

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

Lotem Bassan-Nygate[†]

September 30, 2022

Word Count: 10,000

Abstract

Does shaming human rights violators shape attitudes at home? A growing literature studies the effect of shaming on public attitudes in the target state, but far less is known about its effect in countries initiating the criticism, i.e. the *shamers*. In this article, I theorize that shaming human rights violators shapes both domestic government approval and human rights attitudes. To test my theory, I implement two survey experiments in the U.S. and conduct cross-national observational analysis using data from the Comparative Study of Electoral Systems (CSES) and the World Values Survey (WVS), revealing three findings. First, citizens are more likely to approve of their government when it shames human rights violators. Second, citizens perceive their country as more respectful of human rights when their government shames other countries. At the same time, however, building on moral licensing theory, I show that shaming increases tolerance of domestic violations of human rights. My findings emphasize the mixed consequences of shaming, which are beneficial for governments but have troubling implications for the human rights project.

*Both survey experiments reported in this paper were pre-registered with OSF (DOIs: 10.17605/OSF.IO/PRXJY and 10.17605/OSF.IO/HKNJ2). Study I was posted on: <https://osf.io/prxjy>, and study II was posted on: <https://osf.io/hknj2>. The studies were approved by the University of Wisconsin-Madison IRB. I am grateful to Jessica Weeks for supporting this project and providing invaluable comments. I thank Zoltan Buzas, Adeline Lo, Nadav Shelef, Jonathan Renshon, and Chagai M. Weiss for helpful feedback.

[†]Harvard Kennedy School, University of Wisconsin - Madison, lbassan@wisc.edu, lotembassanygate.com.

Introduction

Naming and shaming is the process of singling out and publicly criticizing a specific government for a negative, or inappropriate behavior. For the most part, research on shaming has studied its effects on *targets of shame*, with a focus on human rights (Hafner-Burton 2008; Murdie and Davis 2012; Krain 2012; Hendrix and Wong 2013; Allendoerfer, Murdie and Welch 2020). Recent novel research studies shaming patterns in the UN (Terman and Voeten 2018; Terman and Byun 2020), and uses experiments to test the impact of shaming on public opinion, considering the important role domestic audiences play in pressuring governments to conform with international norms (Ausderan 2014; Bracic and Murdie 2020; Tingley and Tomz 2021; Greenhill and Reiter 2022). However, far less is known about the extent to which shaming affects public attitudes in the countries initiating the criticism, i.e. the *shamers*.

Drawing attention to shamers is important for several reasons. First, existing research provides only a partial picture by focusing on audiences in the target state, ignoring one of the most important attributes of shaming – that it is public in nature. In fact, shaming is a foreign policy behavior that is observed by, and likely intended for, a variety of different actors – leaders and the public in the shamed country, global human rights groups and activists, and the shamer’s domestic audience. As such, focusing on the effects of shaming on one actor – the target – de-emphasizes its global impact and masks the incentives to shame, which precede the target’s reactions.

Second, focusing on the shamers can provide behavioral foundations for the prevalence of

naming and shaming even when it is ineffective. While shaming has been found to decrease human rights abuses under certain conditions (Murdie and Davis 2012; Hendrix and Wong 2013; Allendoerfer, Murdie and Welch 2020), a growing body of work suggests that shaming, especially in the context of human rights, can create a backlash effect (Ilgit and Prakash 2019; Snyder 2020; Greenhill and Reiter 2022), leading to more violations (Hafner-Burton 2008) and resistance (Adler-Nissen 2014; Terman 2017). It is thus puzzling that governments continue to shame even when shaming backfires. One way to unpack this puzzle is by shifting our focus from targets of shame to its senders, focusing on the domestic political benefits of shaming that accrue even when it does not change state behavior.

Building on these insights, I develop a theory that puts the shamers front and center, explaining how and why naming and shaming affects domestic approval and attitudes towards human rights at home. I argue that by shaming other countries on international platforms, governments demonstrate their dedication to human rights and “virtue signal” to their public back home. In doing so, they appease domestic audiences by appearing more engaged in foreign policy, and improve perceptions of domestic respect for human rights. Governments are thus rewarded with higher approval ratings and more favorable perceptions of respect for human rights.

At the same time, shaming has negative externalities for support for human rights and international law more broadly. My theoretical framework builds on the moral licensing literature, which suggests that engaging in past good deeds liberates individuals to engage in unethical behavior (Merritt, Effron and Monin 2010; Kouchaki 2011; Cascio and Plant 2015; Blanken, van de Ven and Zeelenberg 2015; Lasarov and Hoffmann 2020). I argue that by shaming other countries governments are perceived as more moral by their citizens, build-

ing credentials that license problematic behaviors like violating human rights domestically. Taken together, my theory suggests that shaming is a very effective tool for governments, only perhaps not in the way scholars have traditionally thought about effectiveness. By shaming other countries, governments increase their approval and simultaneously obtain domestic leeway to violate human rights.

To test this theory, I combine three studies – two pre-registered survey-experiments in the U.S. and a cross-national observational analysis. In the first study, I fielded a survey experiment in which respondents were presented with a hypothetical scenario about a foreign country that violates human rights, manipulating whether the U.S. government criticized the country. I find that respondents who were told that the U.S. government shamed the human rights violator were more likely to approve of the U.S. government’s action. Treated respondents were also more likely to believe their own government respects human rights, but at the same time more supportive of violating human rights and international law. In a second experiment, I demonstrate that these effects hold even when the stakes of shaming are potentially high, namely when the targeted state is a U.S. ally.

After providing strong internally valid evidence from the U.S. in support of my theory, I turn to an observational cross-national analysis. I do so to complement my survey experiments, and examine the extent to which some of the effects identified in the U.S. context generalize to a broader population of interest. Analyzing data from the Comparative Study of Electoral Systems (CSES) (CSES 2020), and from the World Values Survey (WVS) (Inglehart et al. 2020), I demonstrate that individuals whose government shamed more countries in the UN Universal Periodic Review (UPR) were more likely to vote for the incumbent in the following year, and to perceive their country as more compliant with human rights.

These observable implications are consistent with findings from my survey experiments.

I make three central contributions to the existing literature. First, by theorizing about the effects of shaming on attitudes at home, I shift the spotlight from targets to the shamers and consider the global effect of shaming. By shedding light on domestic audiences we can begin to unpack the incentives of countries to engage in global criticism in the first place. Doing so may provide micro-foundations (Kertzer 2017) for puzzling phenomena, such as the pervasiveness of shaming, even when it fails to change targets' behaviors.

Second, I extend the literature on shaming by integrating works from public opinion (Guisinger and Saunders 2017; Tomz and Weeks 2020) with recent work from social and moral psychology (Merritt, Effron and Monin 2010; Kouchaki 2011; Lasarov and Hoffmann 2020) to theorize about individual-level motivations. In doing so, I explain how and why shaming other countries shapes domestic public opinion, and provide suggestive evidence for two central mechanisms – perceptions of morality and power – driving the effects I identify.

Third, by providing both experimental and observational evidence for my theory I marry two prominent methodological approaches in the literature. I join the growing literature that uses novel experimental designs to estimate the effects of shaming on individual attitudes (Bracic and Murdie 2020; Tingley and Tomz 2021; Greenhill and Reiter 2022), and complement it with cross-national data that allows me to consider global trends beyond the U.S. (Hafner-Burton 2008; Murdie and Davis 2012; Terman and Byun 2020). Consequently, I do not only consider the generalizability of my results, but also provide stronger causal evidence by using different types of data from different sources (Epstein and King 2002).

My findings have troubling implications for the human rights project. Not only are governments rewarded domestically without making any practical improvement in human

rights, citizens become more tolerant of domestic violations of human rights in light of shaming. This suggests that in the long run, governments might enjoy domestic support while receiving domestic leeway to violate international norms. I further discuss these implications in the concluding section of this article.

Shifting the Spotlight from Targets to Shamers

Governments are often publicly criticized for violating human rights. Past research has focused, for the most part, on the effects of shaming on *targets of shame*. This body of work dates back more than two decades, with the seminal work of [Keck and Sikkink \(1999\)](#) and [Risse et al. \(1999\)](#), who discuss shaming as an effective communicative tool between states that can raise moral consciousness and encourage citizens to take action. Over the years, a rich literature has sprung up to study empirically how shaming shapes domestic policies ([Hafner-Burton 2008](#); [Murdie and Davis 2012](#); [Krain 2012](#); [Hendrix and Wong 2013](#); [Allendoerfer, Murdie and Welch 2020](#)), and public opinion ([Ausderan 2014](#); [Bracic and Murdie 2020](#); [Tingley and Tomz 2021](#); [Greenhill and Reiter 2022](#)) in targeted states.

Shaming has been shown to decrease the prevalence of human rights abuses under certain conditions ([Murdie and Davis 2012](#); [Hendrix and Wong 2013](#); [Allendoerfer, Murdie and Welch 2020](#)). However, a growing body of work highlights its inadequacy, especially in the context of human rights. This research suggests that shaming is a counter-productive strategy ([Snyder 2020](#)), often leading to more violations ([Hafner-Burton 2008](#)), resistance ([Adler-Nissen 2014](#); [Terman 2017, 2020](#)), and a backlash in public attitudes towards human rights ([Greenhill and Reiter 2022](#)). Policy circles hold similar positions, widely accepting the ineffectiveness

of shaming in promoting rights (Neier 2018).

In light of this growing evidence, it is puzzling that governments engage in human rights shaming despite its inability to change domestic attitudes and policies in targeted states. In fact, shaming appears to be a prevalent tactic, used by a variety of states worldwide. This is evident in Figure 1, which demonstrates the global distribution of shamers in the UN Universal Periodic Review (UPR) between 2008-2020, where any government can shame any country after reviewing its human rights record. France is the most frequent shamer, with an average of 190 shaming instances per year, followed by Spain (172) and Canada (157). In fact, most countries appear to engage in human rights shaming to varying degrees.

