:

¹Besides noting that the scenario takes place in 2026, I also explicitly state that the president’s name is “John Richards,” which should further distance the scenario from the Trump presidency in respondents’ minds.

was imposing tough sanctions on Yemen. However, he declined to take any military action against the country. After three years, U.S. intelligence believes that the Yemeni government continues to fund terrorists.

Since it is implausible the US would have zero response to this attempted attack, in the “backing down” condition sanctions—a commonly used foreign policy tool—are imposed on Yemen, but no military action is taken. In other words, the US does respond to the terrorist attack, but backs down from using military force, which Lin-Greenberg (2019) calls “backing up.” That the US backed down from a conflict over the terrorist attack is further primed by the title of the article: “Should the U.S. Have Responded More Forcefully to Attempted Los Angeles Terrorist Attack?” Consistent with backed down conditions in other experimental studies (e.g., Renshon, Dafoe, and Huth 2018), the US does not achieve its objectives as the Yemeni government continues to fund terrorists.

By contrast, in the second condition the US launches a war in response to the terrorist attack and faces a least-likely case for war-weariness because the costs of the war are low, the outcome of the war is victory, and the objective is restraining foreign policy aggression.² Restraining foreign policy aggression is primed by telling respondents that “President Richards declared war on Yemen in order to punish them for their support of the Islamic State and coerce them to break off all relations with terrorists.” In other words, the goal of the intervention was not to overthrow the regime, but to change their foreign policy. Low costs are primed by informing respondents that the war leads to 400 dead US soldiers and \$100 billion spent, which are about the costs of the Gulf War (Daggett 2010; DeBruyne 2018). Finally, victory is primed by telling subjects that:

After three years of fighting, the military operation was a success and is now ending. Yemeni military targets were successfully destroyed, the government agreed to break off all relations with terrorists, and they turned over valuable intelligence to the United States about the Islamic State.

The title of the article is also “Military Operation is a Huge Success” and a picture is included of US soldiers returning home to applause.

In the third condition, the US also launches a war, but faces a most-likely case for war-weariness. Specifically, the costs of the war are high, the outcome of the war is defeat, and the objective is internal political change. Internal political change is primed because President Richards “declared war on Yemen in order to overthrow their government and install a democratic regime in its place.” The costs of the war are high: 40,000 US soldiers die—similar to the Korean and Vietnam Wars—and \$1 trillion is spent—similar to the Vietnam, Afghanistan, and Iraq Wars (Daggett 2010; Trotta 2013; DeBruyne 2018). Lastly, defeat is primed by telling respondents: “With the U.S. now leaving Yemen without having been able to establish a stable democracy, many Americans are questioning whether the war was really worth it. To further prime defeat and high costs, the title of the article is “40,000 Dead and \$1 Trillion Spent. But for What?” A picture is also included of US soldiers returning home in a casket.

²Restraining foreign policy aggression is often (though not always) correlated with lower costs and a lower likelihood of defeat than wars whose goal is internal political change (Downes 2021). For example, the 1991 Gulf War (whose goal was to repel Saddam Hussein’s invasion of Kuwait) was cheaper and its goals were more readily achieved than the 2003 Iraq War (whose goal was regime change). These dynamics help explain why the public is often (though, again, not always) more supportive of restraining foreign policy aggression than internal political change (Jentleson 1992).

Overall, the treatments for the least-likely and most-likely case for war-weariness are relatively strong. However, given the context of this experiment, I believe it is appropriate to have relatively strong treatments. It is quite difficult to experimentally prime the emotions associated with a war involving a most-likely case for war-weariness. Such a conflict would likely involve months or years of television, newspaper, and social media coverage of the war and its costs, and for some members of the public personal connections to dead, injured, or traumatized soldiers. Thus, while strong relative to a more bland survey vignette, these treatments are likely weak compared to the equivalent real-world treatments.

After reading the article, respondents are presented with a subsequent international crisis and asked to what extent they would support US military intervention to resolve the situation. Since support for engaging in a military intervention implicates respondents' willingness to risk and endure costs to achieve their objective, it can be considered a measure of resolve. I also hold constant US military capabilities in all experimental conditions in order to analytically separate resolve and capabilities.

Following Herrmann, Tetlock, and Viser (1999), as well as Flores-Macias and Kreps (2017), there are three different types of international crises that subjects are randomly assigned to: (1) a humanitarian crisis involving ethnic cleansing; (2) a foreign country invading an ally of the US; and (3) the discovery of a clandestine nuclear facility possessed by a country hostile to the US. These three scenarios cover a range of different types of missions on the interest-values continuum and thus should increase the external validity of the experiment by allowing me to probe whether war-weariness from a past conflict can reduce resolve for a variety of different types of future conflicts.

The experiment is a fully crossed 3 x 3 between-subjects design. Per the discussion in the main text, I expect that respondents assigned to the most-likely case for war-weariness condition will be less resolved than those in the backed down condition. In other words, fighting today rather than backing down can reduce a state's actual resolve tomorrow. On the other hand, I do *not* expect subjects in the least-likely case for war-weariness condition to be less resolved than those in the backed down condition.

Results

Table A.1 displays respondents’ support for military action in Crisis II based on the US actions in Crisis I and the outcome of that crisis. In accordance with previous studies, I collapse the 7-point measure of support for military action into a binary measure in order to more clearly illustrate substantive effects. Substantively identical results emerge with the full seven-point measure. In accordance with my expectations, I find that respondents assigned to the most-likely case for war-weariness condition are less resolved than those in the backed down condition. Specifically, they are 11.2 percentage points less likely to support subsequent military action ($p < 0.01$). As anticipated, I find no significant difference in resolve between subjects in the backed down condition and subjects in the least-likely case for war-weariness condition. This provides empirical evidence that the factors outlined in the main text do indeed impact the likelihood of war-weariness developing.

