Supplementary Information

How Does Shaming Human Rights Violators Abroad Shape Attitudes at Home?

Lotem Bassan-Nygate

Contents

A Experiment I

A.1 Descriptive Statistics

Table A1: Descriptive Statistics - Survey Experiment I

Statistic	N	Mean	St. Dev.	Min	Max
Government Approval	1,354	4.167	1.978	1	7
Human Rights Respect	1,326	4.551	1.874	1	7
Oppose Torture	1,350	5.696	1.789	1	7
Legal Obligation	1,337	3.782	0.732	1.000	5.000
Morality	1,332	4.421	1.755	1	7
Power	1,332	4.339	1.710	1	7
Democrat	1,385	0.355	0.479	0	1
Independent	1,385	0.184	0.388	0	1
Republican	1,385	0.461	0.499	0	1
Female	1,385	0.544	0.498	0	1
White	1,323	0.742	0.438	0	1
Black/African American	1,323	0.104	0.306	0	1
Hispanic	1,323	0.067	0.249	0	1
Asian/Asian American	1,323	0.048	0.213	0	1
Native American	1,323	0.013	0.113	0	1
Middle Eastern	1,323	0.001	0.027	0	1
Mixed Race	1,323	0.018	0.134	0	1
Age	1,385	46.926	17.136	18	94

A.2 Main Analyses

In Table A2 I report the main models reported in the manuscript, as well as additional pre-registered models including demographic controls. Overall, my findings remain relatively similar. When examining the effect of my shaming treatment on a broader measure of support for international law (developed by Bayram (2017)), I identify a small but statistically significant positive effect, whereby shaming *increased* respondents' obligation to international law. This finding stands in contrast to H3b and the moral licensing effect I identify on opposition to torture, but is in accordance with H3a. I did not anticipate the shaming treatment to affect these two measures in opposing directions. One explanation may relate to the (in)effectiveness of moral licensing across different domains. Indeed, while there has been some evidence of cross-domain moral licensing (Mazar and Zhong 2010), recent follow-ups were unable to fully replicate this finding (Urban et al. 2019). It is thus possible that the moral licensing effect simply does not extend to the domain of international law. Since I was not able to replicate this finding in the second experiment (see table B10), future research should examine more carefully whether shaming can simultaneously shape favorable attitudes towards international law and adjacent domains.

A.3 Mediation Analysis

As described in the main text, I conduct a mediation analysis using Imai et al.'s 2010 `mediation` package. In Figure A2 , the outcome of interest is government approval, and the mediators are

Table A2: Main Models with and without (pre-treatment) Demographic Controls

	Gov Approval		HR Respect		Oppose Torture		Legal Obligation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	2.312***	2.306***	1.734***	1.721***	−0.207**	−0.198**	0.084**	0.091**
	(0.087)	(0.091)	(0.091)	(0.094)	(0.097)	(0.096)	(0.040)	(0.039)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
N	1,354	1,354	1,326	1,326	1,350	1,350	1,337	1,337

Notes: Demographic controls include: sex, age, race, ethnicity, education, ideology, and region. * $p < .1$, ** $p < .05$, *** $p < .01$

perceptions of U.S. morality (left) and power (right). The average causal mediation effect (ACME), the total effect, and the direct effect are all positive and statistically significant.

In Figure A1 I report results from a mediation analysis where the mediator is the morality indicator and the outcome is opposition to torture. Here I find that the ACME and the total effect are negative and statistically significant, suggesting that respondents who viewed shaming as moral were less likely to oppose torture. However, the direct effect is positive and approaches statistical significance.

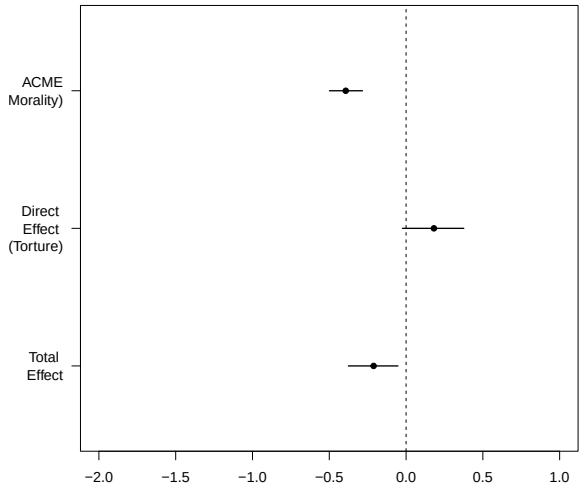

Figure A1: Causal mediation plot. Treatment is shaming manipulation, Mediator is morality (post-treatment), Outcome is opposition to torture. Horizontal lines represent 90% confidence intervals for estimates.

A.4 Sensitivity Analysis

Since the mediators in the mediation analysis conducted in Section A.3 are not randomly assigned, there is a concern for omitted variable bias that accounts for both the mediator and the outcome,

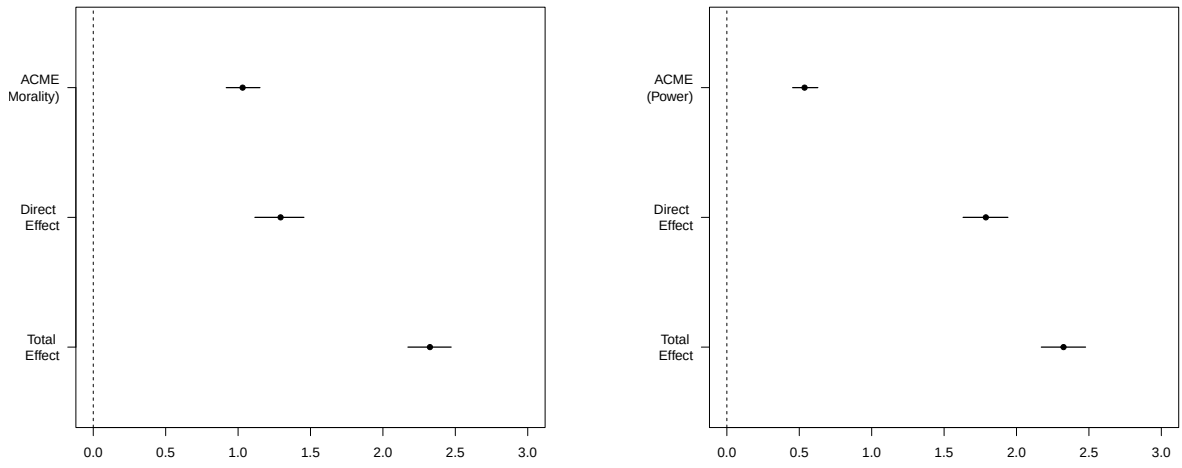

Figure A2: Causal mediation plots. In the figure on the left, morality perception is the mediator. In the figure on the right, morality perception is the mediator. In both figures, Treatment is shaming manipulation and Outcome is government approval. Horizontal lines represent 90% confidence intervals for estimates.

violating the "sequential ignorability" assumption. I thus conduct sensitivity analyses for all three mediation analyses presented in the previous section. The results, presented in Figures A3-A5, present an analysis for two cases: (a) where the omitted variable influences the outcome and the mediator in the same direction, and (b) where it influences them in opposite directions.

