Supplementary Appendix (Online)

A. Survey Instrument

Demographic Module

[Age Measure 1–5]

1. Into which age range do you fall?

- 18 to 29 (1)
- 30 to 44 (2)
- 45 to 59 (3)
- 60 to 74 (4)
- 75 or older (5)

[Gender Measure 0/1]

2. What is your gender? [Order Randomized]

- Male (0)
- Female (1)
- Other (0)

[Race Measure 0/1]

3. What racial/ethnic group best describes you? [Order Randomized]

- White or Caucasian (Proceed to Question #4)
- Black or African American (0)
- Asian or Pacific Islander (0)
- American Indian or Alaska Native (0)
- Mixed Race (0)

[If White or Caucasian.]

4. Do you describe yourself as Hispanic/Latino/Latina/Latinx? [Order Randomized]

- Yes (0)
- No (1)

[Education Measure 1–5]

5. What is your highest degree or level of educational attainment?

- No schooling completed (1)
- Grades 1 through 11 (1)
- 12th grade—no diploma (1)
- High school diploma (2)
- GED or alternative credential (2)
- Some college—no degree (3)
- Associates Degree (for example: AA, AS) (3)
- Bachelor's Degree (for example: BA, BS) (4)
- Master's Degree (for example: MA, MS, MEng, MEd, MSW, MBA) (5)
- Professional degree beyond bachelor's degree (for example: MD, DDS, DVM, LLB, JD) (5)
- Doctorate degree (for example, PhD, EdD) (5)

[Income Measure 1–5]

6. Which of these describes your income last year?

- Less than $25,000 (1)
- $25,000—$49,999 (2)
- $50,000—$74,999 (3)
- $75,000—$99,999 (4)
- $100,000 or more (5)

[Attentiveness Check 1. See Figure A.1 in Appendix Section C.]

[Partisanship Measures D: 0/1; R: 0/1]

7. Generally speaking do you identify yourself as a democrat, a republican, or some other party? [Order Randomized]

- Democrat (D: 0/1)
- Republican (R: 0/1)
- Independent (D/R: 0)
- Other Party (D/R: 0)

Survey Vignette Scenario & Policy Preferences

There is much concern these days about the spread of nuclear weapons. We are going to describe a situation the U.S. could face in the future. For scientific validity the situation is general, and is not about a specific country in the news today. Some parts of the description may strike you as important; other parts may seem unimportant. After describing the situation, we will ask your opinion about a policy option.

Consider the following scenario carefully:

A rogue state is developing nuclear weapons and will have its first nuclear bomb within six months. The country has already refused multiple requests to halt its nuclear program. The country's motives remain unclear, but if it builds nuclear weapons, it will have the power to blackmail or destroy other countries.

Additional details include:

- The rogue state in question has $1/10^{th}$ the military power of the U.S.
- A pre-emptive military strike by the U.S. on the country's nuclear development facilities would be successful in preventing the rogue state from building a nuclear weapon.
- The country in question is not a democracy and shows no sign of becoming a democracy.
- The country in question is led by a single leader – pictured here:

- The country's leader does not have unchallenged authority. Powerful domestic actors have urged the leader to take a more hard-line approach to relations with the United States.
- The leader has repeatedly referred to the United States as a "threat."

1. Given the information provided, would you support a pre-emptive strike on the country in question?

- Yes (1)
- No (0)

2. If the U.S. attacked the country in question, do you believe that the U.S. military would suffer high casualties?

- Yes (1)
- No (0)

3. If the U.S. attacked the country in question, do you believe that it would damage the United States' relations with its allies?

- Yes (1)
- No (0)

4. If the U.S. attacked the rogue state's nuclear facilities, do you believe that the rogue state would attack the United States or its allies?

- Yes (1)
- No (0)

5. Do you believe that a strike would prevent the rogue state from building nuclear weapons in the long-term?

- Yes (1)
- No (0)

Toronto Empathy Questionnaire (0–64 Score)

[Attentiveness Check 2. See Figure A.2 in Appendix Section C.]