Table A.1: Evidence for the Existence of War-Weariness Among the Public

Action in Previous Conflict	Support for Subsequent Military Action (%)	Difference from Backed Down (Percentage Points)
Backed Down	44.1%	—
Stood Firm, Least-Likely Case for War-Weariness	43.2%	-0.9
Stood Firm, Most-Likely Case for War-Weariness	32.9%	-11.2***

Note: Results depict percentage support for engaging in military action to alleviate a humanitarian crisis, protect an ally, or combat a nuclear weapons program. * = $p < 0.10$, ** = $p < 0.05$, and *** = $p < 0.01$, where p-values indicate whether support is statistically different than 0.

Do results vary based on the nature of the follow-up mission? I analyze this question in Table A.2 and utilize the full 7-point scale because it is more representative of the results than the binary measure.³ I find that respondents assigned to the most-likely case for war-weariness condition are less resolved than those in the backed down condition when Crisis II involves a humanitarian crisis or the invasion of an ally. There is no significant difference—though the results are in the expected direction—when Crisis II involves a nuclear proliferation crisis. Given that this condition involves a more direct security threat to the US than a humanitarian crisis or the invasion of an ally, it makes sense that respondents would be more willing to support military action regardless of the outcome of a previous crisis. As expected, Table A.2 also demonstrates that there is no significant difference in resolve between subjects in the backed down condition and

³Substantively similar results emerge when the binary measure is employed.

subjects in the least-likely case for war-weariness condition, regardless of the nature of the follow-up mission.

Table A.2: Does War-Weariness Vary with the Nature of the Follow-Up Mission?

Action in Previous Conflict	Support for Subsequent Military Action	Difference from Backed Down
<i>Humanitarian Crisis</i>		
Backed Down	4.03	—
Stood Firm, Least-Likely Case for War-Weariness	3.88	-0.15
Stood Firm, Most-Likely Case for War-Weariness	3.69	-0.34*
<i>Invasion of an Ally</i>		
Backed Down	4.35	—
Stood Firm, Least-Likely Case for War-Weariness	4.43	0.08
Stood Firm, Most-Likely Case for War-Weariness	3.82	-0.53**
<i>Nuclear Proliferation Crisis</i>		
Backed Down	4.12	—
Stood Firm, Least-Likely Case for War-Weariness	3.79	-0.33
Stood Firm, Most-Likely Case for War-Weariness	3.95	-0.17

Note: Results depict support for engaging in military action on a 7-point scale to alleviate a humanitarian crisis, protect an ally, or combat a nuclear weapons program. * = $p < 0.10$, ** = $p < 0.05$, and *** = $p < 0.01$, where p-values indicate whether support is statistically less than 0.

Robustness

To ensure the robustness of these findings, I take several steps. First, in Table A.3 I show that the results hold when the full 7-point measure of resolve is utilized.

Table A.3: 7-Point Measure of Support for Military Action

Action in Previous Conflict	Support for Subsequent Military Action	Difference from Backed Down
Backed Down	4.17	—
Stood Firm, Least-Likely Case for War-Weariness	4.05	-0.12
Stood Firm, Most-Likely Case for War-Weariness	3.82	-0.34**

Note: Results depict support for engaging in military action on a 7-point scale to alleviate a humanitarian crisis, protect an ally, or combat a nuclear weapons program. * = $p < 0.10$, ** = $p < 0.05$, and *** = $p < 0.01$, where p-values indicate whether support is statistically different than 0.

Table A.4 illustrates that the results remain robust in a regression context when controlling for other factors like respondent party identification, militant assertiveness, gender, education, age, income, and race. Specifically, the negative and statistically significant coefficients in Models 1 and 2 mean that respondents assigned to the most-likely case for war-weariness condition are less willing to engage in military action than those assigned to the backed down condition. There is no similar effect for subjects assigned to the least-likely case for war-weariness condition. I also control for the extent to which respondents believe the US is likely to achieve its military objectives if it decided to involve itself in Crisis II. Although I hold constant military capabilities across all of the experimental conditions, one potential concern is that respondents assigned to the most-likely case for war-weariness condition will believe that the US is less likely to achieve its military objectives in a subsequent military conflict, leading to a lack of information equivalence across experimental conditions (Dafoe, Zhang, and Caughey 2018). If that is the case, then it could be perceptions about the America’s chance of victory rather than respondent resolve per se driving the results. Promisingly, there is no systematic evidence of confounding. Respondents in the most-likely case for war-weariness condition were only slightly less likely than those in the backed down condition to believe the US would achieve its military objectives in a conflict with Country A ($\rho \approx -0.11$). More importantly, the results are robust to holding this factor constant in a regression.

Table A.4: Holding Other Factors Constant

	7-Point DV (1)	Binary DV (2)	7-Point DV (3)	Binary DV (4)
Most-Likely Case for War-Weariness	-0.33** (0.14)	-0.57*** (0.21)		
Least-Likely Case for War-Weariness			-0.08 (0.14)	-0.05 (0.20)
Militant Assertiveness	0.70*** (0.09)	0.85*** (0.15)	0.76*** (0.09)	0.73*** (0.15)
Republican	0.03 (0.15)	0.15 (0.22)	0.08 (0.15)	0.10 (0.22)
Female	-0.10 (0.14)	-0.20 (0.20)	0.001 (0.14)	-0.02 (0.20)
Education	-0.0000 (0.0002)	-0.0001 (0.0004)	-0.0003 (0.0003)	-0.02 (0.06)
Age	-0.07 (0.05)	-0.03 (0.08)	-0.09 (0.06)	-0.11 (0.08)
Income	0.02 (0.04)	0.06 (0.07)	0.05 (0.05)	0.10 (0.07)
White	-0.29 (0.21)	-0.36 (0.30)	-0.35* (0.19)	-0.71** (0.28)
Black	-0.48* (0.27)	-0.82* (0.42)	-0.51* (0.27)	-0.77* (0.40)
Perceived Chance of Victory	0.27*** (0.07)	0.41*** (0.12)	0.14* (0.08)	0.46*** (0.12)
Constant	1.30*** (0.41)	-4.33*** (0.67)	1.50*** (0.41)	-3.74*** (0.67)
Observations	517	517	494	494