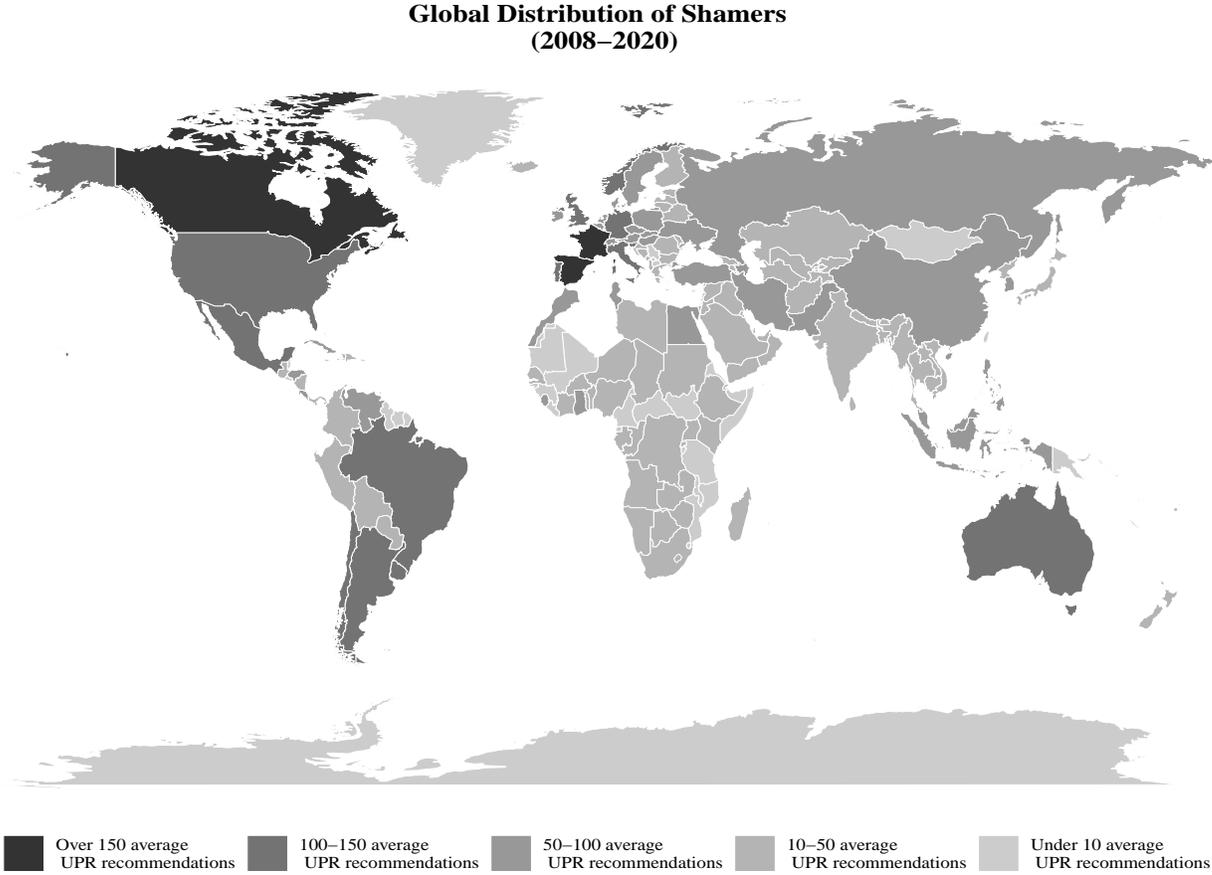


Figure 1: Average number of times each country has shamed other countries in a given year on the Universal Periodic Review (UPR) between 2008-2020.

We can begin to address the puzzling prevalence of shaming, despite its ineffectiveness, by diverting our attention from its targets to the countries that shame. Studying the *domestic* effects of shaming in the shamers' countries may shed light on the incentives that motivate governments to engage in shaming in the first place, independent of its ability to change state behavior or public attitudes in shamed societies. Learning that shaming is effective in domestic realms, rather than international ones, may provide behavioral foundations that explain its pervasiveness.

Indeed, while shaming is often understood to be a tool used internationally, like many diplomatic efforts it is most likely a two-level game (Putnam 1988), intended for a variety of different audiences, including domestic ones. Shaming is primarily a *public* process. Condemnations, whether carried out on international platforms like the UN (Lazaroff 2021), conducted during bilateral meetings (McCurry 2021), or stated in official government reports (State Department 2020), receive media coverage and public attention.

Hence, various actors including publics in the target country, global human rights activists, and the general public back home, are exposed to shaming and may perceive it and react to it in different ways. Thus, limiting our research agenda to target countries de-emphasizes the global impact of shaming, which may have reach well beyond shamed societies, ultimately shaping public attitudes towards human rights internationally. Such impact may contribute to the overall (in)effectiveness of the human rights regime.

While little attention has been allocated to shamers in the literature, some notable exceptions include recent works by Terman and Voeten (2018) and Terman and Byun (2020) who focus on shamers by analyzing shaming patterns in the UPR. These works provide rich theoretical and empirical insights of politicization of the human rights regime, demonstrating

that states shame selectively. They are thus primarily focused with the causes of shaming rather than their domestic consequences. In what follows, I argue that shaming entails certain domestic rewards for shamers. In the long run, these benefits may make international criticism a particularly appealing foreign policy behavior for governments.

Shaming and Government Approval

What are the rewards associated with shaming? [Terman and Byun \(2020\)](#) theorize that countries shame selectively because they face a dilemma – on the one hand, they wish to maintain relationships with their allies, on the other hand they want to collect the social rewards associated with shaming. These rewards, they argue, include appearing to be defenders of human rights and “good citizens” of the international regime.

Indeed, in shaming human rights violators, governments likely appease third parties like global activists, NGOs, and other countries who are dedicated to the human rights project. However, shaming other countries may also include domestic *material* rewards, such as government approval and increased vote share. I argue that by shaming other countries governments can collect domestic rewards because the general public evaluates shaming in positive ways for several reasons.

First, domestic actors within the shamer’s country who advocate human rights issues are particularly likely to support governments that vocalize human rights concerns. Indeed, leaders are often criticized for sidelining issues related to human rights and failing to hold other countries accountable for violations. Consider, for instance, recent criticism against U.S. President Joe Biden’s administration for its reluctance to pressure U.S. allies such as

Egypt, Turkey, and Saudi Arabia in the face of repression (Wootson Jr. 2021), or criticism of European governments for not taking a harder approach towards Israel (Haddad 2021). Domestic human rights activists and other concerned citizens often expect their government to hold human rights violators accountable. When their government publicly condemns violators, they view their leadership as embracing human rights issues.

Naturally, one may expect shaming to play a more significant role in appeasing leftists and doves, who are often considered more supportive of human rights (McFarland and Mathews 2005; Rathbun et al. 2016). When considering, however, the politicized nature of human rights shaming (Krasner 1999), there are reasons to expect shaming to increase government approval across the board. Hawks, for example, may take particular pleasure in stigmatizing adversaries (Ash and Dolan 2021), viewing shaming as a form of global competition. Consider Republican criticism of U.S. President Joe Biden for arguably taking a soft approach towards China's human rights violations (DeMarche 2021). It is possible that these criticisms were motivated by a sense of a missed opportunity to confront a global rival and demonstrate resolve, rather than sincere concern for human rights in China. Thus, while hawks may not be driven by concerns about human rights issues per se, they are also likely to support condemnations, particularly when the target is an adversary.

Indeed, recent experimental work suggests that the public across the political map, at least in the U.S., is equally sensitive to human rights issues when forming foreign policy attitudes (Tomz and Weeks 2020). Hence, although hawks and doves may be ultimately driven by different concerns (Brutger and Kertzer 2018), I argue that they are both likely to evaluate administrations positively in light of shaming. While it is possible that hawks may be driven by concerns about *power*, or 'acting tough' on the international stage, while doves

may be driven by concerns regarding *morality*, there is no reason to expect either group to be particularly *unsupportive* of shaming human rights violators. Hence, I expect the general public to evaluate human rights shaming in positive terms, perceiving their government as more proactive and engaged in foreign policy. It follows that:

Hypothesis 1. *Shaming human rights violators increases domestic government approval.*

Shaming and Human Rights Perceptions and Attitudes

The benefits associated with shaming may surpass government approval. By shaming other countries, governments “virtue signal” their dedication to human rights, which may have different implications for public attitudes and perceptions. Shaming may serve as an elite cue, emphasizing the importance of human rights to domestic audiences. By criticizing other countries’ human rights abuses, leaders may signal to their own public that human rights issues are an important topic that one should draw attention to.

Indeed, research suggests that public opinion is often susceptible to elite cues, especially around foreign affairs, which are distant from most voters’ everyday concerns (Holsti 2004; Guisinger and Saunders 2017). Hence, although shaming may be cheap talk in the sense that shamers do not necessarily respect human rights back home, I argue that citizens are inclined to believe that governments who take foreign policy stances against violators are also more likely to respect human rights themselves:

Hypothesis 2. *Shaming human rights violators increases domestic perceptions that human rights are respected at home.*

At the same time, this elite cue may also shape domestic attitudes towards human rights, increasing domestic support for human rights and international law more broadly. Citizens who are driven by concerns about international image may become more supportive of human rights to avoid appearing hypocritical abroad. Recent experimental evidence suggests that individuals are averse to hypocritical behavior (von Grundherr, Jauernig and Uhl 2021), and care about their country’s international status (Powers and Renshon 2021) and reputation (Brutger and Kertzer 2018). Taken together, these works suggest the public may become more supportive of maintaining a good human rights record at home to avoid criticism for applying double standards.

By criticizing other countries, and especially global rivals, governments likely draw attention to their own human rights practices and invite counter-criticism. Take for instance Beijing’s criticism of U.S. double standards (Fearnow 2021); following criticism over Uyghur human rights violations, the Chinese communist party argued the U.S. has “mountains of human rights problems in its own country”, referencing racial inequality and the murder of George Floyd (Moritsugu 2021). Since such criticism is perceived as damaging of a country’s image (Simmons 2009), the public may become more supportive of norm-confirming behavior that abides to human rights and international law more generally. It follows that:

Hypothesis 3a. *Shaming human rights violators increases domestic support for human rights and international law.*

Alternatively, shaming human rights violators may *decrease* domestic support for human rights. The literature on “moral licensing” in psychology is a robust body of work that suggests that past moral behavior can increase immoral, unethical, or otherwise problematic

behaviors and attitudes (Merritt, Effron and Monin 2010). Individuals' past good deeds liberate them to engage in unethical behaviors without fearing to appear immoral. Individuals thus build "moral credentials" when engaging in moral behavior and draw on these past behaviors when they engage in immorality. Thus, to the extent that shaming human rights violators is perceived as moral, individuals may be *more likely* to support domestic violations of human rights.

Indeed, numerous studies demonstrate how moral behavior, such as supporting a Black political candidate or buying green products, can license problematic and immoral behavior such as engaging in prejudice (Effron, Cameron and Monin 2009), cheating, and stealing (Mazar and Zhong 2010). In fact, the moral licensing effect is so robust that even the prospect of morality, or simply planning to engage in moral behavior in the future, can license current immoral attitudes and behaviors (Cascio and Plant 2015).

One may wonder whether moral licensing, a theory that primarily focuses on individuals rather than groups, is applicable to international politics. After all, in the moral licensing theory, individuals engage in the moral behavior themselves, whereas in my application to IR, individuals observe their leaders' moral actions. However, recent findings from social psychology suggest that learning of one's in-group's past moral actions shapes one's moral attitudes and behaviors (Lasarov and Hoffmann 2020). For example, when learning of their group members' prior non-discriminatory behavior in selecting Hispanic applicants, respondents gave more discriminatory ratings to Hispanic applicants (Kouchaki 2011). This suggests that individuals, who derive a sense of identity from their social associations (Tajfel 2010), build moral credentials based on their in-group's moral behavior.

Applying this theoretical framework to naming and shaming, one might expect that crit-

icizing human rights violators serves as a moral credential, allowing respondents to support otherwise immoral policies, such as engaging in human rights violations, without fearing to appear unethical. Importantly, shaming can serve as a moral license if individuals believe that shaming other countries for their human rights violations constitutes a *moral* deed. Otherwise, it is unlikely that individuals established said moral credential. These perceptions of morality may shape, in turn, individual beliefs about levels of respect for human rights. It thus follows that:

Hypothesis 3b. *Shaming human rights violators decreases domestic support for human rights and international law.*

The Role of the Targeted Country

My theory suggests that shaming may be an effective tool for governments who can increase their approval and perceptions of domestic respect for human rights, while simultaneously obtaining domestic leeway to violate international norms. I assess a scope condition of this theory by considering the role of the targeted country in attenuating these effects. Although it is possible that shaming is effective regardless of the target's identity, I consider whether citizens support their government even when it shames a strategic and economic ally whose shaming has higher costs.

Indeed, shaming may be less appealing when the target is an ally of the shamer. Citizens may be more favorable towards criticizing an adversary with whom their country has no security alliances or economic ties. Such criticism does not impose any threats to existing relationships and may thus be perceived as costless. Alternatively, shaming allies may be

perceived as more genuine and ultimately more effective. In fact, recent work suggests that while criticizing allies is less common, such criticism is ultimately more likely to be accepted by targets (Terman and Voeten 2018).