1. When someone else is feeling excited, I tend to get excited too.

- Never (0)
- Rarely (1)
- Sometimes (2)
- Often (3)
- Always (4)

2. Other people's misfortunes do not disturb me a great deal.

- Never (4)
- Rarely (3)
- Sometimes (2)
- Often (1)
- Always (0)

3. It upsets me to see someone being treated disrespectfully.

- Never (0)
- Rarely (1)
- Sometimes (2)
- Often (3)
- Always (4)

4. I remain unaffected when someone close to me is happy.

- Never (4)
- Rarely (3)
- Sometimes (2)
- Often (1)
- Always (0)

5. I enjoy making other people feel better.

- Never (0)
- Rarely (1)
- Sometimes (2)
- Often (3)
- Always (4)

6. I have tender, concerned feelings for people less fortunate than me.

- Never (0)
- Rarely (1)
- Sometimes (2)
- Often (3)
- Always (4)

7. When a friend starts to talk about his\her problems, I try to steer the conversation towards something else.

- Never (4)
- Rarely (3)
- Sometimes (2)
- Often (1)
- Always (0)

8. I can tell when others are sad even when they do not say anything.

- Never (0)
- Rarely (1)
- Sometimes (2)
- Often (3)
- Always (4)

9. I find that I am “in tune” with other people’s moods.

- Never (0)
- Rarely (1)
- Sometimes (2)
- Often (3)
- Always (4)

10. I do not feel sympathy for people who cause their own serious illnesses.

- Never (4)
- Rarely (3)
- Sometimes (2)
- Often (1)
- Always (0)

11. I become irritated when someone cries.

- Never (4)
- Rarely (3)
- Sometimes (2)
- Often (1)
- Always (0)

12. I am not really interested in how other people feel.

- Never (4)
- Rarely (3)
- Sometimes (2)
- Often (1)
- Always (0)

13. I get a strong urge to help when I see someone who is upset.

- Never (0)
- Rarely (1)
- Sometimes (2)
- Often (3)
- Always (4)

14. When I see someone being treated unfairly, I do not feel very much pity for them.

- Never (4)
- Rarely (3)
- Sometimes (2)
- Often (1)
- Always (0)

15. I find it silly for people to cry out of happiness.

- Never (4)
- Rarely (3)
- Sometimes (2)
- Often (1)
- Always (0)

16. When I see someone being taken advantage of, I feel kind of protective towards him\her.

- Never (0)
- Rarely (1)
- Sometimes (2)
- Often (3)
- Always (4)

B. Sample Power Considerations

In the primary text we discuss how one of the weaknesses of our study is its small sample size. It is then only natural to wonder what the probability of detecting a "true" effect when it exists might be in our study. To alleviate concerns, we calculate the logit power analysis for the naïve model using TEQ scores as single continuous predictors. Since Models 1 through 4 in the main paper all report similar results, showing sufficient power for the naïve model should be sufficient to reassure researchers that our results are broadly valid.

We calculate the sample sizes required to avoid Type II Error rates at different power levels in Table A.1 using the Stata powerlog package from Friendly (1999). When the probability of a Type I error is 10%, our sample size of 121 has roughly an 87% (0.87) chance of avoiding a Type II error. Under these criteria, our sample size demonstrates sufficient power for our results to be broadly acceptable to the academic community at the 90% level (Cohen, 1988). We, therefore, believe our sample to be sufficiently large to be worthy of interpretation as statistically significant at the 90% level without excessive risk of a Type II error.

At the 95% level, however, our sample size of 121 can only be said to avoid a Type II Error with 77% (0.77) certainty. This is below the 80% threshold usually required for academic papers (Cohen, 1988). Still, given just how close it is to said threshold (off

only by 3%), we feel that it is still worth publishing our paper's results and allowing other researchers to determine how seriously to consider them at the 95% level.