Note: *p<0.1; **p<0.05; ***p<0.01

Heterogeneous Effects

I now analyze whether the results vary based on respondent partisan identification and militant assertiveness, which have been shown to be key moderators in previous research (e.g., Herrmann, Tetlock, and Viser 1999; Kertzer and Brutger 2016). *A priori*, we might expect that individuals high in militant assertiveness and Republicans would be less susceptible to war-weariness, as they may be predisposed to discount information that clashes with their worldview; for example, that war may be less beneficial and more costly than previously thought. However, as shown in Tables A.5 and A.6, the basic results hold among both Democrats and Republicans, as well as those high and low in militant assertiveness.⁴ There is also no significant interactions between these variables and the treatments of interest in a regression context. Thus, war-weariness in response to a costly and ineffective war is not a phenomenon restricted solely to Democrats and those low in militant assertiveness. Instead, it is a more generalizable dynamic.

Table A.5: Does War-Weariness Vary by Level of Militant Assertiveness?

Action in Previous Conflict	Support for Subsequent Military Action (%)	Difference from Backed Down (Percentage Points)
<i>Low in Militant Assertiveness</i>		
Backed Down	24.2%	—
Stood Firm, Least-Likely Case for War-Weariness	28.0%	3.9
Stood Firm, Most-Likely Case for War-Weariness	10.1%	-14.1***
<i>High in Militant Assertiveness</i>		
Backed Down	58.5%	—
Stood Firm, Least-Likely Case for War-Weariness	53.8%	-4.6
Stood Firm, Most-Likely Case for War-Weariness	45.5%	-13.0**

Note: Results depict percentage support for engaging in military action to alleviate a humanitarian crisis, protect an ally, or combat a nuclear weapons program. * = $p < 0.10$, ** = $p < 0.05$, and *** = $p < 0.01$, where p-values indicate whether support is statistically less than 0.

⁴I utilize the median split to categorize respondents as low or high in militant assertiveness.

Table A.6: Does War-Weariness Vary by Party Identification?

Action in Previous Conflict	Support for Subsequent Military Action	Difference from Backed Down
<i>Democrats</i>		
Backed Down	3.88	—
Stood Firm, Least-Likely Case for War-Weariness	3.77	-0.11
Stood Firm, Most-Likely Case for War-Weariness	3.54	-0.34**
<i>Republicans</i>		
Backed Down	4.47	—
Stood Firm, Least-Likely Case for War-Weariness	4.35	-0.11
Stood Firm, Most-Likely Case for War-Weariness	4.10	-0.37*

Note: Results depict support for engaging in military action on a 7-point scale to alleviate a humanitarian crisis, protect an ally, or combat a nuclear weapons program. * = $p < 0.10$, ** = $p < 0.05$, and *** = $p < 0.01$, where p-values indicate whether support is statistically less than 0.

Mechanisms

Finally, after measuring the dependent variable, I asked respondents several questions in order to probe the mechanisms underlying my findings. The questions primarily asked respondents about the costs and benefits of fighting in Crisis II, and the results are displayed in Table A.7

Table A.7: What are the Mechanisms of War-Weariness?

Agreement with the statement that...	Most-Likely Case for War-Weariness (%)	Backed Down (%)	Difference (Percentage Points)
The costs of intervening outweigh the benefits	52.3%	39.3%	12.9***
It would cost a substantial or enormous amount to win	56.8%	44.4%	12.3***
Intervening is slightly or not at all in the American interest	39.2%	31.5%	7.7**
Backing down would harm America’s reputation for toughness little or not at all	29.7%	24.1%	5.7*
Victory is unlikely if America intervenes	14.4%	6.4%	8.0***

Note: Results depict percentage support for the statements. * = $p < 0.10$, ** = $p < 0.05$, and *** = $p < 0.01$, where p-values indicate whether support is statistically greater than 0.

Per Kertzer (2016), a greater estimate of the costs and a lower estimate of the benefits of fighting should reduce individual-level resolve. As expected, involvement in a previous war that is a most-likely case for war-weariness increases estimates of a subsequent conflict’s costs and reduces estimates of its benefits relative to backing down. For example, while just 39.3% of respondents assigned to backed down condition agree that the costs of intervening in Crisis II outweigh the benefits, 52.3%—a majority—of individuals in the most-likely case for war-weariness condition believe the costs outweigh the benefits. Similarly, respondents in the most-likely case for war-weariness condition increase their estimates of how costly a subsequent intervention would be and decrease their assessments of how much intervening is in the American interest. Thus, involvement in a past war can shift the public’s estimates of a future conflict’s costs, benefits, and the relative balance of the two.

Interestingly, respondents are also more likely to believe that backing down in Crisis II would not harm America’s reputation for toughness in the most-likely case for war-weariness condition than in the backed down condition. In other words, the reputational costs of backing down in Crisis II are seen as lower for those in the most-likely case for war-weariness condition than the backed down condition. The implication being that fighting is more attractive to those in the backed down condition than in the most-likely case for war-weariness condition. The explanation may be, at least partially, a reflection of the “never again” hypothesis, as individuals may not want to back down twice for fear of being perceived as irresolute (Baser 2024). Finally, respondents in the most-likely case

for war-weariness condition believe that victory is less likely in Crisis II than those in the backed down condition. Since the perceived chance of victory may also be related to capabilities, I show that the main results are robust to controlling for this factor in a regression, and I also controlled for relative capabilities in the scenario itself.