Finally, it is possible that citizens hold both of these perceptions simultaneously and thus support shaming regardless of the identity of the shamed country. In the following sections, I test the hypotheses delineated in this section in three studies. In studies I and II, I provide causal evidence for my theory using survey experiments in the U.S. In study III, I address a limitation of survey experiments relating to external validity, and demonstrate cross-national evidence for two observable implications of my theory, using observational survey data from the CSES and WVS.

Studies I & II: Experimental Evidence

My theory suggests that publicly shaming other countries can shape government approval as well as human rights attitudes and perceptions. I test this theory in two pre-registered survey experiments in the U.S. In study I, I manipulate information regarding the U.S. government's criticism of a country that violates human rights on the UN platform. In study II, I include an additional treatment arm randomizing information regarding the relationship of the shamed country with the U.S. Doing so allows me to examine whether my theory holds when the stakes of shaming are higher. In the following section I present the research designs and discuss the results from both studies.

Research Design

Study I and II were fielded in the U.S. during March 2022 and June 2022, respectively, using Lucid’s survey platform. In study I, 1,200 attentive respondents were recruited, and in study II, 3,000 attentive respondents were recruited. Respondents were matched (through Lucid) to the general U.S. population based on sex, age, race, and region of residence. Descriptive statistics for both samples are reported in Tables [A1](#) and [B7](#).

The surveys were programmed online using Qualtrics. In both experiments, respondents were exposed to the experimental vignette following a screening attention check and a partisanship question (see sections [A.7](#) and [B.6](#)). All respondents were then presented with a briefing text:

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.

In study I, respondents were randomized into a treatment and control condition, depicted in [Figure 2](#). All respondents were told about a hypothetical scenario in which a foreign country engages in human rights violations. Information about whether the U.S. government publicly condemned the country in the U.N. for its behavior was manipulated with probability 0.5. I follow [Kertzer and Brutger \(2016\)](#) and control for policy outcomes by fixing the final result of the scenario across both conditions: the other country continued to engage in violations. Doing so allows me to ensure treated respondents do not assume

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.

Figure 2: **Study I Experimental vignette.** Respondents were randomly assigned into one of the two conditions. Red text signifies treatment condition and blue text signifies control condition.

shaming leads to less violations in the target state, and creates a harder test for my theory – testing the domestic benefits of shaming even when it does not change the target's behavior.

Study II included a second treatment arm, which provided additional information about the shamed target, as depicted in Figure 3. Respondents in the *ally* condition were told that the target country has security and economic ties with the U.S., while respondents in the *non-ally* condition were told that the target country has neither security nor economic ties with the U.S. This results in a fully crossed 2X2 design, where respondents were randomized into one of the four conditions with probability 0.25.

After reading the vignette, respondents in both studies were presented with several questions that I employ as outcomes. To test my first hypothesis regarding government approval, participants were asked how much they approve or disapprove of the way the U.S. government handled the situation in the scenario. Possible responses ranged from 'strongly approve' to 'strongly disapprove' on a 7 point scale, and have been recoded such that higher numbers indicate more support.¹

To test my second hypothesis regarding the effect of shaming on perceptions of domestic

¹For ease of interpretation, all outcomes presented in Figures 4 -6, have been rescaled from 0-100. In the appendix, I demonstrate that results remained the same when outcomes are modeled as a 7 point scale.

respect for human rights, I measure human rights perceptions. Respondents were asked how much they believe the U.S. government in the passage values and respects human rights, on a scale of 1 to 7, where higher numbers indicate more respect for human rights.

To test my third hypothesis regarding the effect of shaming on support for human rights and international law I employ two different measures. First, I follow previous work (Wallace 2013), and measure support for a specific human rights policy which was mentioned in the vignette – the use of torture. Respondents report how much they support or oppose the use of torture by that violate human rights in the U.S. Responses range from ‘strongly approve’ to ‘strongly disapprove’ on a 7 point scale, and have been recoded such that higher numbers indicate opposition to torture. Second, I borrow a more general measure of international legal obligation, developed by Bayram (2017), which captures general support for international law, where respondents were asked about the extent to which they believe the U.S. should respect international law.

I further include questions regarding two potential mechanisms – perceptions of U.S. morality and perceptions of U.S. power, which range from very moral/powerful to very immoral/weak, on a scale of 1 to 7, a manipulation check, and a question that gauges potential confounders by asking whether respondents thought of a specific country when reading the vignette.² The surveys concluded with a host of demographic questions.³

I estimated pre-registered OLS models identifying the effect of the shaming treatment on each outcome of interest (relating to: government approval, perceptions of respect for human rights, and support for human rights). To enhance the precision of my estimates, I employ

²See Sections A.7 and B.6 for the full texts of these questions, and Sections A.4, A.5, B.3 and B.5, for an analysis.

³Note, however, that Lucid provides meta-data on respondents, including demographics collected pre-treatment. I include these as demographic controls.

- 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.

Figure 3: **Study II Experimental vignette.** Respondents were randomly assigned into one of the four conditions. Teal text signifies the ally condition, violet signifies the non-ally condition, red text signifies shaming condition, and blue text signifies the no-shaming condition.

additional models including several demographic controls (sex, age, race, education, state of residence, and partisanship). In study II, I estimate the effect of the shaming treatment, controlling for the identity of the ally to enhance the estimates’ precision. I further estimate the interaction effect of the shaming and ally treatments, considering whether the effects are attenuated by perceptions of the human rights violator.

Study I: Shaming Human Rights Violators Shapes Attitudes at Home

In Figure 4 I report the main results of study I. I look at the average treatment effects (ATEs) of the shaming treatment on all four outcomes of interest, rescaled from 0-100. My results show a large statistically significant effect: learning that the U.S. government

shamed the country violating human rights increases government approval by approximately 40 percentage points, from 32 percentage points to 72 percentage points ($p < 0.0001$).

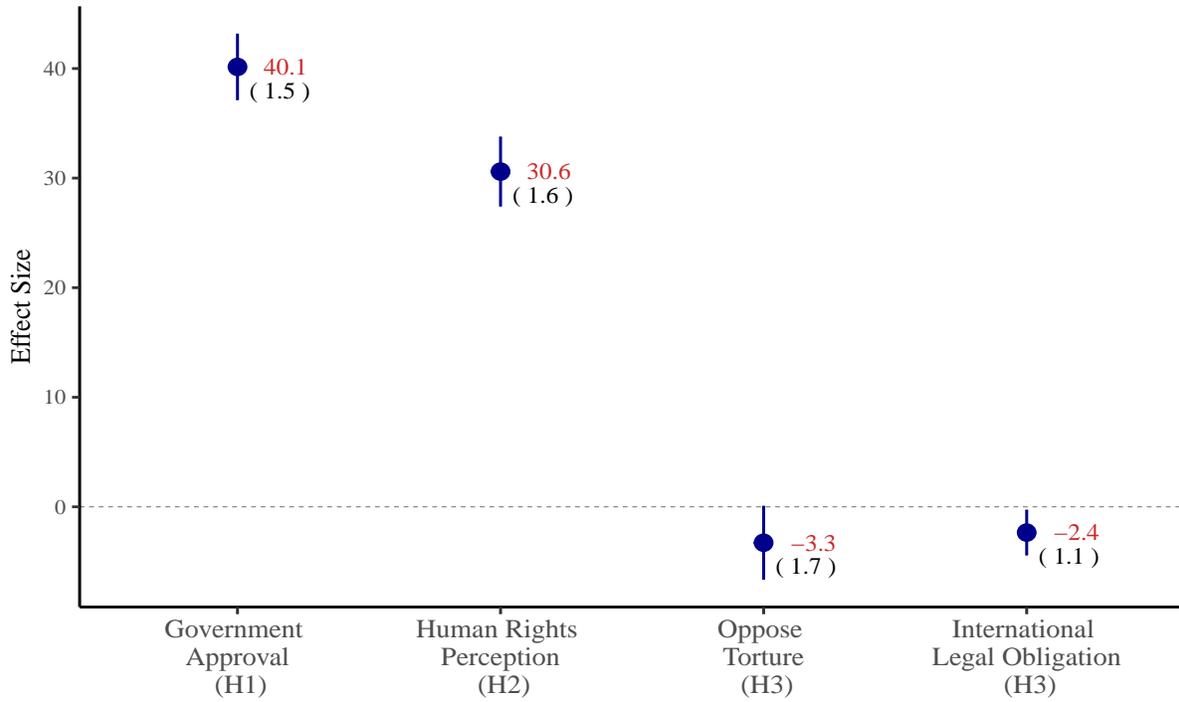


Figure 4: **Treatment effect on main outcomes of interest (Study I).** The X-axis represents four outcomes of interest – government approval, human rights perceptions, opposition of torture, international legal obligation, all of which are rescaled from 0-100. Regression estimates are marked in red, robust standard errors in parentheses.

Next, I consider the effectiveness of shaming human rights violators on human rights perceptions and attitudes. The treatment effect has a similarly sizable effect on human rights perceptions, increasing the belief that the U.S. government respects human rights by approximately 31 percentage points, from 43 percentage points to 74 percentage points ($p < 0.0001$). At the same time, respondents in the shaming treatment were less likely to oppose the use of torture, and reported feeling less obligation to international law and norms more generally. These effects are smaller in size – shaming decreased opposition to torture by approximately 3 percentage points ($p = 0.06$), and decreased international legal obligation

by approximately 2 percentage points ($p < 0.05$).

These results suggest that shaming human rights violators increases support for torture and decreases support for international law. This finding is consistent with H3b and the moral licensing literature, according to which shaming human rights violators serves as a moral credential that frees individuals to develop immoral attitudes and behavior. All the effects I identify and present in Figure 4 remain substantively similar when controlling for various pre-treatment demographic controls including sex, age, race, education, state of residence, and partisanship (see Table A3 in the Appendix).

Perceptions of Morality and Power as Mediators

Thus far, I have shown that shaming human rights violators increases government approval and perceptions of respect for human rights while decreasing support for human rights and international law. However, one may wonder why these patterns emerge. I explore two central mechanisms that may drive some of the effects identified in study I – perceptions of power and morality.

As explicated in the theoretical framework in Section , treated respondents may be more likely to approve of their government simply because they believe shaming human rights violators is the ethical thing to do, and thus find their government to be more moral. An additional mechanism may relate to the U.S. position vis-à-vis other countries, suggesting that shaming is favorable because it provides an opportunity to collide with adversaries and makes the U.S. appear more powerful worldwide.