Figure A3 reports the sensitivity analysis where morality is the mediator and government approval is the outcome. Under the assumption that the confounder influences the mediator and the outcome in the same direction (a), the unobserved confounder must explain about 50% of the variation in morality (M) and 50% of the variation in government approval (Y) for the estimate to be zero. At higher values, the estimated average causal mediation effect will be negative, whereas at lower values the sign of the estimate remains positive. This implies that the proportion of M and Y explained by the unobserved confounder must be relatively high for the original conclusion to be reversed. Under the assumption that the confounder influences the mediator and the outcome in opposite directions, presented in Figure A3(b), the mediation effects attached to each contour line are all positive and grow stronger as more variance is explained.

Figure A4 reports the sensitivity analysis where power is the mediator and government approval is the outcome. Under the assumption that the confounder influences the mediator and the outcome in the same direction (a), the unobserved confounder must explain about 50% of the variation in power (M) and 50% of the variation in government approval (Y) for the estimate to be zero. At higher values, the estimated average causal mediation effect will be negative, whereas at lower values the sign of the estimate remains positive. This implies that the proportion of M and Y explained by the unobserved confounder must be relatively high for the original conclusion to be reversed. Moreover, under the assumption that the confounder influences the mediator and the outcome in opposite directions, presented in Figure A4(b), the mediation effects attached to each contour line are all positive and grow stronger as more variance is explained.

Finally, figure A5 reports the sensitivity analysis where morality is the mediator and torture is the outcome. Under the assumption that the confounder influences the mediator and the outcome

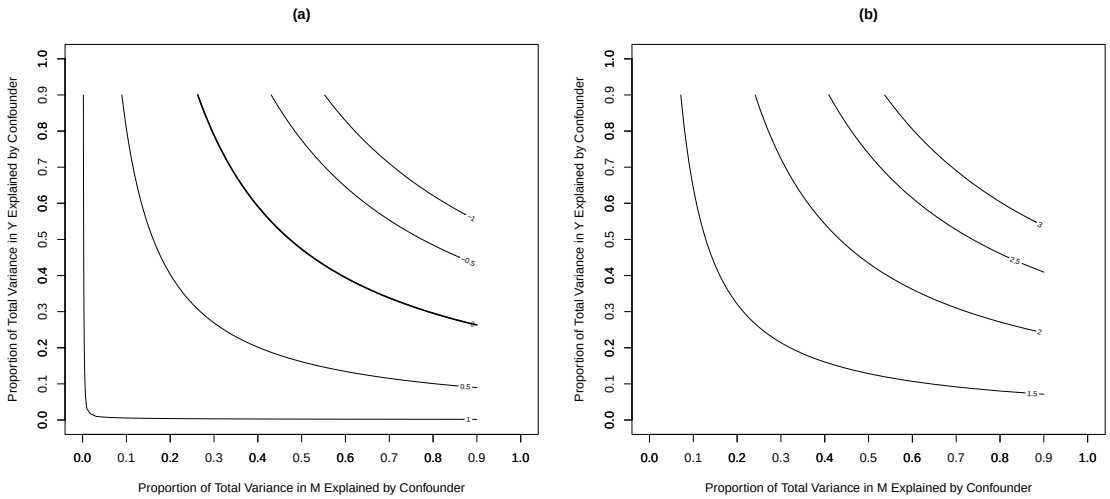

Figure A3: Sensitivity analysis for mediation (Mediator is morality, Outcome is government approval).

in the opposing direction (b), the unobserved confounder must explain about 20% of the variation in power (M) and 15% of the variation in government approval (Y) for the estimate to be zero. Under the assumption that the confounder influences the mediator and the outcome in the same directions, presented in Figure A5(a), the mediation effects attached to each contour line are all negative and grow stronger as more variance is explained.

A.5 Manipulation Check

In this section, I test whether I was able to successfully manipulate information regarding shaming the human rights violators. Overall, 90% of respondents passed the manipulation check. In Table A3 I show that the shaming treatment significantly increased the belief that the U.S. government shamed the other country in the scenario. The results hold when including a host of demographic controls.

Table A3: Manipulation Check

	Belief that the U.S. government criticized the other country	
	(1)	(2)
Treatment	1.689***	1.684***
	(0.026)	(0.027)
Demographic Controls	No	Yes
N	1,358	1,358

Notes: Demographic controls include: sex, age, race, ethnicity, education, ideology, and region. * $p < .1$, ** $p < .05$, *** $p < .01$

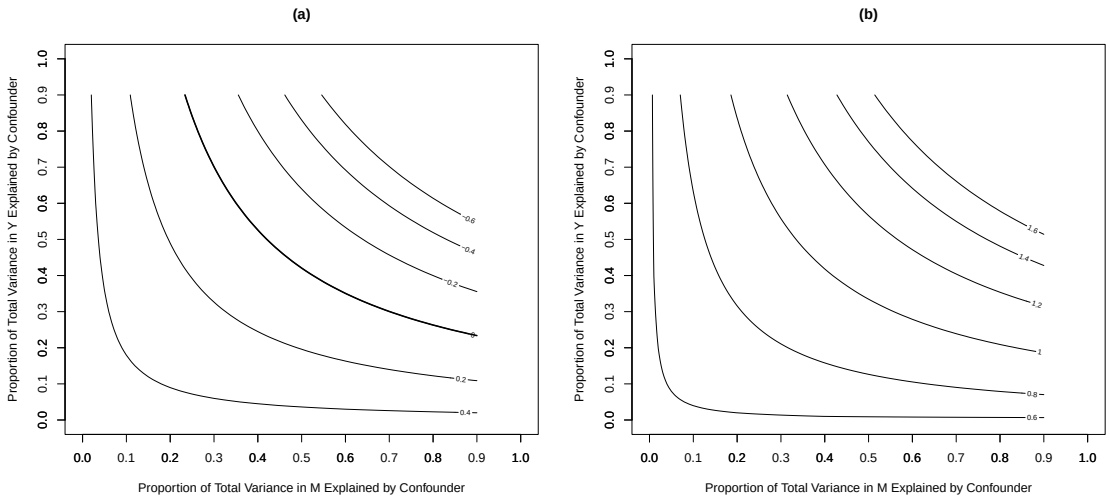

Figure A4: Sensitivity analysis for mediation (Mediator is power, Outcome is government approval).

A.6 Information Leakage

To address a concern relating to information leakage (Dafoe et al. 2018), respondents were asked whether they thought of a specific country when reading the vignette. In Table A4, I show that treated respondents were **not** more likely to think of a specific country when reading the vignette. This, together with the results from the manipulation check presented in Table A3, reduces concerns about internal validity, as it is likely that the treatment successfully manipulated perceptions about shaming, rather than another aspect associated with it.