Finally, utilizing a regression approach to conduct causal mediation analysis, I show in Table A.8 that the above factors do indeed help explain why experiencing a conflict involving a most-likely case for war-weariness reduces future resolve relative to backing down in a previous conflict (Imai, Keele, and Tingley 2010; Imai, Keele, and Yamamoto 2010). Odd-numbered models in Table A.8 analyze the relationship between the key independent variable of interest—whether respondents were assigned to the most-likely case for war-weariness condition or the backed down condition—and each respective mediator. Even-numbered models then analyze the relationship between each mediator and the dependent variable: support for military action in Crisis II. For example, Model 1 shows that respondents in the most-likely case for war-weariness condition were more likely than those in the backed down condition to believe that the costs of intervening in Crisis II outweigh the benefits. Model 2 then shows that respondents who believed the costs outweighed the benefits were less likely to support military action in Crisis II.⁵ These results illustrate the contribution of this project relative to Kertzer (2016). While he establishes that the costs and benefits of war can impact a state’s resolve, he does not consider whether or how past actions and experiences—such as war—affect these estimates. I show that the experience of a prior war can increase estimates of future conflict’s cost and decrease estimates of its benefits, thereby reducing an individual’s resolve.

Table A.8: Mediation Analysis

	Costs Outweigh Benefits (1)	Mil Action (2)	High Costs (3)	Mil Action (4)	Low Interests (5)	Mil Action (6)	Low Reputation Interests (7)	Mil Action (8)	Victory Unlikely (9)	Mil Action (10)
Most-Likely Case for War-Weariness	0.298*** (0.0975)	-0.262** (0.130)	0.290*** (0.0900)	-0.332** (0.136)	0.117 (0.0934)	-0.357*** (0.133)	0.122 (0.0956)	-0.368*** (0.135)	0.267*** (0.0808)	-0.326** (0.136)
Costs Outweigh Benefits		-0.456*** (0.0588)								
High Costs				-0.227*** (0.0666)						
Lower Interests						-0.348*** (0.0631)				
Low Reputation Interests								-0.250*** (0.0625)		
Victory Unlikely										-0.270*** (0.0741)
Militant Assertiveness	-0.375*** (0.0635)	0.639*** (0.0869)	-0.160*** (0.0586)	0.773*** (0.0886)	-0.414*** (0.0608)	0.666*** (0.0902)	-0.477*** (0.0622)	0.691*** (0.0925)	-0.421*** (0.0526)	0.696*** (0.0931)
Republican	-0.188* (0.106)	-0.00636 (0.140)	-0.126 (0.0975)	0.0509 (0.146)	-0.114 (0.101)	0.0397 (0.144)	-0.316*** (0.103)	0.000419 (0.147)	-0.167* (0.0875)	0.0342 (0.146)
Female	-0.185* (0.0971)	-0.256** (0.129)	0.143 (0.0896)	-0.140 (0.135)	-0.0241 (0.0930)	-0.181 (0.132)	0.0944 (0.0951)	-0.149 (0.134)	0.250*** (0.0804)	-0.105 (0.135)
Education	-0.000227 (0.000178)	-0.0000955 (0.000236)	0.0000266 (0.000164)	0.0000141 (0.000246)	-0.000108 (0.000170)	-0.0000294 (0.000242)	0.000218 (0.000174)	0.0000624 (0.000246)	-0.0000828 (0.000147)	-0.0000143 (0.000246)
Age	-0.0481 (0.0391)	-0.0969* (0.0519)	-0.0734** (0.0361)	-0.0917* (0.0544)	-0.0531 (0.0375)	-0.0935* (0.0533)	0.0427 (0.0383)	-0.0644 (0.0540)	0.0209 (0.0324)	-0.0694 (0.0541)
Income	0.0168 (0.0324)	0.0327 (0.0429)	-0.0139 (0.0299)	0.0219 (0.0449)	-0.0289 (0.0310)	0.0150 (0.0441)	-0.0163 (0.0317)	0.0210 (0.0447)	-0.0343 (0.0268)	0.0158 (0.0448)
White	0.257* (0.149)	-0.196 (0.197)	-0.00970 (0.137)	-0.315 (0.206)	-0.00767 (0.142)	-0.316 (0.202)	0.0618 (0.146)	-0.298 (0.205)	0.0699 (0.123)	-0.294 (0.205)
Black	0.314 (0.196)	-0.381 (0.260)	-0.0194 (0.181)	-0.528* (0.272)	-0.221 (0.188)	-0.601** (0.267)	0.118 (0.192)	-0.494* (0.271)	0.154 (0.163)	-0.482* (0.271)
Constant	4.529*** (0.264)	4.026*** (0.439)	4.207*** (0.243)	2.916*** (0.460)	4.755*** (0.253)	3.616*** (0.468)	4.271*** (0.258)	3.029*** (0.451)	3.566*** (0.219)	2.926*** (0.450)
Observations	517	517	517	517	517	517	517	517	517	517

Notes: Standard errors in parentheses. *p<0.10; **p< 0.05; ***p<0.01

⁵Note that while the mediators in Models 5 and 7 fall just short of conventional levels of statistical significance ($p < 0.21$), both are significantly associated with lower support for military action.

Elite Experiment on UK MPs

Robustness Tests

I now proceed to analyze the robustness of the results displayed in the main text and to probe whether there are any heterogeneous treatment effects. I first do so for war-weariness and backing down reputation costs, which are the within-subject treatment effects. Model 3 in Table A.9 analyzes whether the finding that war-weariness and backing down reputation costs exist holds when controlling for covariates like party identification and gender. The negative and statistically significant coefficients on *War-Weariness* and *Backed Down* indicate that these two treatments both reduce Country B's reputation for resolve relative to the baseline condition, even when controlling for other factors. Model 4 adds survey weights to ensure the sample is representative based on party identification, gender, region, and electoral cohort. The findings are robust to the inclusion of weights. Model 5 shows the results also hold when both controls and weights are included.