To explore whether perceptions of morality and power are potential mechanisms, I em-

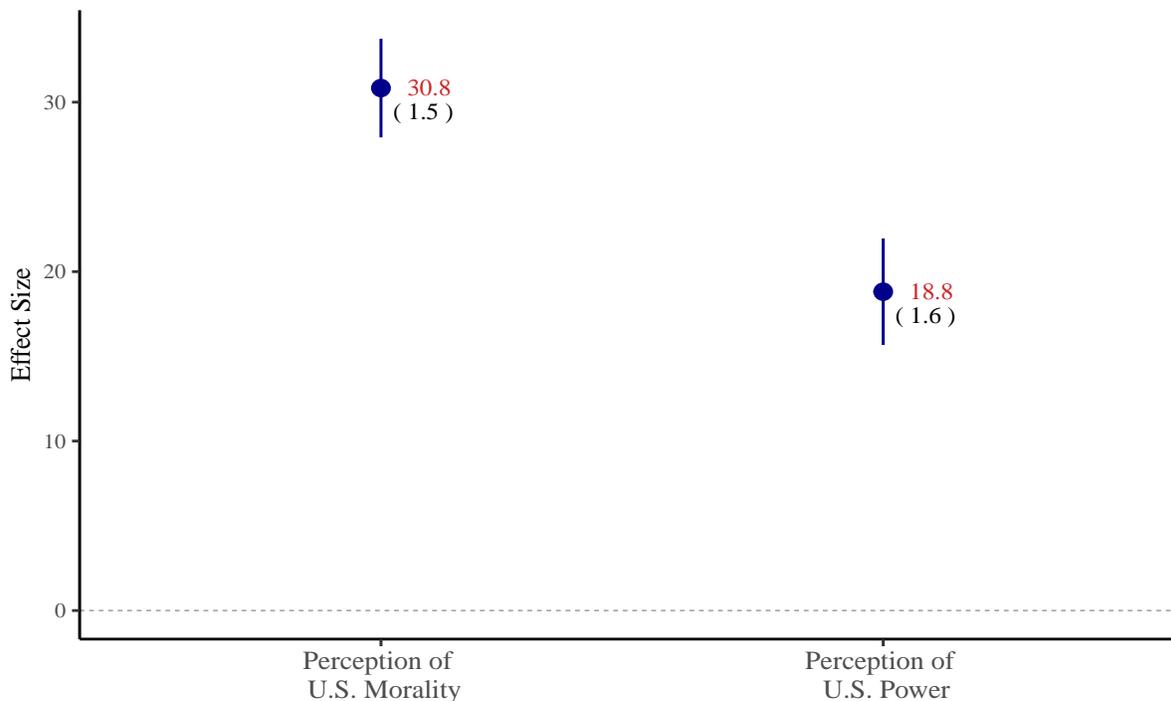


Figure 5: **Treatment effect on potential mechanisms (Study I)**. The X-axis represents two outcomes of interest – perceptions of U.S. morality and power, rescaled from 0-100. The Y-axis represents the effect size for each model. Regression estimates are marked in red, robust standard errors in parentheses.

ployed two post-treatment questions about U.S. morality and power (see Section A.7 of the appendix), and analyzed whether responses were shaped by my shaming treatment. As demonstrated in Figure 5, I find that treated respondents were more likely to view the U.S. government in the story as more moral and more powerful. Learning that the U.S. shamed the human rights violator increases perceptions of U.S. morality by approximately 31 percentage points, from 40 percentage points to 71 percentage points ($p < 0.0001$), and increases perceptions of U.S. power by approximately 19 percentage points, from 45 percentage points to 64 percentage points ($p < 0.0001$). In the Appendix, I conduct a mediation analysis utilizing Imai, Keele and Tingley’s 2010 mediation package. The results of the analyses are depicted in Figure A2. In both models the average causal mediation effect (ACME), the

direct effect and the total effect are positive and statistically significant at $p < 0.001$.

I further consider whether perceptions of morality drive the effect of my shaming treatment on opposition of international law. Since I hypothesize that shaming human rights violators operates as a moral credential, I expect the morality indicator to mediate the effect between shaming and decreased support for domestic compliance with human rights as well. In other words, I expect that shaming will license immoral behavior if it is perceived to be moral. I thus further conduct a mediation analysis where perception of morality is the mediator and opposition of torture is the outcome. The result, reported in Figure A1, suggests that the average causal mediation effect and the total effect are negative and statistically significant at $p < 0.05$.

Importantly, all three mediators described above are not randomly assigned and thus there is a potential concern for omitted variable bias that accounts for both the mechanism and outcome, violating the “sequential ignorability” assumption necessary for mediation analysis (Imai, Keele and Tingley 2010). I discuss these concerns in Section A.3, and conduct a sensitivity analysis, calculating the mediation effect for different magnitudes of a potential confounder’s effect on the mediator and the outcome. I also test whether study I faces concerns that relate to confounding due to information leakage (Dafoe, Zhang and Caughey 2018) (see Section A.5) or attrition (See Section A.6). I do not find evidence in favor of these concerns.

Study II: Shaming is Effective when Targets are both Allies and Non-Allies

Although there seems to be little concern relating to information leakage in study I, meaning that the treatment did not differentially cause respondents to think of a specific country (see Table A5), the findings from study I leave open questions regarding the scope of the theory. Figure A6 plots the names of countries that respondents revealed they thought may be the hypothetical human rights violator in the vignette. Respondents in both conditions mentioned countries that are not allies of the U.S., with adversaries like Russia and China being the most common responses.

While this finding does not pose a threat to internal validity, it raises questions regarding the scope of the theory. Namely, *will these findings replicate when the stakes of shaming are higher and the country is a U.S. ally?* I address this issue by varying the target of shaming in study II. As evident from Figure 6, the shaming treatment had substantively similar effects on respondents in both the ally and non-ally conditions, with no statistically significant differences between them (see heterogeneous treatment effect reported in Tables B8-B10).

Overall, the effects of the shaming treatment remained substantively similar to the ones identified in study I. The shaming treatment in study II increased government approval by approximately 24 percentage points, and perceptions of respect for human rights by approximately 17 percentage points ($p < 0.0001$). The shaming treatment decreased opposition to torture by approximately 3 percentage points ($p < 0.0001$), and international legal obligation by approximately 3 percentage points. All of these effects remain substantively similar when controlling for the target's identity treatment and various demographic controls including

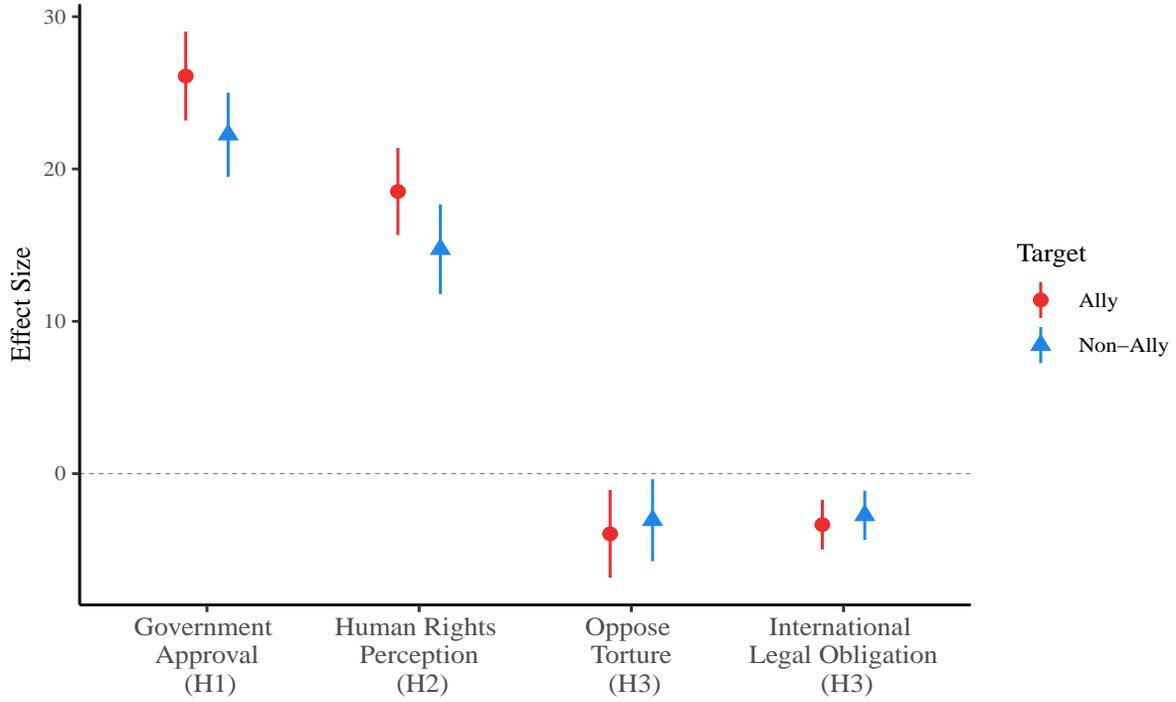


Figure 6: **The effect of the shaming treatment by target conditions (Study II).** The X-axis represents four outcomes of interest – government approval, human rights perceptions, opposition of torture, international legal obligation, rescaled from 0-100. The shaming treatment effect for respondents in the *ally* condition corresponds to the **red circle**. The shaming treatment effect for respondents in the *non-ally* condition corresponds to the **blue triangle**.

sex, age, race, education, state of residence, and partisanship. Taken together, these independent results from a separate sample, increase our confidence that shaming human rights violators can shape individual attitudes, even when the stakes of shaming are higher and the target is an ally.

Study III: Cross-national Evidence

While studies I and II provide causal evidence regarding the ability of shaming to shape government approval and human rights attitudes, they face certain limitations relating to

external validity. Indeed, a central critique of survey experiments relates to the extent to which the effects identified in one sample will generalize to a broader population of interest (Berinsky, Huber and Lenz 2012; Krupnikov and Levine 2014; Coppock 2019). To reduce some of these concerns, and to increase the confidence in my findings, I complement these studies with cross-national evidence analyzing global trends relating to two implications of my theory. To do so, I use data on government shaming in the UN Universal Periodic Review (UPR) combined with relevant outcomes collected in the Comparative Study of Electoral Systems (CSES) survey and in the World Values Survey (WVS).

UPR Shaming

To estimate the effect of human rights shaming I follow Terman and Byun (2020) and 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.

Although UPR recommendations have been used to measure government shaming patterns in the past, doing so poses certain challenges. First, while some recommendations include very harsh language, others are more encouraging or general in nature and may not be considered as criticism. Luckily, a new feature of UPR info classifies recommendations

into 5 categories of severity. My main analysis excludes weak recommendations that emphasize continuity which may be considered as praise. I thus focus on the three harshest categories coded by UPR info. In the appendix, I demonstrate that my results are robust to a more noisy estimate that includes *all UPR* recommendations.

Second, while government activity in the UPR is publicly available and occasionally publicized by human rights NGOs and media outlets (Amnesty 2022; Gavilan 2022), it is unlikely that most citizens follow it closely. Hence, I interpret UPR recommendations to be a proxy measuring the extent to which a country, in general, is likely to engage in shaming.

Outcomes

To test the effect of UPR shaming on public attitudes, I use 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 2008-2020,⁴ 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 2008-2020,⁵ and reported participants' perceptions of domestic respect for human rights. Respondents were

⁴See Table C20, for complete list of countries covered by the CSES in these years.

⁵See Table C21, for complete list of countries covered by the WVS in these years.

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, which 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 every observation is an individual’s survey response. For each 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 C18 and C19 present the descriptive statistics for relevant variables when datasets are merged.

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 who are 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. 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 physical integrity rights in each country in a given year (Fariss 2014), as this is a central theoretical confounder.

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, and η signifies physical integrity rights, measured at the country level at time $y - 1$.

$$y_{icy} = \beta UPR_{c,y-1} + \gamma_c + \delta_y + \zeta X_i + \eta Physical_{c,y-1} + \epsilon_{icy} \quad (1)$$

Results

Table 1 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 is statistically significant ($p < 0.05$) when including various individual-level demographic controls, and when controlling for

ideological distance from the incumbent and the respondent’s country’s physical integrity rights score at time $t - 1$.

Table 1: The effect of UPR Shaming on incumbent voting (CSES)

	Incumbent vote		
	(1)	(2)	(3)
UPR Shaming (t-1)	0.0004 (0.0003)	0.001 (0.0003)	0.001 (0.0003)
Individual-level Controls	No	Yes	Yes
Physical Integrity (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).