Table A4: The effect of shaming treatment on thinking of a specific country

	Did you think of a specific country?	
	(1)	(2)
Treatment	0.013	0.015
	(0.027)	(0.027)
Demographic Controls	No	Yes
N	1,385	1,385

Notes: Demographic controls include: sex, age, race, ethnicity, education, ideology, and region. * $p < .1$, ** $p < .05$, *** $p < .01$

Another concern, however, is that while respondents did not think of specific countries differentially (i.e. by experimental condition), all respondents thought of one specific country. This concern relates to external validity, as it suggests that the results speak to one particular context and may not generalize beyond it. I take two measures to reduce concerns about this issue. First, I note that less than 50% of respondents mentioned that they thought of a specific country when reading the vignette. Second, as demonstrated in Figure A6 – of the subsample who said they thought of

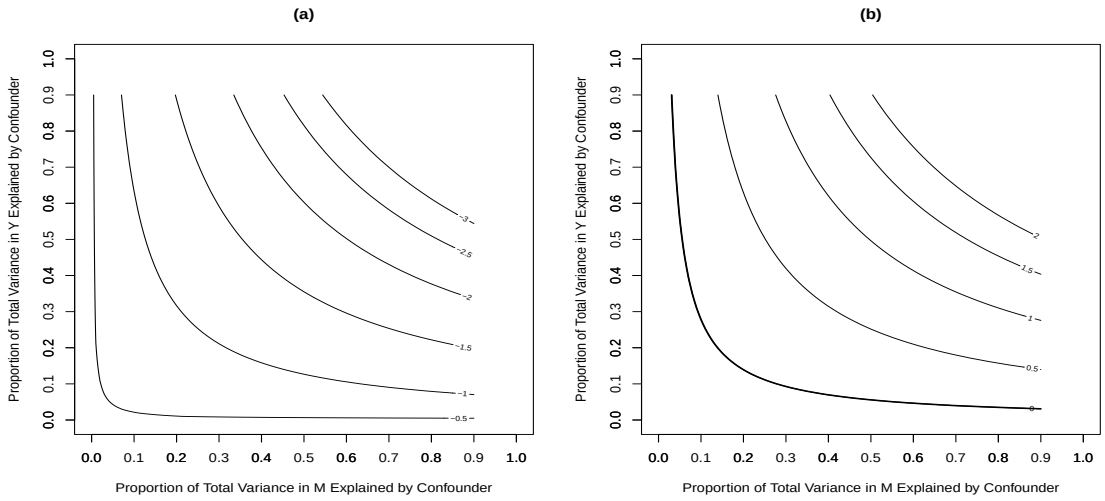

Figure A5: Sensitivity analysis for mediation (Mediator is morality, Outcome is opposing torture).

a specific country, respondents were not confined to one context but rather mentioned a host of countries, with most subjects reporting Russia and China.

A.7 Exploratory Analysis

Utilizing responses to the open-ended follow-up question described in the preceding section, I explore variation in my average treatment effect along specific target states. In Figure A7 I report the ATEs by the two main countries reported by respondents: Russia and China, other countries, and no countries. This yields interesting variation which can be further explored in future research. Interestingly, I find that while similar dynamics appear across different subgroups, the treatment effect on government approval is stronger for respondents who think of specific countries, particularly China, demonstrating perhaps the strong cross-partisan support for the condemnation of China in the U.S.

Two points should be underscored, however. First, I am most likely underpowered to detect any meaningful differences across the groups, particularly for small effect sizes like H3, although future work should certainly consider cross-pressures depending on the context and respondents' individual preferences. Second, this additional analysis should be read as suggestive and exploratory because it conditions the ATE on a post-treatment variable.

A.8 Attrition

Out of the 1,385 respondents who passed the attention check, 62 have attrited, which accounts for approximately 4.5% of the sample. Although this is a relatively small proportion, I test whether my treatment causes respondents to attrite. If my treatment was to account for attrition, this could pose a threat to inference (Coppock 2021). As reported in Figure A5, I do not find a statistically significant association between my shaming treatment and attrition.

A.9 Experiment I Survey Instrument

Italics text signifies coding text or text that was not present to respondents.

A.9.1 Pre-treatment attention check + hawkishness

An attention check was embedded within the hawkishness grid. Respondents who answered it incorrectly were removed from the survey.

Figure A6: Wordcloud of the countries respondents thought of when reading the vignette (note that only 50% thought of a specific country, and that this was not associated with treatment).

Table A5: The effect of shaming treatment on attrition

	Attrition	
	(1)	(2)
Treatment	−0.016	−0.012
	(0.011)	(0.012)
Demographic Controls	No	Yes
N	1,385	1,385

Notes: Demographic controls include: sex, age, race, ethnicity, education, ideology, and region. * $p < .1$, ** $p < .05$, *** $p < .01$

- Below, you will see a series of statements. Please tell us whether you agree or disagree with each statement.
  - In the United States, our people are not perfect, but our culture is superior to others.
  - The use of military force only makes problems worse.
  - I would rather be a citizen of America than of any other country, click disagree regardless of your opinion.
  - Going to war is unfortunate, but sometimes the only solution.
  - The best way to ensure peace is through military strength.

A.9.2 Pre-treatment covariates

- **Partisanship:** Generally thinking, do you think of yourself as a: [Republican, Democrat, Independent, Another party (fill in), No preference]

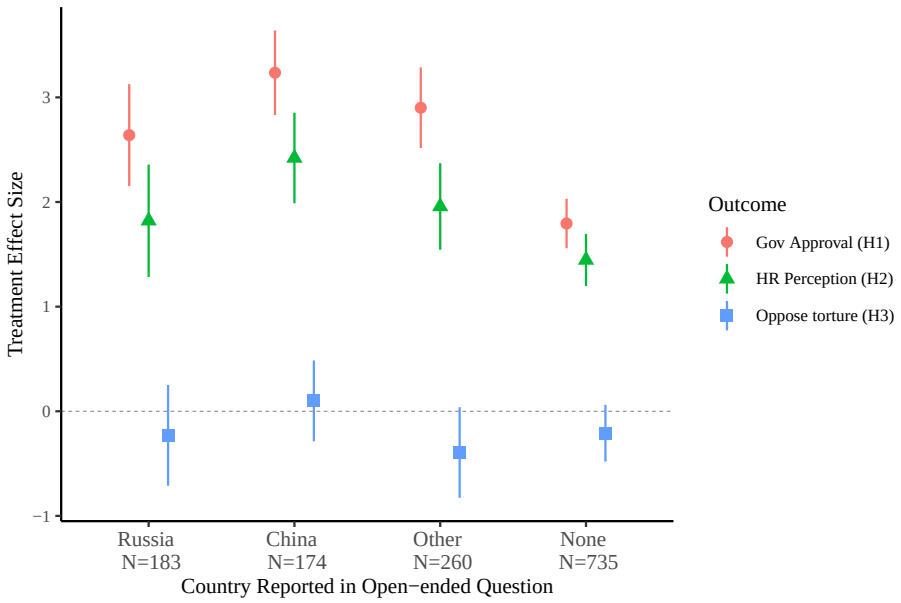

Figure A7: Average treatment effect by country reported by respondents in the open-ended question. Colors and shapes correspond to different outcome questions, relevant to each hypothesis.

- *If Republican:* Would you call yourself a: [Strong Republican, Not very strong Republican]
- *If Democrat:* Would you call yourself a: [Strong Democrat, Not very strong Democrat]
- *If neither Democrat nor Republican:* Do you think yourself as closer to the: [Republican Party, Democratic Party, Neither Party]

A.9.3 Experimental vignettes

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 administrations have handled the situation in different ways. We will describe one approach the U.S. government has taken and ask whether you approve or disapprove.

A country violates human rights; it imprisons and tortures some of its citizens because of their beliefs and silences human rights defenders and activists. The U.S. government [called the country out and publicly criticized it for its violation of human rights in the United Nations./ did not criticize the country in the United Nations or make any public statement about the country's human rights abuses.] The country continued to violate human rights.