Table A.1 Sample Sizes Required to Avoid a Type II Error at Different Power Levels

	n Required	
Power	$\alpha = 0.1$	$\alpha = 0.05$
0.75	81	115
0.76	84	118
0.77	87	121
0.78	89	125
0.79	92	128
0.80	95	132
0.81	99	135
0.82	102	139
0.83	106	143
0.84	109	148
0.85	113	152
0.86	118	157
0.87	122	163
0.88	127	168
0.89	132	174
0.90	138	181

C. Attentiveness Check & Sample Cleaning Considerations

As noted in the main treatise, three steps were taken to ensure response quality on Mechanical Turk (MTurk) during sample collection: (1) Respondents were required to be masters; (2) respondents were required not to share IP addresses; and (3) respondents were required to pass two visual attentiveness checks. No studies to date indicate that the first of these three requirements might affect sample quality negatively, and while some evidence exists that deleting duplicate IP addresses may be insufficient to protect sample quality independently of other security measures, there is little compelling evidence that deleting duplicates reduces sample quality in a non-random manner. Yet there is some evidence that our third security measure, attentiveness checks may skew results when poorly administered. We, therefore, seek to address concerns here by defending our method on four grounds.

First, because we used images that asked for respondents to complete very simple

counting tasks, we believe that only bots, which are unable to read Portable Network Graphics (PNG) image files, could fail our attentiveness checks—especially since MTurk workers have historically outperformed other populations such as college students on attentiveness checks (Hauser and Schwarz, 2016). Our first screening question (Figure A.1) asked respondents to count dots in a picture. We presuppose that is should not be difficult for any human being to count to ten. Our second attentiveness check (Figure A.2) asked respondents how many triangles were shown in a picture, and both four and five were acceptable answers. Once again, we expect almost every respondent to be able to count triangles.

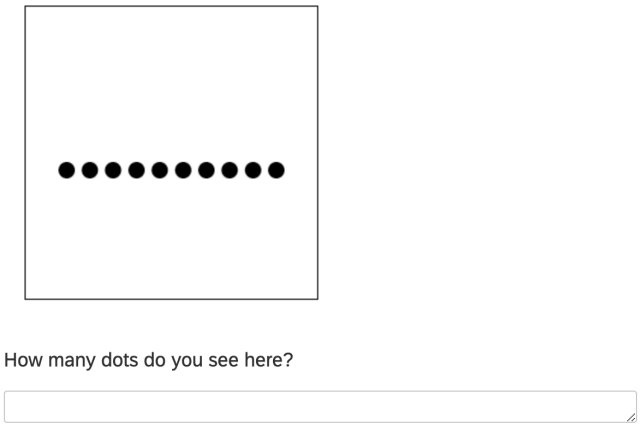

Figure A.1. Attentiveness Question 1

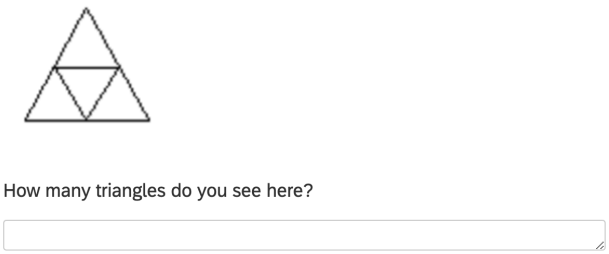

Figure A.2. Attentiveness Question 2

Second, while some scholars (e.g. Berinsky, Margolis, and Sances, 2014) have found that attentiveness check passage rates correlate with select political variables such as race and education, we control for these demographic variables in regression models 2 through 4 in the main text. Since models 2 through 4 show no significant divergence from

the naïve Model 1, we would suggest that demographic confounding – assuming it exists – does not have a significant effect on the treatise's primary finding.

Third, we would suggest that the attentiveness checks that scholars typically take issue with are much more complex than those checks that are presented in our survey. For instance, Berinsky, Margolis, and Sances (2014, 740) criticizes attentiveness checks for biasing respondents. However, the checks to which that study makes reference are significantly more complex than those presented in our survey. Consider the lengthy prompt to one attentiveness check in Berinsky, Margolis, and Sances (2014, 740):

> "When a big news story breaks people often go online to get up-to-the-minute details on what is going on. We want to know which websites people trust to get this information. We also want to know if people are paying attention to the question. To show that you've read this much, please ignore the question and select ABC News and The Drudge Report as your two answers."