Table 2 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 respondents’ ideology and the respondent’s country’s physical integrity rights score at time $t - 1$.

In the appendix, 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

Table 2: The effect of UPR shaming on perceptions of human rights (WVS)

	Human Rights Perceptions		
	(1)	(2)	(3)
UPR Shaming (t-1)	0.001 (0.001)	0.001 (0.0005)	0.001 (0.0005)
Individual-level Controls	No	Yes	Yes
Physical Integrity (t-1)	No	No	Yes
Year FEs	Yes	Yes	Yes
Country FEs	Yes	Yes	Yes
<i>N</i>	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).

test, presented in Table C14, suggest that UPR 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 C16 and C17)). 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 C15). Finally, my results remain substantively similar (with slightly increased standard errors in some models), when alternative clustering at the country level (see section C.1).

To conclude, I find that governments' likelihood to engage in shaming, proxied by gov-

ernments' tendency to shame other countries in the UPR, is associated with an increase in incumbent voting and perceptions of domestic respect for human rights. The findings presented in Tables 1 and 2 reveal similar patterns to the experimental evidence presented in the previous section regarding hypotheses 1 and 2. Due to lack of available data, I am unable to test my third hypothesis regarding attitudes towards human rights and international law using cross-national data.

Conclusion

In this paper, I theorize about the domestic effects of shaming human rights violators. I find that criticizing other countries' human rights record: (1) increases government support; (2) increases perceptions of domestic respect for human rights; and (3) decreases support for human rights domestically and international norms and laws more generally.

I test my theory in three different studies. In studies I and II, I fielded survey experiments in the U.S., manipulating information about the U.S. government's criticism of a country that violates human rights. I found that information on shaming had a sizable effect on government approval and perceptions of U.S. respect for human rights. My shaming treatment also increased respondents' tolerance for the use of a method that violates human rights – torture, and decreased obligation to international law. These results remained substantively similar even when the stakes of shaming were potentially higher, with the target being a U.S. ally. Taken together, my findings suggest that shaming is an effective tool for governments – increasing approval while providing domestic leeway to violate human rights.

In a third study, I test some observable implications of my theory using cross-national

data from the CSES and WVS. I demonstrate that citizens whose governments shamed more countries in the UPR were more likely to vote for the incumbent and to perceive their government as more respectful of human rights in the following year. These results are robust to the inclusion of country and year fixed effects, and when controlling for human rights conditions and a host of individual-level demographics.

I contribute to the literature on several fronts. First, by studying the domestic effects of shaming I shift the focus from targets of shaming to the shamers. Doing so not only emphasizes the global impact of shaming, which exceeds shamed societies, it also provides behavioral foundations for the puzzling prevalence of naming and shaming, even when these strategies backfire (Hafner-Burton 2008; Adler-Nissen 2014; Terman 2017; Ilgit and Prakash 2019; Snyder 2020; Greenhill and Reiter 2022).

Second, my theory integrates works from moral psychology to explain the effects of shaming. Doing so both provides foundations for my IR theory and contributes to the work on moral licensing which has only recently begun to consider the role of groups in licensing individual behavior (Lasarov and Hoffmann 2020). Due to the nature of naming and shaming, the vignettes in my survey-experiment focus on elite behavior, providing a distinct case of moral licensing in which the actors performing the moral act are several levels removed from ordinary survey respondents. I thus demonstrate that moral licensing is robust even in such unique cases, when the in-group is one's national identity.

Third, by providing both experimental and observational evidence for my theory, I join the growing literature that uses novel experiments to test the effect of shaming (Ausderan 2014; Bracic and Murdie 2020; Tingley and Tomz 2021; Greenhill and Reiter 2022), while considering global trends that travel beyond the U.S. context. Doing so allows me to consider

the generalizability of my results while drawing attention to the important role that public opinion plays in international criticism.

Finally, my findings have several policy implications. I find that shaming is a beneficial tool for governments. Political elites who want to appease domestic audiences can engage in human rights shaming. However, what is good news for governments is bad news for the human rights project. By shaming other countries, leaders are able to shape domestic perceptions of respect for human rights without doing any of the ground work necessary; neither adopting new policies that improve human rights domestically, nor making official commitments to human rights. Shaming thus constitutes a cheap talk (at least domestically)⁶ that increases government approval without tying their hands by committing to human rights domestically. These implications are even more alarming when considering the negative externalities of shaming other countries towards international norms and legal obligation. My findings suggest that shaming decreases domestic support for human rights and international law, providing governments domestic leeway to violate international norms at home.

Despite these contributions, my paper faces some limitations that motivate future research. First, while I am able to provide cross-national evidence for hypotheses 1 and 2, regarding government approval and perceptions of domestic respect for human rights, I do not provide cross-national evidence for my third hypothesis regarding decreased support for human rights. This relationship is demonstrated in two U.S.-based survey experiments using two different samples, which increases our confidence in the finding. However, future research should strive to implement similar designs beyond the U.S. context.

⁶As highlighted by [Terman and Byun \(2020\)](#), shaming may bare international costs by souring existing relationships, especially if the target is an ally.

Second, although I provide a theoretical discussion of mechanisms that may drive my identified effects, as well as some suggestive empirical evidence, my empirical analysis falls short of providing direct evidence regarding causal mechanisms. As explicated in this paper, these mechanisms may relate to perceptions of morality and power. Future work should adapt rigorous designs that are capable of manipulating not only treatments, but also potential mechanisms (Imai et al. 2011), to advance our understanding of the effects of shaming.

References

- Adler-Nissen, Rebecca. 2014. “Stigma management in international relations: Transgressive identities, norms, and order in international society.” *International Organization* 68(1):143–176.
- Allendoerfer, Michelle Giacobbe, Amanda Murdie and Ryan M Welch. 2020. “The Path of the Boomerang: Human Rights Campaigns, Third-Party Pressure, and Human Rights.” *International Studies Quarterly* 64(1):111–119.
- Amnesty, International. 2022. ““Thailand: Mock fashion show protester sentenced to two years for ‘insulting the monarchy’”.”
- URL:** <https://www.amnesty.org/en/latest/news/2022/09/thailand-mock-fashion-show-protester-sentenced-to-two-years-for-insulting-the-monarchy/>
- Ash, Konstantin and Thomas M Dolan. 2021. “Politics of interconfessional empathy and Schadenfreude in Lebanon.” *Conflict Management and Peace Science* 38(6):718–741.
- Ausderan, Jacob. 2014. “How naming and shaming affects human rights perceptions in the shamed country.” *Journal of Peace Research* 51(1):81–95.
- Bayram, A Burcu. 2017. “Due deference: Cosmopolitan social identity and the psychology of legal obligation in international politics.” *International Organization* 71(S1):S137–S163.

- Berinsky, Adam J, Gregory A Huber and Gabriel S Lenz. 2012. "Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk." *Political analysis* 20(3):351–368.
- Blanken, Irene, Niels van de Ven and Marcel Zeelenberg. 2015. "A meta-analytic review of moral licensing." *Personality and Social Psychology Bulletin* 41(4):540–558.
- Bracic, Ana and Amanda Murdie. 2020. "Human rights abused? Terrorist labeling and individual reactions to call to action." *Political Research Quarterly* 73(4):878–892.
- Brutger, Ryan and Joshua D Kertzer. 2018. "A dispositional theory of reputation costs." *International Organization* 72(3):693–724.
- Cascio, Jessica and E Ashby Plant. 2015. "Prospective moral licensing: Does anticipating doing good later allow you to be bad now?" *Journal of Experimental Social Psychology* 56:110–116.
- Coppock, Alexander. 2019. "Generalizing from survey experiments conducted on Mechanical Turk: A replication approach." *Political Science Research and Methods* 7(3):613–628.
- Coppock, Alexander. 2021. "Visualize as You Randomize." *Advances in Experimental Political Science* p. 320.
- CSES. 2020. "CSES INTEGRATED MODULE DATASET (IMD)." *The Comparative Study of Electoral Systems (www.cses.org)*. .
- Dafoe, Allan, Baobao Zhang and Devin Caughey. 2018. "Information equivalence in survey experiments." *Political Analysis* 26(4):399–416.
- DeMarche, Edmund. 2021. "Biden takes heat for different norms remark while discussing China, human rights." .
URL: <https://www.foxnews.com/politics/biden-takes-heat-for-different-norms-remark-while-discussing-china-human-rights>
- Effron, Daniel A, Jessica S Cameron and Benoit Monin. 2009. "Endorsing Obama licenses favoring whites." *Journal of experimental social psychology* 45(3):590–593.

- Epstein, Lee and Gary King. 2002. "The rules of inference." *The University of Chicago Law Review* pp. 1–133.
- Fariss, Christopher J. 2014. "Respect for human rights has improved over time: Modeling the changing standard of accountability." *American Political Science Review* 108(2):297–318.
- Fearnow, Benjamin. 2021. "China Blasts US 'Human Rights Violations Double Standard': 'Robber Acting Like a Cop'." .
URL: <https://www.newsweek.com/china-blasts-us-human-rights-violations-double-standard-robber-acting-like-cop-1590144>
- Gavilan, Jodesz. 2022. "Why the upcoming UN review of human rights under Duterte matters" .
URL: <https://www.rappler.com/nation/things-to-know-united-nations-review-philippines-human-rights-record-rodrigo-duterte/>
- Greenhill, Brian and Dan Reiter. 2022. "Naming and shaming, government messaging, and backlash effects: Experimental evidence from the Convention Against Torture." *Journal of Human Rights* pp. 1–20.
- Guisinger, Alexandra and Elizabeth N Saunders. 2017. "Mapping the boundaries of elite cues: How elites shape mass opinion across international issues." *International Studies Quarterly* .
- Haddad, Benjamin. 2021. "How Europe Became Pro-Israel." .
URL: <https://foreignpolicy.com/2021/05/20/how-europe-became-pro-israel/>
- Hafner-Burton, Emilie M. 2008. "Sticks and stones: Naming and shaming the human rights enforcement problem." *International organization* 62(4):689–716.
- Hendrix, Cullen S and Wendy H Wong. 2013. "When is the pen truly mighty? Regime type and the efficacy of naming and shaming in curbing human rights abuses." *British Journal of Political Science* 43(3):651–672.
- Holsti, Ole R. 2004. *Public opinion and American foreign policy*. University of Michigan Press.
- Ilgit, Asli and Deepa Prakash. 2019. "Making human rights emotional: a research agenda to recover shame in "naming and shaming" ." *Political Psychology* 40(6):1297–1313.

- Imai, Kosuke, Luke Keele and Dustin Tingley. 2010. "A general approach to causal mediation analysis." *Psychological methods* 15(4):309.
- Imai, Kosuke, Luke Keele, Dustin Tingley and Teppei Yamamoto. 2011. "Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies." *American Political Science Review* 105(4):765–789.
- Inglehart, R., C. Haerper, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin and B. Puranen. 2020. "World Values Survey: All Rounds – Country-Pooled Datafile." *World Values Survey* .
- Keck, Margaret E and Kathryn Sikkink. 1999. "Transnational advocacy networks in international and regional politics." *International social science journal* 51(159):89–101.
- Kertzer, Joshua D. 2017. "Microfoundations in international relations." *Conflict Management and Peace Science* 34(1):81–97.
- Kertzer, Joshua D and Ryan Brutger. 2016. "Decomposing audience costs: Bringing the audience back into audience cost theory." *American Journal of Political Science* 60(1):234–249.
- Kouchaki, Maryam. 2011. "Vicarious moral licensing: the influence of others' past moral actions on moral behavior." *Journal of personality and social psychology* 101(4):702.
- Krain, Matthew. 2012. "J'accuse! Does naming and shaming perpetrators reduce the severity of genocides or politicides?" *International Studies Quarterly* 56(3):574–589.
- Krasner, Stephen D. 1999. *Sovereignty*. Princeton University Press.
- Krupnikov, Yanna and Adam Seth Levine. 2014. "Cross-sample comparisons and external validity." *Journal of Experimental Political Science* 1(1):59–80.
- Lasarov, Wassili and Stefan Hoffmann. 2020. "Social moral licensing." *Journal of Business Ethics* 165(1):45–66.