A.9.4 Manipulation checks

- In the passage you read, did the other country violate human rights? [yes, no, not specified]
- In the passage you read, did the U.S. government criticize the other country? [yes, no, not specified]

A.9.5 Outcomes

- **Government approval:** How much do you approve or disapprove of the way the U.S. government handled the situation? [Strongly approve/Approve/Slightly approve/ Neutral/Slightly disapprove/Disapprove/Strongly disapprove]

- **Torture:** How much do you support or oppose government use of practices like torture that violate human rights **in the U.S.**? [Strongly support/Support/Slightly support/ Neutral / Slightly oppose / Oppose /Strongly oppose]
- **International legal obligation:** How strongly do you agree or disagree with the following statements: [Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree]
  - It is important to me personally that the US will comply with international law.
  - Complying with international law is an important value.
  - Complying with international law is important, even if it contradicts the national interest.
  - I feel uncomfortable when the US violates international laws and norms.
  - If the US defies international laws and norms, criticism from other countries is justified.
- Many things may be desirable, but not all of them are essential characteristics of democracy. How essential is each of the following things as a characteristic of democracy? Use this scale where 1 means "not at all an essential characteristic of democracy" and 10 means it definitely is "an essential characteristic of democracy"
  - Civil rights protect people's liberty against oppression.
  - Women have the same rights as men.
- **Morality:** In the passage you read, how moral and ethical do you believe the U.S. government to be? [Very moral/ moral/ slightly moral/ neutral/ slightly immoral/ immoral/ very immoral] *This was recoded such that higher variables indicate more moral)*
- **Power:** In the passage you read, how powerful or weak do you believe the U.S. government to be? [Very powerful/ powerful/ slightly powerful/ neutral/ slightly weak/ weak/ very weak] *This was recoded such that higher variables indicate more power)*
- **HR perception:** In the passage you read, how much does the U.S. government value human rights? Please respond on a scale of 1 to 7, where 1 indicates the U.S. doesn't value human rights at all, and 7 indicates it values human rights a lot.
- **Hypocrisy:** In the passage you read, how hypocritical would it be if the U.S. violated human rights itself? Please respond on a scale of 1 to 7, where 1 indicates the not hypocritical at all, and 7 indicates very hypocritical.
- **Placebo:** Did you think of a specific country when you read about the "other country" in the passage? If so, please specify. [Yes (open end), No]

A.9.6 Post-treatment covariates

Note that I collected these to ensure representativeness, but Lucid provides demographic metadata on all respondents, which was used as demographic controls in the models reported in Table A2.

- What is your age?
- Are you [Male/ Female /Other (insert answer)]
- What racial or ethnic group best describes you? [White, Black or African American, Hispanic or Latino, Asian or Asian American, Native American, Middle Eastern, Mixed race, Other (insert answer)]

- What is your highest level of education? [Less than high school / High school graduate/ Some college but no degree / Bachelor's / Master's / Doctoral / Professional (JD or MD)]
- What is your state of residence?

B Experiment II

B.1 Descriptive Statistics

Table B6: Descriptive Statistics - Survey Experiment II

Statistic	N	Mean	St. Dev.	Min	Max
Government Approval	3,054	4.299	1.840	1	7
Human Rights Respect	3,011	4.385	1.792	1	7
Oppose Torture	3,043	5.450	1.891	1	7
Legal Obligation	3,030	3.739	0.734	1.000	5.000
Morality	3,018	4.428	1.692	1	7
Power	3,018	4.505	1.649	1	7
Democrat	2,929	0.410	0.492	0	1
Independent	2,929	0.289	0.453	0	1
Republican	2,929	0.301	0.459	0	1
Female	3,144	0.519	0.500	0	1
White	2,998	0.719	0.449	0	1
Black/African American	2,998	0.116	0.321	0	1
Hispanic	2,998	0.075	0.264	0	1
Asian/Asian American	2,998	0.047	0.212	0	1
Native American	2,998	0.009	0.094	0	1
Middle Eastern	2,998	0.001	0.037	0	1
Mixed Race	2,998	0.024	0.152	0	1
Age	3,144	45.587	17.069	18	98

B.2 Pre-registered Models

Tables B7-B10 present the results of the main models. A detailed discussion of these findings is reported in the manuscript.

Table B7: Treatment effects on government approval (H1)

	Government Approval			
	(1)	(2)	(3)	(4)
Shaming treatment	1.442***	1.444***	1.478***	1.363***
	(0.061)	(0.061)	(0.060)	(0.085)
Ally treatment		−0.428***	−0.449***	−0.564***
		(0.061)	(0.060)	(0.085)
Shaming*Ally				0.230*
				(0.120)
Demographic Controls	No	No	Yes	Yes
N	3,054	3,054	3,054	3,054

Notes: Demographic controls include: sex, age, race, ethnicity, education, ideology, and region. * $p < .1$, ** $p < .05$, *** $p < .01$

Table B8: Treatment effects on human rights perceptions (H2)

	Human Rights Perceptions			
	(1)	(2)	(3)	(4)
Shaming treatment	0.995***	0.996***	1.002***	0.906***
	(0.063)	(0.062)	(0.062)	(0.088)
Ally treatment		−0.393***	−0.405***	−0.501***
		(0.062)	(0.062)	(0.088)
Shaming*Ally				0.192
				(0.124)
Demographic Controls	No	No	Yes	Yes
N	3,011	3,011	3,011	3,011

Notes: Demographic controls include: sex, age, race, ethnicity, education, ideology, and region. * $p < .1$, ** $p < .05$, *** $p < .01$

Table B9: Treatment effects on opposing torture (H3)

	Oppose Torture			
	(1)	(2)	(3)	(4)
Shaming	−0.079	−0.079	−0.110*	−0.201**
	(0.069)	(0.069)	(0.064)	(0.091)
Ally		0.003	0.016	−0.075
		(0.069)	(0.064)	(0.091)
Shaming*Ally				0.182
				(0.128)
Demographics	No	No	Yes	Yes
N	3,043	3,043	3,043	3,043

Notes: Demographic controls include: sex, age, race, ethnicity, education, ideology, and region. * $p < .1$, ** $p < .05$, *** $p < .01$

Table B10: Treatment effects on international law (pre registered)

	International Legal Obligation			
	(1)	(2)	(3)	(4)
Shaming	0.016	0.016	0.017	0.043
	(0.027)	(0.027)	(0.026)	(0.036)
Ally		0.019	0.015	0.041
		(0.027)	(0.026)	(0.036)
Shaming*Ally				−0.052
				(0.051)
Demographics	No	No	Yes	Yes
N	3,030	3,030	3,030	3,030

Notes: Demographic controls include: sex, age, race, ethnicity, education, ideology, and region. * $p < .1$, ** $p < .05$, *** $p < .01$

B.3 Supplemantary analyses to Explore Moral Licensing

Figure 6 reveals that the shaming treatment is effective in decreasing opposition to torture for respondents in the 'non-ally' condition, but not in the 'ally' condition. Here, I explore why shaming would decrease opposition to torture (moral licensing) when the target is a non-ally, but not when the target is an ally of the U.S. My theoretical framework suggests that shaming could increase tolerance for problematic attitudes when it establishes 'moral credentials' and reinforces beliefs about the morality of one's ingroup. If shaming decreases opposition of torture by increasing moral credentials, we should expect respondents in the 'non-ally' condition to perceive shaming as more moral than respondents in the 'ally' condition.