Given the 100-point scale utilized to measure reputation for resolve, one problematic possibility is that outliers could be driving the findings for war-weariness and backing down costs. This is a particularly significant concern for this elite experiment because the small sample size means one or a small handful of MPs that rate Country B's resolve as extremely high (close to a 100% chance of standing firm) or extremely low (close to a 0% chance of standing firm) could skew the average. To probe the robustness of my results to outliers, I drop the top and bottom 10% of dependent variable responses in Model 6. The core findings hold for this test.

Models 7 and 8 then test for heterogeneous effects. That is, they test for whether war-weariness or backing down reputation costs are higher or lower depending on different respondent characteristics. The quantities of interest for these analyses are not the regression coefficients themselves, but the differences depicted in the bottom rows of the table. Model 7 analyzes heterogeneous effects based on party identification. There is little statistically significant evidence that war-weariness or backing down costs differ based on party identification. Model 8 examines heterogeneous effects based on whether the MP identifies a male or female. No significant differences emerge for this analysis either.

Table A.9: Within-Subject Effects: Robustness & Heterogeneous Effects

	Reputation for Resolve Relative to the Baseline							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
War-Weariness	-17.84*** (2.86)	-29.41*** (7.57)	-21.01*** (6.50)	-18.26*** (3.57)	-22.28*** (6.28)	-18.60*** (3.55)		
Backed Down	-15.76*** (3.54)	-24.49*** (8.52)	-18.58*** (6.09)	-16.09*** (3.85)	-19.43*** (5.77)	-17.32*** (3.75)		
Conservative Party			4.15 (6.65)		4.86 (6.50)	0.89 (3.99)		4.00 (7.33)
Labour Party			4.32 (7.26)		6.14 (6.26)	1.26 (3.95)		6.95 (8.26)
Female			-1.72 (5.32)		-2.76 (6.45)	0.95 (3.16)	-1.55 (6.28)	
War-Weariness; Labour Party							-22.52*** (6.09)	
War-Weariness; Conservative Party							-15.32*** (5.58)	
War-Weariness; Other Party							-14.57*** (4.52)	
Backed Down; Labour Party							-5.73 (5.86)	
Backed Down; Conservative Party							-17.58*** (5.21)	
Backed Down; Other Party							-23.89*** (7.69)	
War-Weariness; Female								-30.56*** (9.27)
War-Weariness; Male								-19.40** (7.50)
Backed Down; Female								-16.98* (8.60)
Backed Down; Male								-21.97*** (7.04)
<i>Difference Lab/Cons</i>							WW: -7.20 BD: -11.85	
<i>Difference Lab/Other</i>							WW: -7.95 BD: 18.16**	
<i>Difference Cons/Other</i>							WW: -0.75 BD: 6.31	
<i>Difference Male/Female</i>								WW: 11.16 BD: -4.99
Weights	×	×	×	✓	✓	×	✓	✓
Binary DV	×	✓	×	×	×	×	×	×
Observations	100	100	100	100	100	77	100	100

Notes: *p<0.10; **p< 0.05; ***p<0.01. WW = War-Weariness; BD = Backed Down.

Next, I analyze robustness and heterogeneous effects when directly comparing the war-weariness and backed down treatments, which is the between-subjects effect in this experiment. Model 2 in Table A.10 tests whether Country B’s reputation for resolve in the war-weariness condition is still lower than in the backed down condition when controlling for covariates. The coefficient on *War-Weariness* is negative and statistically significant, indicating that reputation for resolve is indeed lower among UK MPs when Country B shows signs of war-weariness compared to when they back down. Model 3 shows this result also holds when including weights to increase representativeness, and Model 4 demonstrates it is also robust to including both weights and covariates. Model 5 addresses the possibility of outliers by trimming the top and bottom 10% of responses.

Model 6 adopts a different approach in order to compare the war-weariness and backed down treatments. Instead of conducting a purely between-subjects comparison of reputation for resolve in the two treatments, Model 6 also controls for reputation for resolve in the baseline condition. This estimate is therefore a kind of hybrid between and within-subjects estimate; it incorporates both the within-subjects estimate of perceived reputation for resolve in the war-weariness/backed down treatment relative to the baseline, and the between-subjects estimate of perceived reputation for resolve in the war-weariness treatment relative to the backed down treatment. The core result holds for this test as well. Overall, these findings provide evidence for the strongest contention of my argument: that choosing to fight in the moment of decision can undermine a country’s reputation for resolve relative to backing down. Models 7 and 8 analyze heterogeneous effects through a standard interaction approach. No significant heterogeneous effects emerge based on party identification (Model 7) or gender (Model 8).

Table A.10: Between-Subject Effects: Robustness & Heterogeneous Effects

	Reputation for Resolve							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
War-Weariness	-7.93** (4.65)	-8.74** (4.75)	-9.03* (5.89)	-9.93** (5.95)	-6.91** (3.93)	-5.50* (4.24)	-6.00 (6.52)	-8.92* (6.55)
War-Weariness x Labour Party							-12.60 (14.57)	
War-Weariness x Female								-3.00 (13.64)
Conservative Party		10.35* (6.23)		9.86* (6.88)	2.07 (5.97)	7.17* (5.30)		9.70* (6.99)
Labour Party		9.11* (6.57)		11.48* (8.17)	-1.96 (6.20)	6.65 (5.37)	8.57 (11.11)	11.63* (8.44)
Female		3.59 (4.88)		4.11 (5.22)	-1.24 (4.18)	0.87 (4.84)	4.92 (5.45)	5.55 (9.29)
Other Party							-9.10* (6.93)	
Resolve Baseline						0.51*** (0.11)		
Constant	38.78*** (3.34)	29.79*** (5.41)	41.35*** (4.37)	31.33*** (6.04)	37.51*** (5.16)	4.96 (7.13)	38.91*** (4.88)	30.86*** (6.34)
Weights	×	×	✓	✓	×	×	✓	✓
Observations	100	100	100	100	77	100	100	100