Lazaroff, Tovah. 2021. "157 nations call on Israel to halt exploitation of Palestinian resources."

URL: <https://www.jpost.com/arab-israeli-conflict/157-nations-call-on-israel-to-halt-exploitation-of-palestinian-resources-685487>

Mazar, Nina and Chen-Bo Zhong. 2010. "Do green products make us better people?" *Psychological science* 21(4):494–498.

McCurry, Justin. 2021. "US and China publicly rebuke each other in first major talks of Biden era."

URL: <https://www.theguardian.com/world/2021/mar/19/us-china-talks-alaska-biden-blinken-sullivan-wang>

McFarland, Sam and Melissa Mathews. 2005. "Who cares about human rights?" *Political Psychology* 26(3):365–385.

Merritt, Anna C, Daniel A Effron and Benoît Monin. 2010. "Moral self-licensing: When being good frees us to be bad." *Social and personality psychology compass* 4(5):344–357.

Moritsugu, Ken. 2021. "China bashes US over racism, inequality, pandemic response."

URL: <https://abcnews.go.com/US/wireStory/china-bashes-us-racism-inequality-pandemic-response-76648636>

Murdie, Amanda M and David R Davis. 2012. "Shaming and blaming: Using events data to assess the impact of human rights INGOs." *International Studies Quarterly* 56(1):1–16.

Neier, Aryeh. 2018. "'Naming and shaming': still the human rights movement's best weapon."

URL: <https://www.openglobalrights.org/Naming-and-shaming-still-the-human-rights-movements-best-weapon/?lang=English>

Peyton, Kyle, Gregory A Huber and Alexander Coppock. 2020. "The generalizability of online experiments conducted during the COVID-19 pandemic." *Journal of Experimental Political Science* pp. 1–16.

Powers, Ryan and Jonathan Renshon. 2021. "International Status Concerns and Domestic Support for Political Leaders." *American Journal of Political Science* .

- Putnam, Robert D. 1988. "Diplomacy and domestic politics: the logic of two-level games." *International organization* 42(3):427–460.
- Rathbun, Brian C, Joshua D Kertzer, Jason Reifler, Paul Goren and Thomas J Scotto. 2016. "Taking foreign policy personally: Personal values and foreign policy attitudes." *International Studies Quarterly* 60(1):124–137.
- Risse, Thomas, Thomas Risse-Kappen, Stephen C Ropp and Kathryn Sikkink. 1999. *The power of human rights: International norms and domestic change*. Vol. 66 Cambridge University Press.
- Simmons, Beth A. 2009. *Mobilizing for human rights: international law in domestic politics*. Cambridge University Press.
- Singh, Shane P and Jaroslav Tir. 2019. "The Effects of Militarized Interstate Disputes on Incumbent Voting Across Genders." *Political behavior* 41(4):975–999.
- Snyder, Jack. 2020. "Backlash against naming and shaming: The politics of status and emotion." *The British Journal of Politics and International Relations* 22(4):644–653.
- State Department, U.S. 2020. "2020 Country Reports on Human Rights Practices."
URL: <https://www.state.gov/reports/2020-country-reports-on-human-rights-practices/>
- Tajfel, Henri. 2010. *Social identity and intergroup relations*. Vol. 7 Cambridge University Press.
- Terman, Rochelle. 2017. "Rewarding resistance: Theorizing defiance to international norms." *Unpublished Manuscript* .
- Terman, Rochelle. 2020. "The positive side of negative identity: Stigma and deviance in backlash movements." *The British Journal of Politics and International Relations* 22(4):619–630.
- Terman, Rochelle and Erik Voeten. 2018. "The relational politics of shame: Evidence from the universal periodic review." *The Review of International Organizations* 13(1):1–23.

- Terman, Rochelle and Joshua Byun. 2020. "Punishment and Politicization in the International Human Rights Regime." *American Political Science Review* pp. 1–18.
- Tingley, Dustin and Michael Tomz. 2021. "The effects of naming and shaming on public support for compliance with international agreements: an experimental analysis of the Paris Agreement." *International Organization* pp. 1–24.
- Tomz, Michael R and Jessica LP Weeks. 2020. "Human rights and public support for war." *The Journal of politics* 82(1):182–194.
- von Grundherr, Michael, Johanna Jauernig and Matthias Uhl. 2021. "To condemn is not to punish: an experiment on hypocrisy." *Games* 12(2):38.
- Wallace, Geoffrey PR. 2013. "International law and public attitudes toward torture: An experimental study." *International Organization* 67(1):105–140.
- Wootson Jr., Cleve R. 2021. "Biden: U.S. leadership on human rights depends on our record at home."

Supplementary Information

How Does Shaming Human Rights Violators Shape Attitudes at Home?

Contents

A Study I	SI-1
A.1 Descriptive Statistics	SI-1
A.2 Mediation Analysis	SI-2
A.3 Sensitivity Analysis	SI-3
A.4 Manipulation Check	SI-5
A.5 Information Leakage	SI-6
A.6 Attrition	SI-6
A.7 Study I Survey Instrument	SI-6
A.7.1 Pre-treatment attention check + hawkishness	SI-6
A.7.2 Pre-treatment covariates	SI-7
A.7.3 Experimental vignettes	SI-8
A.7.4 Manipulation checks	SI-8
A.7.5 Outcomes	SI-8
A.7.6 Additional outcomes	SI-8
A.7.7 Post-treatment covariates	SI-9
B Study II	SI-9
B.1 Descriptive Statistics	SI-9
B.2 Pre-registered Models	SI-9
B.3 Manipulation Check	SI-9
B.4 Attrition	SI-12
B.5 Information Leakage	SI-12
B.6 Study II Survey Instrument	SI-12
B.6.1 Pre-treatment attention check + hawkishness	SI-13
B.6.2 Pre-treatment covariates	SI-13
B.6.3 Experimental vignettes	SI-13
B.6.4 Manipulation checks	SI-14
B.6.5 Outcomes	SI-14
B.6.6 Additional outcomes	SI-14
B.6.7 Post-treatment covariates	SI-15
C Study III	SI-15
C.1 Robustness Checks	SI-15
C.2 Descriptive Statistics	SI-18

A Study I

A.1 Descriptive Statistics

Table A1: Descriptive Statistics - Survey Experiment

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Government Approval	1,195	4.189	1.989	1.000	2.000	6.000	7.000
Torture	1,192	5.689	1.794	1.000	5.000	7.000	7.000
Legal Obligation	1,183	2.209	0.730	1.000	1.800	2.800	5.000
Human Rights Perceptions	1,175	4.580	1.882	1.000	3.000	6.000	7.000
Morality	1,179	4.425	1.767	1.000	3.000	6.000	7.000
Power	1,179	4.329	1.711	1.000	3.000	6.000	7.000
Gvmnt Approval (rescaled)	1,195	53.152	33.149	0.000	16.667	83.333	100.000
Torture (rescaled)	1,192	78.146	29.894	0.000	66.667	100.000	100.000
Legal Obligation (rescaled)	1,183	30.232	18.245	0.000	20.000	45.000	100.000
HR Perceptions (rescaled)	1,175	59.660	31.359	0.000	33.333	83.333	100.000
Morality (rescaled)	1,179	57.082	29.442	0.000	33.333	83.333	100.000
Power (rescaled)	1,179	55.485	28.524	0.000	33.333	83.333	100.000
Democrat	1,220	0.353	0.478	0.000	0.000	1.000	1.000
Republican	1,220	0.452	0.498	0.000	0.000	1.000	1.000
Female	1,173	0.533	0.499	0.000	0.000	1.000	1.000
White	1,173	0.736	0.441	0.000	0.000	1.000	1.000
Black/African American	1,173	0.108	0.311	0.000	0.000	0.000	1.000
Hispanic	1,173	0.066	0.249	0.000	0.000	0.000	1.000
Asian/Asian American	1,173	0.047	0.211	0.000	0.000	0.000	1.000
Native American	1,173	0.014	0.120	0.000	0.000	0.000	1.000
Middle Eastern	1,173	0.001	0.029	0.000	0.000	0.000	1.000
Mixed Race	1,173	0.019	0.136	0.000	0.000	0.000	1.000
Age	1,172	46.950	16.837	17.000	33.000	61.000	94.000

Table A2: Models with original scale

	Government Approval		HR Perception		Torture		Legal Obligation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	2.409 (0.092)	2.426 (0.096)	1.836 (0.096)	1.835 (0.100)	-0.197 (0.104)	-0.192 (0.103)	-0.094 (0.042)	-0.100 (0.043)
Demographics	No	Yes	No	Yes	No	Yes	No	Yes
N	1,195	1,195	1,175	1,175	1,192	1,192	1,183	1,183

Notes: Outcome variables with 7 point scales.

Table A3: Main Models with (pre-treatment) Demographic Controls

	Government Approval	HR Perception	Torture	Legal Obligation
	(1)	(2)	(3)	(4)
Treatment	40.436 (1.608)	30.585 (1.662)	-3.208 (1.724)	-2.497 (1.063)
Demographic Controls	Yes	Yes	Yes	Yes
<i>N</i>	1,195	1,175	1,192	1,183

Notes: Outcome variables rescaled from 0-100 to ease interpretation and match figures in main text.

A.2 Mediation Analysis

As described in the main text, I conduct a mediation analysis using [Imai, Keele and Tingley’s 2010 mediation](#) package. In [Figure A2](#), the outcome of interest is government approval, and the mediators are perceptions of U.S. morality (left) and power (right). The average causal mediation effect (ACME), or the indirect effect of morality/power, respectfully, 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.

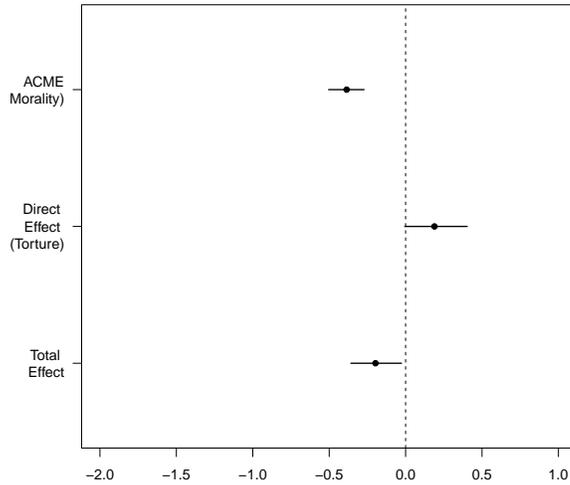


Figure A1: Causal mediation plot. Treatment is shaming manipulation, Mediator is morality (post-treatment), Outcome is opposition to torture. Horizontal lines represent 90% confidence intervals for estimates.