Figure B8 confirms this assumption and shows that shaming of non-allies was perceived as more moral than shaming of allies. In Table B11 I report the effect of the 'ally' condition (compared to 'non-ally') on perceptions of U.S. morality, showing that respondents in the ally condition were less likely to perceive the U.S. as moral ($\beta = -0.4, p < 0.001$). While I cannot definitively say, this effect may be driven in part by respondents' disappointment in their government for having allies that violate human rights in the scenario. As for the small effect size, Figure B9 demonstrates that respondents in both conditions were compelled to object torture, suggesting that my survey design may have introduced social desirability bias that created a harder test for my theory.

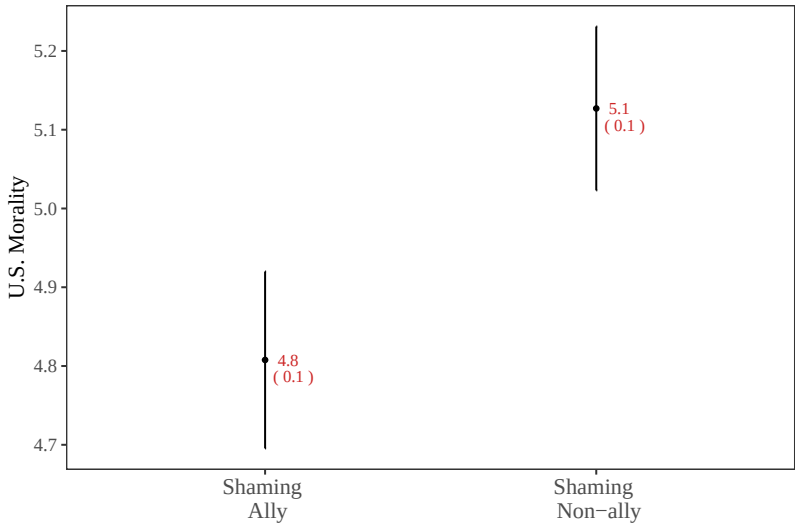

Figure B8: The X-axis marks experimental conditions. Y-axis marks the level of perception of U.S. morality, scaled from 1-7. To the right of the corresponding circles, means are marked in red and standard errors are in parentheses below.

Next, I explore whether the effect of shaming on torture is driven by beliefs about morality more systematically. Specifically, I utilize Imai et al.'s 2010 `mediation` package. The result of this analysis is depicted in Figure B10, where the mediator is perceptions of morality, the treatment is shaming, and the outcome is opposition to the government's use of torture. I find evidence that the average causal mediation effect (ACME) is negative and statistically significant. Interestingly, the design of the second experiment (unintentionally) randomizes levels of the mediator independently from the shaming treatment through the 'ally/non-ally' treatment arm. This results in a more robust test of the downstream effect of the mediator on the outcome (Acharya et al. 2018), when compared to the results of the mediation in study I (see Figure A1). Nonetheless, both analyses reveal similar results.

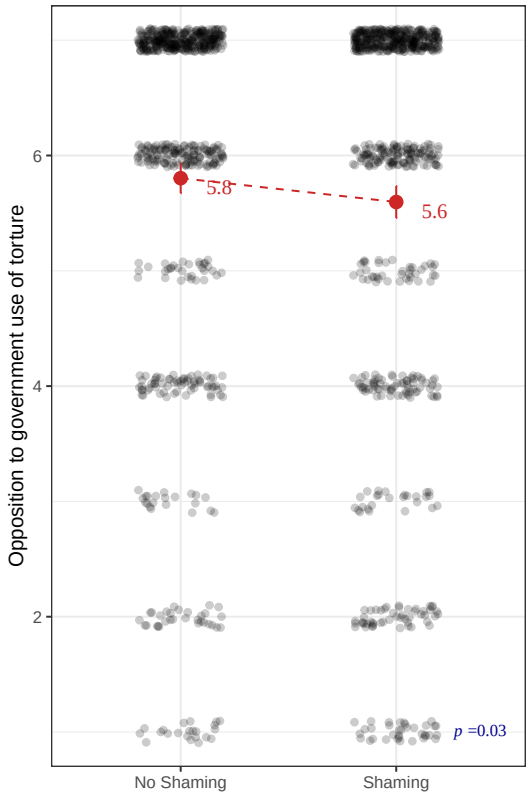

Figure B9: Distribution of responses to the outcome question measuring opposition to the U.S. government's use of torture, by treatment condition (Study I). To the right of the corresponding circles, means are marked in red.

Table B11: The Effect of Ally-shaming on Perceptions of U.S. Morality

	Government Approval	
	(1)	(2)
Ally	-0.418^{***}	-0.433^{***}
	(0.061)	(0.061)
Demographic Controls	No	Yes
N	3,018	3,018

Notes: Demographic controls include: sex, age, race, ethnicity, education, ideology, and region. * $p < .1$, ** $p < .05$, *** $p < .01$

B.4 Manipulation Check

In this section, I test whether I was able to successfully manipulate the information randomized in the vignettes. Overall, 87% of respondents passed the shaming manipulation check and 86% of respondents passed the target manipulation check. In Table B12 I show that the shaming treatment significantly increased the belief that the U.S. government shamed the other country in the scenario and that the ally treatment significantly increased the belief that the other country in the scenario

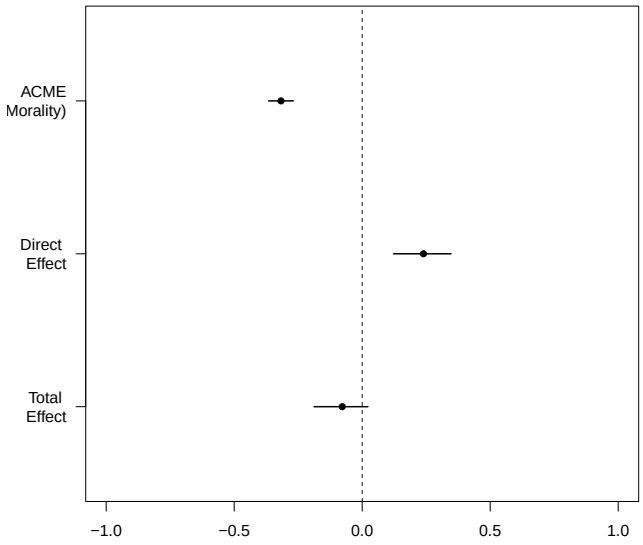

Figure B10: **Causal mediation plot.** Morality is the mediator, shaming manipulation is treatment and Outcome is opposition of torture. Horizontal lines represent 90% confidence intervals for estimates. Analysis includes the following pretreatment demographic controls: gender, age, education, ethnicity, party ID, and residence.

is a US ally. The results remain similar when including a host of demographic controls.

Table B12: Manipulation Check

	US criticized the other country		Other country is a US ally	
	(1)	(2)	(3)	(4)
Shaming treatment	1.596***	1.597***		
	(0.020)	(0.020)		
Ally treatment			1.576***	1.572***
			(0.020)	(0.020)
Demographic Controls	No	Yes	No	Yes
N	3,062	3,062	3,060	3,060

Notes: Demographic controls include: sex, age, race, ethnicity, education, ideology, and region. * $p < .1$, ** $p < .05$, *** $p < .01$

B.5 Attrition

Out of the 3,144 respondents who passed the attention check, 146 have attrited, which accounts for approximately 4.6% of the sample. Although this is a relatively small proportion, I test whether

my treatment causes respondents to attrite. As reported in Figure B13, I do not find a statistically significant association between my shaming treatment and attrition.