Notes: *p<0.10; **p< 0.05; ***p<0.01

Open-Ended Responses

As is standard in the literature, the main dependent variable in this experiment was a forced-choice closed-ended question that asked MPs to rate Country B's resolve on a 100-point scale. However, this measurement strategy has downsides. Most notably, closed-ended questions do not allow respondents to *explain* their answer, which makes it difficult to understand the logic driving results (Krosnick 1999; Roberts et al. 2014). For example, even if the results from closed-ended questions are in accordance with my argument, we cannot know for sure whether they are driven by the logic of my theory (e.g., a previous conflict leads to a reduction in public and/or policymaker support for war and that leads to a reduction in their reputation for resolve) or some other rationale. Furthermore, closed-ended questions cannot necessarily disentangle the decisive factor(s) impacting respondents' opinions (Lazarsfeld 1944). For instance, if signs of war-weariness reduce Country B's reputation for resolve, is that primarily due to the war-weariness among the general public or among decision-making elites? To address this issue, I included an open-ended question that asked MPs to explain why their estimates of Country B's probability of standing firm did or did not differ between the first and second scenarios.

Given the small sample size of this experiment, I conducted a qualitative analysis of the open-ended responses rather than estimating a more formal topic model (Roberts et al. 2014). Beginning with MPs randomly assigned to the war-weariness treatment, it is clear that the results were driven by the logic outlined by my argument. The most popular explanation for why MPs assessed Country B's resolve to be lower in the war-weariness treatment relative to the baseline treatment was reduced public support for war following B's previous conflict:

- *Labour MPs*
 - "Expect that as a democracy they will be concerned by public opinion and that will influence their decision."
 - "Public opinion."
- *Conservative MPs*
 - "Because of pressure on the government by public opinion and the media."
 - "Public opinion in Country B."
 - "Public opinion."⁶
- *Other MPs*
 - "The optics for country B were poor in scenario A but in scenario B Country B, a democracy, possesses a spurned demos after previous military exploits – therefore their % is even lower."
 - "Public opinion will influence a policy maker..."

⁶Two MPs answered the open-ended question this way.

Many respondents also pointed to the views of Country B’s leadership or the interaction between public opinion and leader views. This is in accordance with my argument surrounding war-weariness among decision-making elites. Furthermore, it is notable that none of the respondents articulated that they downgraded estimates of Country B’s reputation for resolve because leadership turnover combined with the inherent uncertainty of a new leader’s resolve led them to expect regression to the mean. Although certainly not dispositive, this non-finding combined with the above open-ended responses related to the power of public opinion provides reassurance that the findings are not being driven by leadership turnover itself. Moreover, the logic respondents articulate is more consistent with my contention that expectations of a new leader’s resolve are likely to be shaped by the choices State B made in the past and their consequences.

- “The losses suffered in the previous conflict and the unpopularity of conflict make it very unlikely that a democratically elected leader will choose war.”
- “Government was elected on an anti-conflict mandate so won’t have the stomach to escalate.”
- “Because of the basis upon which the leader was elected and the mood of the country following the previous conflict.”
- “Public opinion very important and new leader would make different decisions.”
- “In the second scenario it’s clear there is significant public feeling against the war and that new leadership also feels the war was mistake so it sounds like they are more likely to back down than in scenario one.”

Overall, the open-ended responses suggest that the findings in this experiment related to war-weariness reputation costs are driven by the logic outlined by my theory, as around 85% of the respondents assigned to the war-weariness condition used logic consistent with my argument to justify their response. These open-ended responses also provide additional qualitative evidence that public opinion matters to policymakers, as prior literature has demonstrated. The few respondents that did not outline logic consistent with my theory instead advocated a position closer to those of reputation skeptics, which is that past actions are not particularly informative about future resolve:

- “Impossible to predict the behaviour of the Country B so kept both at 50%.”
- “The facts leading up to the dispute remain.”⁷

Moving on to respondents assigned to the backed down condition, a review of the open-ended responses suggest backing down costs are driven by the logic outlined by the conventional wisdom. Specifically, in contrast to the general contention of those skeptical of reputation dynamics altogether, past actions do seem to influence expectations of future behavior. In fact, UK MPs in this experiment said this quite directly:

- “Previous actions.”
- “Previous behavior.”

⁷This respondent did not assess Country B’s resolve as different in the war-weariness treatment and the baseline treatment.

- “History of backing down.”
- “The history of Country B’s actions.”
- “Based on their history of how they dealt with the conflict previously.”
- “Possible propensity to back down in a dispute based on recent history.”
- “History.”
- “They have form for folding!”

One MP had a particularly strong implicit criticism of the contention of reputation skeptics that past actions are relatively uninformative: “I would have thought the answer was obvious!” Nevertheless, some MPs assigned to the backed down treatment did indicate views consistent with more skeptical reputation theories, though this was clearly a minority opinion:

- “I think the response will largely be the same.”
- “Both scenarios suggest the same probability.....”
- “I rated them the same. I went 50% to try to indicate my neutral view on the questions!”

Finally, one MP outlined logic consistent with the “never again” (Mercer 1996, 40) or “reassertion” (Stapleton 2014, 23-25) hypothesis: “Backing down didn’t work last time, so as a democracy country B has to stand strong or risk wrath of domestic population.”