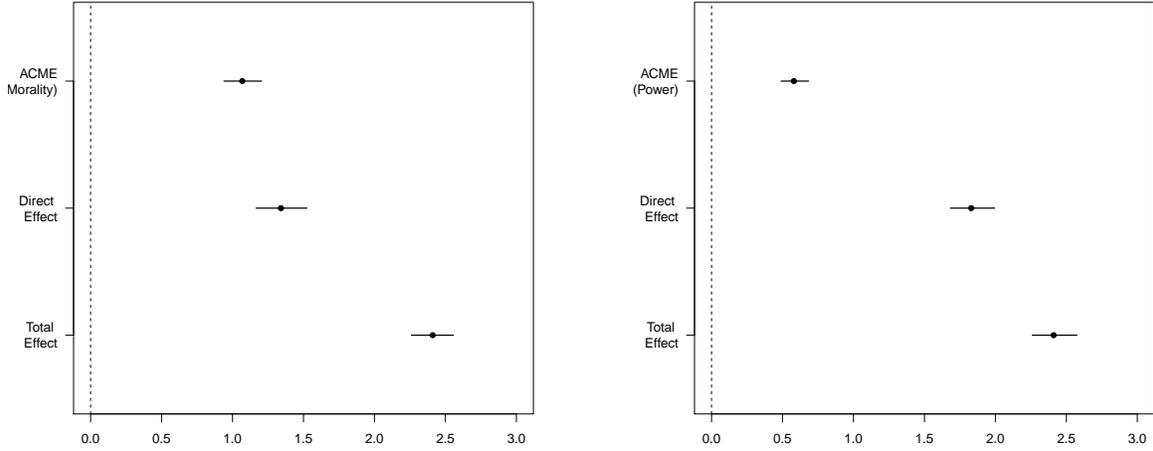


Figure A2: Causal mediation plots. In the figure on the left, morality perception is the mediator. In the figure on the right, morality perception is the mediator. In both figures, Treatment is shaming manipulation and Outcome is government approval. Horizontal lines represent 90% confidence intervals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

A.3 Sensitivity Analysis

Since the mediators in the mediation analysis conducted in Section A.2 are not randomly assigned, there is a concern for omitted variable bias that accounts for both the mediator and the outcome, 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 policy support and morality 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 60% of the variation in morality (M) and 50% of the variation in government approval (Y) for the estimate to be zero. At higher lower 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 A3(b), the mediation effects attached to each contour line which 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 60% of the variation in power (M) and 40% of the variation in government approval (Y) for the estimate to be zero. At higher lower 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

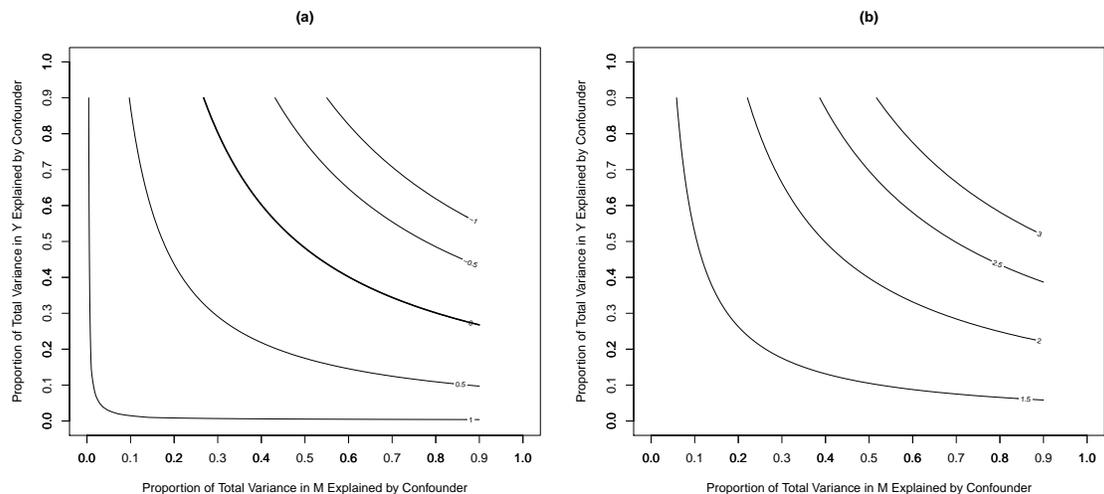


Figure A3: Sensitivity analysis for mediation (Mediator is morality, Outcome is government approval).

contour line which are all positive and grow stronger as more variance is explained.

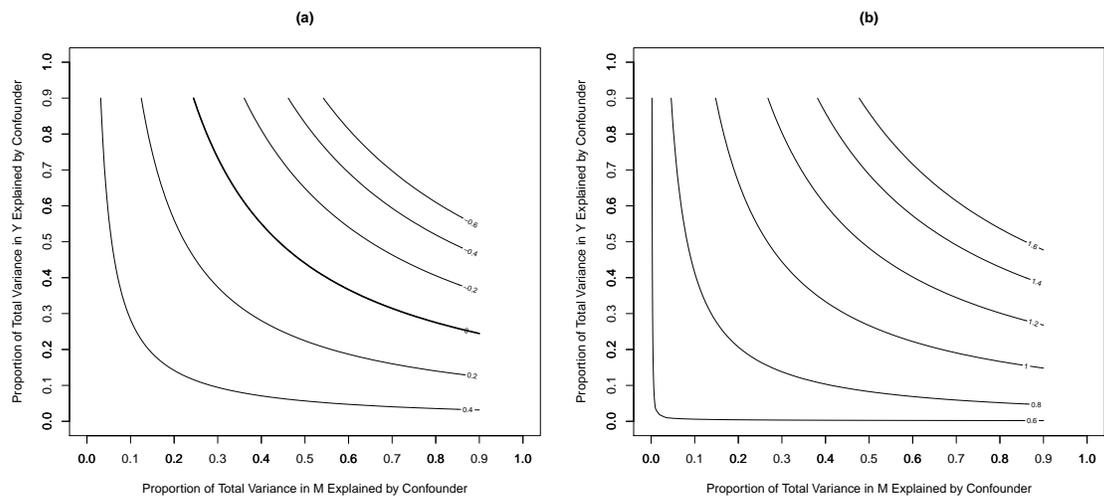


Figure A4: Sensitivity analysis for mediation (Mediator is power, Outcome is government approval).

Figure A5 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 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 which are all negative and grow stronger as more variance is explained.

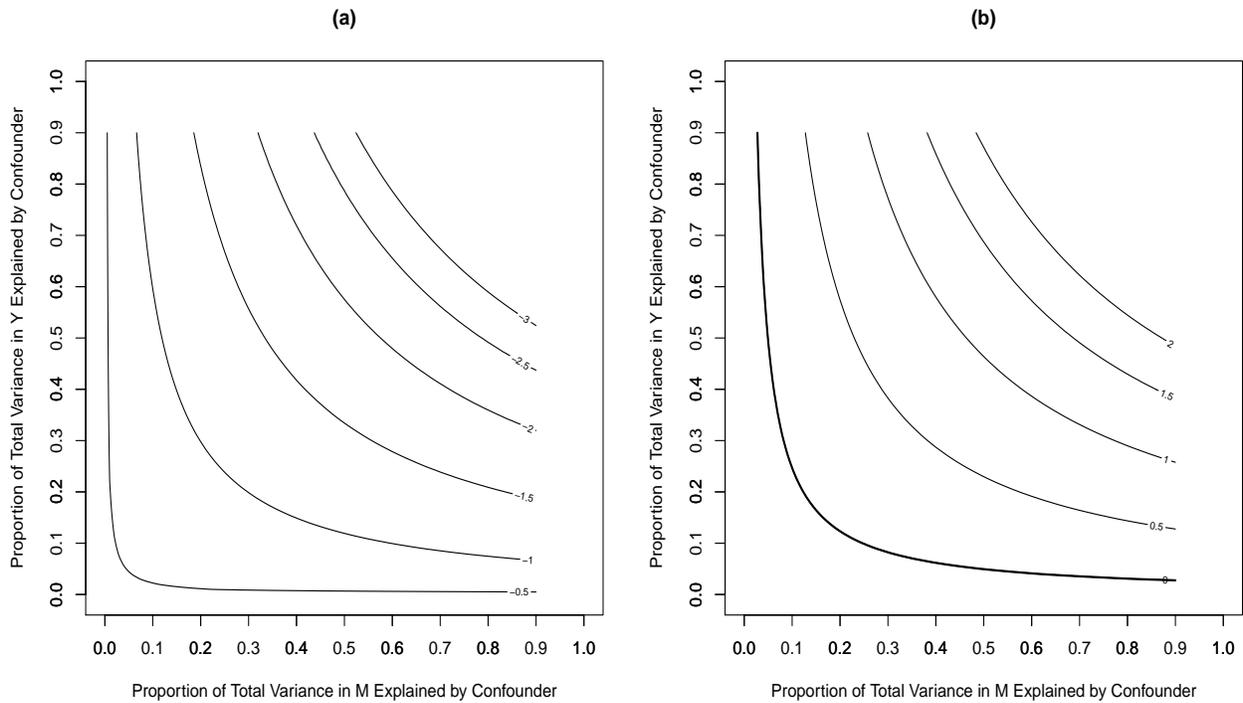


Figure A5: Sensitivity analysis for mediation (Mediator is human rights perception, Outcome is oppose torture).

A.4 Manipulation Check

In this section I test whether I was able to successfully manipulate information regarding shaming the human rights violators. The shaming treatment increased the belief that the U.S. government shamed the other country by 56.7% points, more than doubling, from 40% points to 96.6% points ($p < 0.0001$). As demonstrated in Table A4, these results hold when including a host of demographic controls.

Table A4: Manipulation Check

	Belief that the U.S. government criticized the other country	
	(1)	(2)
Treatment	1.681 (0.028)	1.677 (0.030)
Demographic Controls	No	Yes
<i>N</i>	1,199	1,199

A.5 Information Leakage

To address a concern relating to information leakage (Dafoe, Zhang and Caughey 2018), respondents were asked whether they thought of a specific country when reading the vignette. In Table A5, 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 A4, reduce concerns about internal validity, as it is likely that the treatment successfully manipulated perceptions about shaming, rather than another aspect associated with it.

Table A5: The effect of shaming treatment on thinking of a specific country

	Did you think of a specific country?	
	(1)	(2)
Treatment	-0.003 (0.029)	-0.008 (0.030)
Demographic Controls	No	Yes
<i>N</i>	1,174	1,174

Another concern, however, is that while respondents did not think of a countries differentially, all respondents thought of one specific country. This concern relates to external validity, as it suggests that the results speak to one particular context. 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 demonstrate in Figure A6, that of the subsample who admitted to thinking of 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.6 Attrition

Out of the 1,220 respondents, 47 have attrited, which accounts for approximately 3.9% 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 A6, I do not find a statistically significant association between my shaming treatment and attrition.

A.7 Study I Survey Instrument

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

A.7.1 Pre-treatment attention check + hawkishness

An attention check was embedded within the hawkishness grid. Respondents who answer it incorrectly were removed from the survey. About 46% of respondents passed the attention check, which is in line with recent survey experiments in the U.S. (Peyton, Huber and Coppock 2020).

- 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.



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 A6: The effect of shaming treatment on attrition

	Attrition	
	(1)	(2)
Treatment	-0.010 (0.011)	-0.009 (0.012)
Demographic Controls	No	Yes
<i>N</i>	1,220	1,220

- 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.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]

A.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 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.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, did the U.S. government criticize the other country? [yes, no, not specified]

A.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 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.

A.7.6 Additional outcomes

- 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.7.7 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 A3.