Table B13: The effect of shaming treatment on attrition

	Attrition	
	(1)	(2)
Shaming treatment	−0.002	0.0001
	(0.007)	(0.007)
Demographic Controls	No	Yes
N	3,133	3,133

Notes: Demographic controls include: sex, age, race, ethnicity, education, ideology, and region. * $p < .1$, ** $p < .05$, *** $p < .01$

B.6 Information Leakage

Finally, I explore whether respondents who received the shaming treatment were more likely to think of a particular country. As reported in Table B14, the shaming treatment is weakly associated with leakage ($\beta = 0.03, p = 0.094$). This effect becomes weaker and statistically insignificant when including demographic controls. I thus assert that differential leakage does not appear to be a major issue in study II.

Table B14: The effect of shaming treatment on thinking of a specific country

	Did you think of a specific country?	
	(1)	(2)
Shaming treatment	0.029*	0.024
	(0.017)	(0.017)
Demographic Controls	No	Yes
N	3,133	3,133

Notes: Demographic controls include: sex, age, race, ethnicity, education, ideology, and region. * $p < .1$, ** $p < .05$, *** $p < .01$

B.7 Experiment II Survey Instrument

Italics text signifies coding text or text that was not present to respondents.

B.7.1 Pre-treatment attention check + hawkishness

An attention check was embedded within the hawkishness grid. Respondents who answered it incorrectly were removed from the survey.

- Below, you will see a series of statements. Please tell us whether you agree or disagree with each statement.
  - In the United States, our people are not perfect, but our culture is superior to others.
  - The use of military force only makes problems worse.
  - I would rather be a citizen of America than of any other country. Please click "disagree" regardless of your opinion.
  - Going to war is unfortunate, but sometimes the only solution.
  - The best way to ensure peace is through military strength.

B.7.2 Pre-treatment covariates

- **Partisanship:** Generally thinking, do you think of yourself as a: [Republican, Democrat, Independent, Another party (fill in), No preference]
- *If Republican:* Would you call yourself a: [Strong Republican, Not very strong Republican]
- *If Democrat:* Would you call yourself a: [Strong Democrat, Not very strong Democrat]
- *If neither Democrat nor Republican:* Do you think yourself as closer to the: [Republican Party, Democratic Party, Neither Party]

B.7.3 Experimental vignettes

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 administrations have handled the situation in different ways. We will describe one approach the U.S. government has taken and ask whether you approve or disapprove.

- A country violates human rights; it imprisons and tortures some of its citizens because of their beliefs and silences human rights defenders and activists.
- The country is [a U.S. ally. It has signed a military alliance with the U.S. and has high levels of trade with the U.S. / not a U.S. ally. It has not signed a military alliance with the U.S. and does not have high levels of trade with the U.S.]
- The U.S. government [called the country out and publicly criticized it for its violation of human rights in the United Nations./ did not criticize the country in the United Nations or make any public statement about the country's human rights abuses.]
- The country continued to violate human rights.

B.7.4 Manipulation checks

- In the passage you read, did the other country violate human rights? [yes, no, not specified]
- In the passage you read, was the other country an ally of the U.S.? [yes, no, not specified]
- In the passage you read, did the U.S. government criticize the other country? [yes, no, not specified]

B.7.5 Outcomes

- **Government approval:** How much do you approve or disapprove of the way the U.S. government handled the situation? [Strongly approve/Approve/Slightly approve/ Neutral/Slightly disapprove/Disapprove/Strongly disapprove]

- **Torture:** How much do you support or oppose government use of practices like torture that violate human rights **in the U.S.**? [Strongly support/Support/Slightly support/ Neutral / Slightly oppose / Oppose /Strongly oppose]

- **International legal obligation:** How strongly do you agree or disagree with the following statements: [Strongly disagree, Disagree, neither agree nor disagree, Agree, strongly agree]

  - It is important to me personally that the US will comply with international law.
  - Complying with international law is an important value.
  - Complying with international law is important, even if it contradicts the national interest.
  - I feel uncomfortable when the US violates international laws and norms.
  - If the US defies international laws and norms, criticism from other countries is justified.

- **HR perception:** How much do you believe that the U.S. government *in the passage* respects human rights? Please respond on a scale of 1 to 7, where 1 indicates the U.S. doesn't respect human rights at all, and 7 indicates it respects human rights a lot.

- **Morality:** In the passage you read, how moral and ethical do you believe the U.S. government to be? [Very moral/ moral/ slightly moral/ neutral/ slightly immoral/ immoral/ very immoral] *This was recoded such that higher variables indicate more moral)*

- **Power:** In the passage you read, how powerful or weak do you believe the U.S. government to be? [Very powerful/ powerful/ slightly powerful/ neutral/ slightly weak/ weak/ very weak] *This was recoded such that higher variables indicate more power)*

- **Placebo:** Did you think of a specific country when you read about the "other country" in the passage? If so, please specify. [Yes (open end), No]

B.7.6 Post-treatment covariates

Note that I collected these to ensure representativeness, but Lucid provides demographic meta data on all respondents, which was used as demographic controls in the models reported in Table A2.

- What is your age?

- Are you [Male/ Female /Other (insert answer)]

- What racial or ethnic group best describes you? [White, Black or African American, Hispanic or Latino, Asian or Asian American, Native American, Middle Eastern, Mixed race, Other (insert answer)]

- What is your highest level of education? [Less than high school / High school graduate/ Some college but no degree / Bachelor's / Master's / Doctoral / Professional (JD or MD)]

- What is your state of residence?

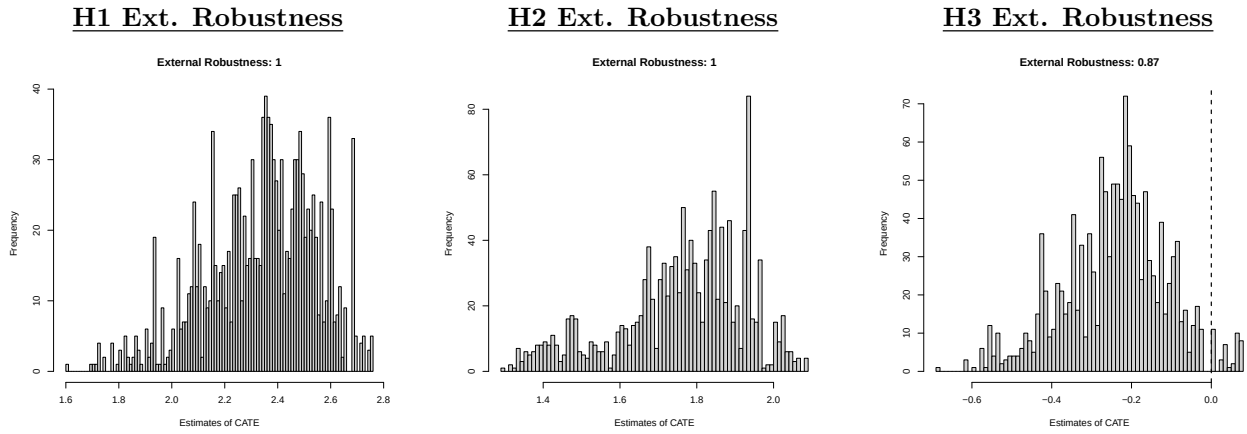

Figure C11: External Robustness and Distribution of Estimated CATEs when the outcomes are: government approval (left), favorable perceptions of respect for human rights (middle), and opposition of torture (right). The treatment is shaming, controlling for all pre-treatment covariates.