In sum, an analysis of the open-ended question from the elite experiments suggests the main results are driven by logic consistent with my argument and the conventional wisdom respectively.

Survey Text

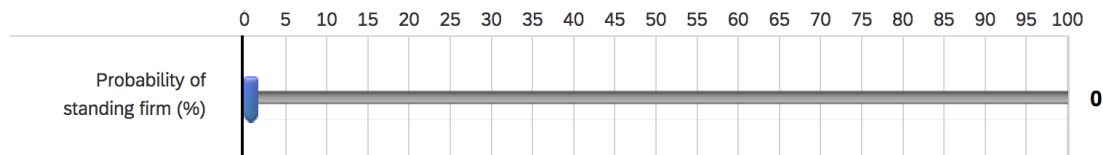
Baseline Scenario⁸

Imagine the following situation, which could occur at some point in the future:

- Your country—the United Kingdom—is involved in a dispute with Country B
- The dispute began with a collision between a US shipping vessel and a ship registered to Country B
- During the collision, injuries were reported on both sides
- Additionally, both countries maintain that their ship was carrying sensitive military technology and are suspicious of the motives of the other side, leading to a tense standoff at sea
- The United Kingdom's military forces are estimated to be about 2.5 times stronger than Country B's
- Country B is a democracy
- During its most recent major international crisis, Country B did not back down and the crisis led to war. According to most impartial observers, Country B did not achieve its aims

Baseline Dependent Variable Measure

Given the information available, what is your best estimate about whether Country B will stand firm in this dispute, ranging from 0% to 100%?



⁸All respondents are presented with this scenario.

Backed Down Scenario⁹

Now, we would like to ask you a question about a different, alternative version of the scenario you just read. Suppose the basic details remain the same:

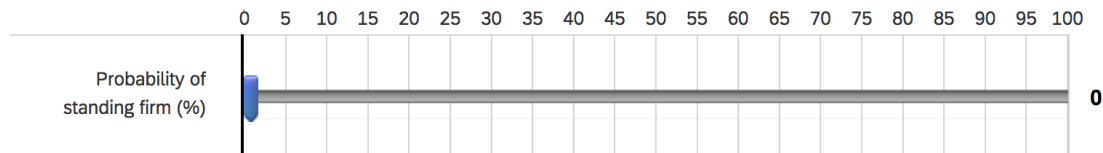
- Your country—the United Kingdom—is involved in a dispute with Country B
- The dispute began with a collision between a US shipping vessel and a ship registered to Country B
- During the collision, injuries were reported on both sides
- Additionally, both countries maintain that their ship was carrying sensitive military technology and are suspicious of the motives of the other side, leading to a tense standoff at sea
- The United Kingdom’s military forces are estimated to be about 2.5 times stronger than Country B’s
- Country B is a democracy

But this time, suppose that...

During its most recent major international crisis, Country B backed down and failed to achieve its aims

Backed Down Dependent Variable Measure

Given the information available, what is your best estimate about whether Country B will stand firm in this dispute, ranging from 0% to 100%?



⁹Half of respondents are randomly assigned to this scenario.

War-Weariness Scenario¹⁰

Now, we would like to ask you a question about a different, alternative version of the scenario you just read. Suppose the basic details remain the same:

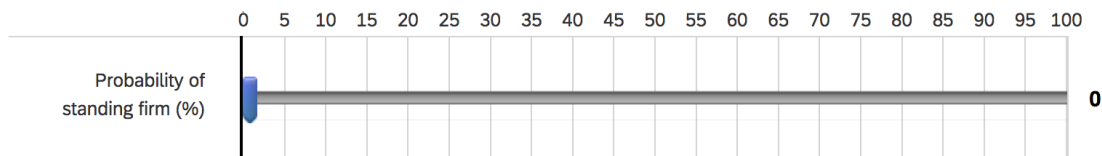
- Your country—the United Kingdom—is involved in a dispute with Country B
- The dispute began with a collision between a US shipping vessel and a ship registered to Country B
- During the collision, injuries were reported on both sides
- Additionally, both countries maintain that their ship was carrying sensitive military technology and are suspicious of the motives of the other side, leading to a tense standoff at sea
- The United Kingdom’s military forces are estimated to be about 2.5 times stronger than Country B’s
- Country B is a democracy

But this time, suppose that...

During its most recent major international crisis, Country B did not back down and the crisis led to war. According to most impartial observers, Country B did not achieve its aims, and the war led to the deaths of thousands of Country B’s soldiers and cost Country B a significant amount of money. According to public polling, about 70% of Country B’s people believe that the war was not worth it and that the experience should make Country B more cautious about using military force in the future. Country B’s new leader—who was elected after this conflict—has also said the war was “mistake” and that the country should focus more on domestic issues going forward.

War-Weariness Dependent Variable Measure

Given the information available, what is your best estimate about whether Country B will stand firm in this dispute, ranging from 0% to 100%?



Open-Ended Question

Please explain why your estimates of Country B’s probability of standing firm did or did not differ between the first and second scenario.

¹⁰Half of respondents are randomly assigned to this scenario.