After this lengthy prompt, the study then presents the question in Figure A.3.

When there is a big news story, which is the one news website you would visit first? (Please only choose one)

☐ New York Times website	☐ The Drudge Report	☐ The Associated Press (AP) website
☐ Huffington Post	☐ Google News	☐ Reuters website
☐ Washington Post website	☐ ABC News website	☐ National Public Radio (NPR) website
☐ CNN.com	☐ CBS News website	☐ USA Today website
☐ FoxNews.com	☐ NBC News website	☐ New York Post Online
☐ MSNBC.com	☐ Yahoo! News	☐ None of these websites

Figure A.3. Example Attentiveness Check from Berinsky, Margolis, and Sances (2014, 740)

Even a cursory glance at Figure A.3 in relation to Figures A.2 and A.1 (our attentiveness checks) makes it plainly apparent that our survey's attentiveness checks are less demanding than those criticized by Berinsky, Margolis, and Sances (2014). Our prompts and tasks are both shorter in length and simpler to complete. We thus suggest that general criticisms of attentiveness checks are unlikely to apply to our specific study.

Finally, we re-run our analysis from Table 1 in the main text here using our full sample. This follows the advice of Berinsky, Margolis, and Sances (2014). We find that the inclusion of those who failed the attentiveness checks does not fundamentally alter our main findings regardless of model specification. However, their inclusion does make some demographic variables, such as gender and partisanship, that we might expect to be associated with either an increased or decreased willingness to strike (Barnhart et al., 2020; Maxey, 2022) no longer statistically significant. Meanwhile, age is now the only statistically significant demographic predictor of willingness to strike. Given that there is little theoretical support for these changes in the existing literature, we believe that our simple image-based attentiveness checks are actually playing a vital role in preventing MTurk bots from biasing our findings. As such, we do believe that the attentiveness checks should be included even if their inclusion risks some small loss of statistical power.

Table A.2. Primary Results Reproduced Without Attentiveness Checks Removed

	strike			
	(1)	(2)	(3)	(4)
teq	−0.056***	−0.045**	−0.094***	−0.078**
	(0.019)	(0.021)	(0.031)	(0.033)
white		0.499	0.437	0.754
		(0.429)	(0.443)	(0.521)
age2		−0.422*	−0.459**	−0.491**
		(0.224)	(0.222)	(0.239)
dem		−0.017	−0.016	−0.007
		(0.471)	(0.479)	(0.546)
rep		0.434	0.580	0.571
		(0.554)	(0.577)	(0.647)
education2		0.002	0.098	−0.020
		(0.265)	(0.277)	(0.310)
female		0.175	−3.803**	−2.769
		(0.437)	(1.722)	(1.886)
I(teq ∗female)			0.098**	0.076*
			(0.041)	(0.045)
believe_casualties				0.604
				(0.519)
hurt_relations				−1.024**
				(0.520)
counterattack				−0.113
				(0.542)
prevent_prolif_longterm				1.748***
				(0.464)
N	136	136	136	136
Log Likelihood	−85.174	−82.530	−79.577	−69.335
AIC	174.347	181.061	177.155	164.671

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

D. Appendix References

Barnhart JN, Trager RF, Saunders EN, Dafoe A (2020) The Suffragist Peace. *International Organization* 74(4): 633–670.

Berinsky AJ, Margolis MF, and Sances MW (2014) Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-administered Surveys. *American Journal of Political Science* 58(3): 739–753.

Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillside, NJ: Lawrence Erlbaum Associates.

Friendly M (1999). Visualizing Categorical Data. pp. 319–438 in *Cognition and Survey Research*, edited by M Sirken, D Herrmann, S Schechter, et al. New York, NY: John Wiley and Sons.

Hauser DJ and Schwarz N (2016) Attentive Turkers: Mturk Participants Perform Better On Online Attention Checks Than Do Subject Pool Participants. *Behavior Research Methods* 48: 400–407.

Maxey S (2022) Finding the water's edge: when negative partisanship influences foreign policy attitudes. *International Politics* 59(5): 802–826.