C Probing External Validity

C.1 Estimating External Robustness

In this section, I follow Devaux and Egami (2022) and test the estimated external robustness of my findings. In doing so, I estimate how different a population should be from my experimental sample to explain away the Target Population Average Treatment Effect (T-PATE). External robustness is estimated at 1 (the highest level) when the outcomes are perceptions of human rights and government approval. This implies that the target population average treatment effect (T-PATE) estimate is robust to populations that are highly different from the experimental sample and that my study exhibits low treatment effect heterogeneity. This is in line with my theoretical argument which suggests that individuals across the board are favorable of shaming violators.

The estimated external robustness when the outcome is opposition to torture is 0.87, as demonstrated by the distribution of estimated CATEs in Figure C11(right). The estimate is slightly lower than when the outcome accounts for perceptions/approval of the U.S. government but is still much higher than Devaux and Egami's proposed upper bound benchmark for moderate external robustness (0.57). This implies that the target population average treatment effect (T-PATE) estimate is robust to populations that are relatively different from the experimental sample. As a comparison, the causal estimates in my experiment is estimated at equal to zero only when the experimental sample is as different from a hypothetical population as MTurk samples are from the U.S. general population.

C.2 Awareness of shaming

A central assumption discussed in the manuscript relates to public awareness of human rights shaming. To assess the centrality of human rights shaming amongst the general public, I conducted a follow-up survey (see descriptive statistics in Table C15) in December 2023. Respondents were asked the following question: "Have you heard of any instance in the past month or so where U.S. officials publicly criticized another country's human rights policies?" A vast majority – 86.3% – of the sample said they heard of such instances in the past month.

Table C15: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Recall HRs shaming	102	0.863	0.346	0	1
Female	102	0.206	0.406	0	1
Male	102	0.794	0.406	0	1
Democrat	102	0.451	0.500	0	1
Independent	102	0.157	0.365	0	1
Republican	102	0.392	0.491	0	1
Age	102	32.971	8.117	23	71
Education	102	4.873	1.012	2	7

C.3 Observational Analysis (UPR)

Next, I examine whether shaming patterns in the UN Universal Periodic Review (UPR) are correlated with incumbent voting and favorable perceptions of human rights as reported by cross-national surveys. To estimate human rights shaming I use data from UPR info, an NGO that documents the activity of the UN Universal Periodic Review (UPR), covering three waves (2008-2011,2012-2016, 2017-2022). The UPR is a state-driven mechanism of the human rights regime, in which governments review each other's human rights records and can make targeted recommendations to specific countries, calling on them to improve specific human rights conditions and address violations. My main independent variable of interest is coded as the number of times a government shamed other countries in the UPR in a given year.

C.3.1 Outcomes

To test the correlation of UPR shaming with public attitudes, I used survey data from the CSES and WVS. When testing hypothesis 1 my outcome of interest is *incumbent vote*. The CSES surveyed 120,277 respondents from 51 countries between 2008-2020 and reported whether each respondent cast a ballot for the outgoing incumbent. In cases of a presidential election, the variable refers to the president and/or the president's party, in all other cases the variable refers to the parties which were part of the outgoing cabinet. This results in a binary variable, where 1 indicates a vote for the incumbent, and 0 indicates no vote for the incumbent (mean of $\mu = 0.41$ and a standard deviation of $\sigma = 0.49$).

When testing hypothesis 2 my outcome of interest is *perceptions of respect for human rights*. The WVS surveyed 154,761 respondents from 73 countries between 2008 and 2020 and reported participants' perceptions of domestic respect for human rights. Respondents were asked "How much respect is there for individual human rights nowadays in this country?" and answers ranged from 1 (a lot of respect) to 4 (no respect). This variable has been reverse coded such that higher numbers indicate more respect for human rights (mean of $\mu = 2.66$ and a standard deviation of $\sigma = 0.87$).

I merged each of these two outcomes with the UPR data, such that each observation is an individual's survey response. For every respondent, *UPR shaming* refers to the number of times the respondent's government has shamed other countries in the UPR in a given year. Tables C22 and C23 present the descriptive statistics for relevant variables.

C.3.2 Estimation Strategy

Many differences between governments may account for their ability to issue recommendations shaming other countries in the UPR *and* for their citizens' approval and human rights attitudes.

For instance, governments more respectful of human rights may also be more active in the UPR. As a result, citizens may be more likely to (accurately) say their country respects human rights. Additionally, countries with higher levels of state capacity, which are more respectful of human rights (Cole 2015), may be more capable of issuing condemning recommendations in the UPR, and simultaneously more favored by constituents. I take several steps to control for this omitted variable bias.

First, I employ country and year fixed effects in all of my models. In doing so, I isolate time-invariant country attributes and temporal attributes, considering how within-country changes in shaming affect incumbent voting and perceptions of respect for human rights. Second, although differences between countries are controlled for by the inclusion of the fixed effects, I increase the precision of my estimate by controlling for two central theoretical confounders: physical integrity rights (Fariss 2014) and state capacity (Hanson and Sigman 2021).

I also include several controls measured at the individual level to account for variation across survey respondents. I control for all the respondent demographics collected by the CSES or WVS, including age, gender, level of education, reported ideology, and income. When testing hypothesis 1, I also include a variable that accounts for the ideological distance between the respondent's reported ideology and the incumbent party's ideology, which is a predictor of incumbent voting (Singh and Tir 2019).

My preferred specification is the OLS model presented in equation 1, where I employ country (γ) and year (δ) fixed effects and cluster robust standard errors at the country-year level in order to estimate the association of *UPR shaming* conducted by country c at time $y-1$ (β) with my outcomes of interest, *incumbent voting* and *human rights perception* (y_{icy}) for respondent i from country c interviewed in year y. ζ signifies demographic controls, measured at the respondent level, η signifies physical integrity rights, and θ signifies state capacity, measured at the country level at time $y-1$.

$$y_{icy} = \beta UPR_{c,y-1} + \gamma_c + \delta_y + \zeta X_i + \eta Physical_{c,y-1} + \theta Capacity_{c,y-1} + \epsilon_{icy} \tag{1}$$

C.3.3 Main Observational Trends

Table C16 reports the results of models estimating the effect of UPR shaming on incumbent voting, using data from the CSES. I find that respondents whose government shamed more countries in a given year were more likely to vote for the incumbent in the following year. This modest effect is equivalent to 0.2% of a standard deviation and approaches statistical significance ($p = 0.055$) when including various individual-level demographic controls, and when controlling the respondent's country's physical integrity rights score and state capacity at time $t-1$.

Table C17 reports the results of models estimating the effect of UPR shaming on perceptions of respect for human rights, using data from the WVS. I find that respondents whose government shamed more countries in a given year were more likely to perceive their country as more respectful of human rights in the following year. This modest effect is equivalent to 0.12% of a standard deviation and is statistically significant ($p < 0.05$) when including various individual-level demographic controls, and when controlling for the respondent's country's physical integrity rights score and state capacity at time $t-1$.