References

1. Arnold, Anthony. 1993. *The Fateful Pebble: Afghanistan's Role in the Fall of the Soviet Empire*. Presidio Press.
2. Baser, Ekrem T. 2024. "Reputations and Change in International Relations." *International Studies Quarterly*.
3. Bennett, Andrew. 2005. "The Guns That Didn't Smoke: Ideas and the Soviet Non-Use of Force in 1989." *Journal of Cold War Studies* 7(2), 81-109.
4. Bloom, Jakes. 2004. *Seeing Through the Eyes of the Polish Revolution: Solidarity and the Struggle Against Communism in Poland*. Chicago, IL: Haymarket Books.
5. Chafetz, Glenn R. 1993. *Gorbachev, Reform, and the Brezhnev Doctrine: Soviet Policy Toward Eastern Europe, 1985-1990*. Westport, CT: Praeger.
6. Chen, Jidong, Jennifer Pan, and Yiqing Xu. 2016. "Sources of Authoritarian Responsiveness: A Field Experiment in China." *American Journal of Political Science* 60(2), 383-400.
7. Coppock, Alexander, and Oliver A. McClellan. 2019. "Validating the Demographic, Political, Psychological, and Experimental Results Obtained from a New Source of Online Survey Respondents." *Research & Politics* 6(1), 1-14.
8. Dafoe, Allan, Baobao Zhang, and Devin Caughey. 2018. "Information Equivalence in Survey Experiments." *Political Analysis* 26(4): 399-416.
9. Daggett, Stephen. 2010. "Costs of Major U.S. Wars." *Congressional Research Service*, 29 June.
10. DeBruyne, Nese. 2018. "American War and Military Operations Casualties: Lists and Statistics." *Congressional Research Service*, 14 September.
11. Downes, Alexander B. 2021. *Catastrophic Success Why Foreign-Imposed Regime Change Goes Wrong*. Cornell University Press.
12. Flores-Macías, Gustavo A., and Sarah E. Kreps. 2017. "Borrowing Support for War: The Effect of War Finance on Public Attitudes toward Conflict." *Journal of Conflict Resolution* 61(5), 997-1020.
13. Herrmann, Richard K., Phillip E. Tetlock, and Penny S. Viser. 1999. "Mass Public Decisions to Go to War: A Cognitive-Interactionist Framework." *American Political Science Review* 93(3), 553-573.
14. Imai, Kosuke, Luke Keele, and Dustin Tingley. 2010. "A General Approach to Causal Mediation." *Psychological Methods* 15(4), 309-334.
15. Imai, Kosuke, Luke Keele, and Teppei Yamamoto. 2010. "Identification, Inference, and Sensitivity Analysis for Causal Mediation Effects." *Statistical Sciences* 25(1), 51-72.

16. Jentleson, Bruce W. 1992. "The Pretty Prudent Public: Post Post-Vietnam America Opinion on the Use of Military Force." *International Studies Quarterly* 36(1): 49-73.
17. Jones, Calvert W. 2019. "Adviser to the King: Experts, Rationalization, and Legitimacy." *World Politics* 71(1): 1-43.
18. Kertzer, Joshua D. 2016. *Resolve in International Politics*. Princeton University Press.
19. 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): 234-49.
20. Krosnick, Jon A. 1999. "Survey Research." *Annual Review of Psychology* 50, 537–567.
21. Lazarsfeld, Paul F. 1944. "The Controversy Over Detailed Interviews—An Offer for Negotiation." *Public Opinion Quarterly* 8(1), 38–60.
22. Li, Xiaojun, and Dingding Chen. 2021. "Public Opinion, International Reputation, and Audience Costs in an Authoritarian Regime." *Conflict Management and Peace Science* 38(5), 543-560.
23. Lin-Greenberg, Erik. 2019. "Backing Up, Not Backing Down: Mitigating Audience Costs through Policy Substitution." *Journal of Peace Research* 56(4), 559-574.
24. Lueders, Hans. 2022. "Electoral Responsiveness in Closed Autocracies: Evidence from Petitions in the former German Democratic Republic." *American Political Science Review* 116(3), 827-842.
25. MacDonald, Paul K., and Joseph M. Parent. 2018. *Twilight of the Titans: Great Power Decline and Retrenchment*. Ithaca, NY: Cornell University Press.
26. Meng, Tianguang, Jennifer Pan, and Ping Yang. 2017. "Conditional Receptivity to Citizen Participation: Evidence From a Survey Experiment in China." *Comparative Political Studies* 50(4), 399-433.
27. Mercer, Jonathan. 1996. *Reputation and International Politics*. Cornell University Press.
28. Peyton, Kyle, Gregory A. Huber, and Alexander Coppock. 2021. "The Generalizability of Online Experiments Conducted During The COVID-19 Pandemic." *Journal of Experimental Political Science*.
29. Press, Daryl G., Scott D. Sagan, and Benjamin A. Valentino. 2013. "Atomic Aversion: Experimental Evidence on Taboos, Traditions, and the Non-Use of Nuclear Weapons." *American Political Science Review* 107(1), 188-206.
30. Reilly, James. 2011. *Strong Society, Smart State: The Rise of Public Opinion in China's Japan Policy*. New York, NY: Columbia University Press.

31. Renshon, Jonathan, Allan Dafoe, and Paul Huth. 2018. "Leader Influence and Reputation Formation in World Politics." *American Journal of Political Science* 62(2), 325-339.
32. 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):1064–82.
33. Schwartz, Joshua A., and Christopher W. Blair. 2020. "Do Women Make More Credible Threats? Gender Stereotypes, Audience Costs, and Crisis Bargaining." *International Organization* 74(4), 872-895.
34. Splidsboel-Hansen, Flemming. 2002. "Past and Future Meet: Aleksandr Gorchakov and Russian Foreign Policy." *Europe-Asia Studies* 54(3), 377-396.
35. Stapleton, Ian Bradford. 2014. "The Strategic Consequences of Military Quagmires: An Examination of War-Weariness Theory." *PhD Dissertation*.
36. Trotta, Daniel. 2013. "Iraq War Costs More than \$2 Trillion: Study." *Reuters*, 14 March.
37. Truex, Rory. 2016. *Making Autocracy Work: Representation and Responsiveness in Modern China*. Cambridge, UK: Cambridge University Press.
38. Tuminez, Astrid. 2000. *Russian Nationalism Since 1856: Ideology and the Making of Foreign Policy*. Lanham, Md: Rowman and Littlefield.
39. Weeks, Jessica L. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62(1): 35-64.