C.3.4 Robustness Checks

I conduct several robustness checks to reduce concerns about selection bias. Indeed, it is possible that leaders who have higher government approval also have more political capital to shame in the first place. I include a placebo test in which I regress UPR shaming at time $t+1$ over each of the dependent variables at time t. The results of this test, presented in Table C18, suggest that UPR

Table C16: The effect of UPR Shaming on incumbent voting (CSES)

	Incumbent vote		
	(1)	(2)	(3)
UPR Shaming (t-1)	0.0004	0.001**	0.001*
	(0.0003)	(0.0003)	(0.0003)
Individual-level Controls	No	Yes	Yes
Physical Integrity (t-1)	No	No	Yes
State Capacity (t-1)	No	No	Yes
Year FEs	Yes	Yes	Yes
Country FEs	Yes	Yes	Yes
N	120,277	120,277	120,277

Notes: Robust standard errors are clustered at the country-year level. Missing data for control variables is imputed (average value by country-year). * $p < .1$, ** $p < .05$, *** $p < .01$

Table C17: The effect of UPR shaing on perceptions of human rights (WVS)

	Human Rights Perceptions		
	(1)	(2)	(3)
UPR Shaming (t-1)	0.001**	0.001**	0.001**
	(0.001)	(0.0005)	(0.001)
Individual-level Controls	No	Yes	Yes
Physical Integrity (t-1)	No	No	Yes
State Capacity (t-1)	No	No	Yes
Year FEs	Yes	Yes	Yes
Country FEs	Yes	Yes	Yes
N	154,761	154,761	154,761

Notes: Robust standard errors are clustered at the country-year level. Missing data for control variables is imputed (average value by country-year). * $p < .1$, ** $p < .05$, *** $p < .01$

shaming in the future is not associated with current incumbent vote ($p = 0.89$) or current perceptions of human rights ($p = 0.21$). If I were to find a significant correlation, it would suggest that UPR shaming is not independent, but rather confounded by other variables which may explain my main results. While this test does not fully rule out the issue of selection bias, it reduces some concerns.

I also demonstrate that my results remain robust to several alternative modeling. First, results remain substantively similar when I use a noisier estimate for my independent variable, using *all UPR recommendations*, rather than focusing only on harsher categories (Tables C20 and C21)). Second, since incumbent voting is a binary variable, I demonstrate that my results are robust to the inclusion of a generalized linear model (GLM) (Table C19). Finally, my results remain substantively similar (with slightly increased standard errors in some models), when alternative clustering at the country level.

Table C18: Placebo test: the effect of UPR shaming at t+1 on dependent variables at t

	Incumbent vote	Human rights perceptions
	(1)	(2)
UPR Shaming (t+1)	−0.00004	−0.001
	(0.0002)	(0.001)
Year FEs	Yes	Yes
Country FEs	Yes	Yes
N	119,394	134,460

Notes: Robust standard errors are clustered at the country-year level.

	Model 1	Model 2	Model 3
UPR Shaming (t-1)	0.0004	0.0006**	0.0006*
	(0.0003)	(0.0003)	(0.0003)
Individual-level Controls	No	Yes	Yes
Physical Integrity (t-1)	No	No	Yes
Imputed Capacity (t-1)	No	No	Yes
Year FEs	Yes	Yes	Yes
Country FEs	Yes	Yes	Yes
Num. obs.	120277	120277	120277
Num. groups: year	12	12	12
Num. groups: country.x	45	45	45
Deviance	27267.5466	25257.1062	25242.9471
Log Likelihood	−81415.1107	−76809.1223	−76775.3994
Pseudo R^2	0.0455	0.0994	0.0997

$^{***}p < 0.001$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table C19: The effect of UPR shaming on incumbent voting (CSES). Logistic Regression Models. Robust standard errors are clustered at the country-year level. Missing data for control variables is imputed (average value by country-year).

Table C20: Alternative Modeling (CSES)

	Incumbent vote					
	(1)	(2)	(3)	(4)	(5)	(6)
UPR Shaming (t-1)	0.001*	0.001*	0.001*			
	(0.0004)	(0.0004)	(0.0003)			
ALL UPR Recommendations (t-1)				0.0004	0.001**	0.001*
				(0.0003)	(0.0003)	(0.0003)
Individual-level Controls	No	Yes	Yes	No	Yes	Yes
Physical Integrity (t-1)	No	No	Yes	No	No	Yes
State Capacity (t-1)	No	No	Yes	No	No	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Country FEs	Yes	Yes	Yes	Yes	Yes	Yes
Clustering at country level	Yes	Yes	Yes	No	No	No
N	120,277	120,277	120,277	120,277	120,277	120,277

*Notes:*Missing data for control variables is imputed (average value by country-year). * $p < .1$, ** $p < .05$, *** $p < .01$

Table C21: Alternative modeling (WVS)

	Human Rights Perceptions					
	(1)	(2)	(3)	(4)	(5)	(6)
UPR Shaming (t-1)	0.001	0.001	0.001*			
	(0.001)	(0.001)	(0.001)			
All UPR Recommendations (t-1)				0.001**	0.001**	0.001**
				(0.0005)	(0.0005)	(0.0004)
Individual-level Controls	No	Yes	Yes	No	Yes	Yes
Physical Integrity (t-1)	No	No	Yes	No	No	Yes
State Capacity (t-1)	No	No	Yes	No	No	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Country FEs	Yes	Yes	Yes	Yes	Yes	Yes
Clustering at country level	Yes	Yes	Yes	No	No	No
N	154,761	154,761	154,761	154,761	154,761	154,761

*Notes:*Missing data for control variables is imputed (average value by country-year).

C.3.5 Descriptive Statistics

Table C22: Descriptive Statistics - CSES (H1)

Statistic	N	Mean	St. Dev.	Min	Max
Incumbent Vote	135,096	0.411	0.492	0	1
UPR Shaming	176,302	78.354	56.623	0	230
Physical integrity rights	188,950	1.748	1.589	−1.606	5.160
Ideological distance	198,930	2.500	1.599	0.000	9.000
Age	198,930	48.620	17.353	16.000	115.000
Gender (Female)	198,930	0.476	0.499	0	1
Education	198,930	2.547	1.131	0.000	4.000

Table C23: Descriptive Statistics - WVS (H2)

Statistic	N	Mean	St. Dev.	Min	Max
Human Rights Perceptions	163,328	2.659	0.867	1	4
UPR Shaming	160,260	44.545	40.575	0	241
Physical integrity rights	163,311	0.267	1.456	−1.816	4.541
Ideology	173,235	5.673	2.063	1.000	10.000
Age	173,235	42.548	16.486	16.000	103.000
Gender (Female)	173,235	0.474	0.499	0	1
Education	173,235	4.955	1.623	1.000	8.000
Income	173,235	3.305	0.970	1.000	5.000

References

Acharya, A., Blackwell, M., and Sen, M. (2018). Analyzing causal mechanisms in survey experiments. *Political Analysis*, 26(4):357–378.

Cole, W. M. (2015). Mind the gap: State capacity and the implementation of human rights treaties. *International Organization*, 69(2):405–441.

Coppock, A. (2021). Visualize as you randomize. *Advances in Experimental Political Science*, page 320.

Fariss, C. J. (2014). Respect for human rights has improved over time: Modeling the changing standard of accountability. *American Political Science Review*, 108(2):297–318.

Hanson, J. K. and Sigman, R. (2021). Leviathan's latent dimensions: Measuring state capacity for comparative political research. *The Journal of Politics*, 83(4):1495–1510.

Singh, S. P. and Tir, J. (2019). The effects of militarized interstate disputes on incumbent voting across genders. *Political behavior*, 41(4):975–999.