- 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 Study II

B.1 Descriptive Statistics

B.2 Pre-registered Models

Tables B8-B10 present results of the pre-registered models. As discussed in the main text, these remain substantively similar to those reported in Study I.

B.3 Manipulation Check

In this section I test whether I was able to successfully manipulate the information randomized in the vignettes. Respondents who received the shaming treatment were more likely to recall that the country in the other country was shamed by the US, and respondents in the ally treatment were more likely to recall that the other country was a US ally. The shaming treatment had an effect size equivalent to 170% of a standard deviation ($p < 0.0001$) and the ally treatment had an effect size equivalent to 166% of a standard deviation ($p < 0.0001$). As demonstrated in Table B11, these results hold when including a host of demographic controls.

Table B7: Descriptive Statistics - Survey Experiment II

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Government Approval	2,992	4.299	1.845	1.000	3.000	6.000	7.000
Torture	2,978	5.600	1.923	1.000	4.000	7.000	8.000
Legal Obligation	2,978	2.347	0.737	1.000	2.000	2.800	5.600
Human Rights Perceptions	2,972	4.383	1.793	1.000	3.000	6.000	7.000
Gvmnt Approval (rescaled)	2,992	54.980	30.743	0.000	33.333	83.333	100.000
Torture (rescaled)	2,978	65.710	27.472	0.000	42.857	85.714	100.000
Legal Obligation (rescaled)	2,978	29.287	16.014	0.000	21.739	39.130	100.000
HR Perceptions (rescaled)	2,972	56.382	29.875	0.000	33.333	83.333	100.000
Democrat	3,021	0.358	0.479	0.000	0.000	1.000	1.000
Republican	3,021	0.459	0.498	0.000	0.000	1.000	1.000
Female	2,968	0.520	0.500	0.000	0.000	1.000	1.000
White	2,968	0.721	0.449	0.000	0.000	1.000	1.000
Black/African American	2,968	0.116	0.320	0.000	0.000	0.000	1.000
Hispanic	2,968	0.074	0.261	0.000	0.000	0.000	1.000
Asian/Asian American	2,968	0.047	0.212	0.000	0.000	0.000	1.000
Native American	2,968	0.009	0.095	0.000	0.000	0.000	1.000
Middle Eastern	2,968	0.001	0.037	0.000	0.000	0.000	1.000
Mixed Race	2,968	0.023	0.151	0.000	0.000	0.000	1.000
Age	2,968	45.898	16.973	18.000	31.000	61.000	99.000

Table B8: Treatment effects on government approval (H1)

	Government Approval			
	(1)	(2)	(3)	(4)
Shaming treatment	24.131 (1.034)	24.169 (1.026)	24.695 (1.019)	22.894 (1.435)
Ally treatment		-7.099 (1.026)	-7.473 (1.017)	-9.273 (1.434)
Shaming*Ally				3.601 (2.021)
Demographic Controls	No	No	Yes	Yes
<i>N</i>	2,992	2,992	2,992	2,992

Notes: Outcome variable is rescaled from 0-100 to ease interpretation and match figures in main text.

Table B9: Treatment effects on human rights perceptions (H2)

	Human Rights Perceptions			
	(1)	(2)	(3)	(4)
Shaming treatment	16.594 (1.053)	16.624 (1.047)	16.650 (1.046)	15.047 (1.473)
Ally treatment		-6.423 (1.047)	-6.624 (1.044)	-8.225 (1.471)
Shaming*Ally				3.206 (2.075)
Demographic Controls	No	No	Yes	Yes
<i>N</i>	2,972	2,972	2,972	2,972

Notes: Outcome variable is rescaled from 0-100 to ease interpretation and match figures in main text.

Table B10: Treatment effects on opposing torture and legal obligation (H3)

	Oppose Torture				International Legal Obligation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Shaming	-3.492 (1.005)	-3.504 (1.004)	-3.967 (0.950)	-3.836 (1.338)	-3.037 (0.584)	-3.043 (0.584)	-3.062 (0.570)	-2.825 (0.803)
Ally		2.873 (1.004)	3.118 (0.948)	3.249 (1.337)		1.517 (0.584)	1.556 (0.569)	1.792 (0.803)
Shaming*Ally				-0.262 (1.886)				-0.473 (1.132)
Demographics	No	No	Yes	Yes	No	No	Yes	Yes
<i>N</i>	2,978	2,978	2,978	2,978	2,978	2,978	2,978	2,978

Notes: Outcome variables rescaled from 0-100 to ease interpretation and match figures in main text.

Table B11: Manipulation Check

	US criticized the other country		Other country is a US ally	
	(1)	(2)	(3)	(4)
Shaming treatment	1.598 (0.020)	1.600 (0.020)		
Ally treatment			1.585 (0.020)	1.579 (0.020)
Demographic Controls	No	Yes	No	Yes
<i>N</i>	2,993	2,993	2,984	2,984

B.4 Attrition

Out of the 3,025 respondents who passed the attention check, 57 have attrited, which accounts for approximately 1.8% 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 B12, I do not find a statistically significant association between my shaming treatment and attrition.

Table B12: The effect of shaming treatment on attrition

	Attrition	
	(1)	(2)
Shaming treatment	-0.001 (0.005)	-0.0002 (0.005)
Demographic Controls	No	Yes
<i>N</i>	3,020	3,020

B.5 Information Leakage

Finally, I explore whether respondents who received the shaming treatment were more likely to think of a particular country. This would suggest that certain information may have leaked and confounded my findings. As reported in Table B13, there is no statistically significant association between my shaming treatment and attrition.

B.6 Study II Survey Instrument

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

Table B13: The effect of shaming treatment on thinking of a specific country

	Did you think of a specific country?	
	(1)	(2)
Shaming treatment	-0.006 (0.018)	-0.003 (0.018)
Demographic Controls	No	Yes
<i>N</i>	2,970	2,970

B.6.1 Pre-treatment attention check + hawkishness

An attention check was embedded within the hawkishness grid. Respondents who answer it incorrectly were removed from the survey. About 63% of respondents passed the attention check, which is higher than reported in recent survey experiments in the U.S. (Peyton, Huber and Coppock 2020).

- 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.6.2 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.6.3 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.6.4 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.6.5 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.

B.6.6 Additional outcomes

- **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.6.7 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 A3.

- 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?

C Study III

C.1 Robustness Checks

Table C14: 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 (0.0002)	-0.001 (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 (0.0003)	0.0006 (0.0003)	0.0007 (0.0003)
Individual-level Controls	No	Yes	Yes
Physical Integrity (t-1)	No	No	Yes
Year FEs	Yes	Yes	Yes
Country FEs	Yes	Yes	Yes
pseudo.r.squared	0.0461	0.1001	0.1001
nobs	120277	120277	120277
AIC	162944.2214	153740.2447	153739.2705
BIC	163496.9819	154331.7954	154340.5187
Log Likelihood	-81415.1107	-76809.1223	-76807.6352

Table C15: 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 C16: Alternative Modeling (CSES)

	Incumbent vote					
	(1)	(2)	(3)	(4)	(5)	(6)
UPR Shaming (t-1)	0.001 (0.0004)	0.001 (0.0004)	0.001 (0.0004)			
ALL UPR Recommendations (t-1)				0.0004 (0.0003)	0.001 (0.0003)	0.001 (0.0003)
Individual-level Controls	No	Yes	Yes	No	Yes	Yes
Physical Integrity (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
<i>N</i>	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).

Table C17: 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 (0.0005)	0.001 (0.0005)	0.001 (0.0005)
Individual-level Controls	No	Yes	Yes	No	Yes	Yes
Physical Integrity (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
<i>N</i>	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.2 Descriptive Statistics

Table C18: Descriptive Statistics - CSES (H1)

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Incumbent Vote	135,096	0.411	0.492	0.000	0.000	1.000	1.000
UPR Shaming	176,302	78.354	56.623	0.000	35.000	114.000	230.000
Physical integrity rights	188,950	1.748	1.589	-1.606	0.442	2.954	5.160
Ideological distance	198,930	2.500	1.599	0	1.1	3	9
Age	198,930	48.620	17.353	16	34	62	115
Gender (Female)	198,930	0.524	0.499	0	0	1	1
Education	198,930	2.547	1.131	0	2	3.5	4

Table C19: Descriptive Statistics - WVS (H2)

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Human Rights Perceptions	163,328	2.659	0.867	1.000	2.000	3.000	4.000
UPR Shaming	160,260	44.545	40.575	0.000	16.000	60.000	241.000
Physical integrity rights	163,311	0.267	1.456	-1.816	-1.058	1.096	4.541
Ideology	173,235	5.673	2.063	1	5	6.5	10
Age	173,235	42.548	16.486	16	29	55	103
Gender (Female)	173,235	0.526	0.499	0	0	1	1
Education	173,235	4.955	1.623	1	4	6	8
Income	173,235	3.305	0.970	1	3	4	5

Table C20: Countries and years covered by H1 (CSES) analysis.

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Argentina								X					
Australia	X					X						X	
Austria		X				X				X			
Belarus		X											
Belgium												X	
Brazil			X				X				X		
Bulgaria								X					
Canada	X			X				X				X	
Chile		X	X							X	X		
Costa Rica												X	
Croatia	X												
Czech Republic			X			X							
Denmark	X											X	
Estonia				X									
Finland				X				X				X	
France					X					X			
Germany		X				X				X			
Great Britain								X		X			
Greece		X				X		X	X				
Hungary											X		
Iceland		X				X			X	X	X		
Ireland				X					X				
Israel						X							X
Italy											X		
Japan						X					X		
Kenya						X							
Latvia			X	X			X						
Lithuania									X				
Mexico		X			X			X					
Montenegro						X			X	X			
Netherlands			X							X			
New Zealand	X	X		X	X		X	X		X	X		X
Norway		X	X			X	X			X			
Peru				X					X				
Philippines			X						X				
Poland				X									
Portugal		X						X				X	
Republic of Korea	X				X				X				
Romania		X			X	X	X						
Serbia					X	X							
Slovakia			X						X				X
Slovenia		X			X								
South Africa		X						X					
Spain	X												
Sweden							X				X		
Switzerland				X								X	X
Thailand				X								X	
Tunisia													X
Turkey				X				X		X			
United States of America	X				X	X			X	X			X
Uruguay			X										X

Table C21: Countries and years covered by H2 (WVS) analysis.

	2008	2009	2010	2011	2012	2013	2014	2016	2017	2018	2019	2020
Algeria							X					
Andorra										X		
Argentina						X			X			
Armenia				X								
Australia					X					X		
Azerbaijan				X								
Bangladesh										X		
Belarus				X								
Bolivia									X			
Brazil							X			X		
Canada												X
Chile					X					X		
China						X				X		
Colombia					X					X		
Cyprus				X							X	
Ecuador						X				X		
Egypt	X					X				X		
Estonia				X								
Ethiopia												X
Georgia		X					X					
Germany						X				X		
Ghana					X							
Greece									X			
Guatemala												X
Haiti								X				
Hungary		X										
India					X							
Indonesia										X		
Iran												X
Iraq						X				X		
Japan			X								X	
Jordan							X			X		
Kazakhstan				X						X		
Kuwait							X					
Kyrgyzstan				X								X
Lebanon						X				X		
Libya							X					
Malaysia					X					X		
Mexico					X					X		
Morocco				X								
Myanmar (Burma)												X
Netherlands					X							
New Zealand				X								X
Nicaragua												X
Nigeria					X					X		
North Macedonia												X
Pakistan					X					X		
Peru					X					X		
Philippines					X						X	
Poland					X							
Qatar			X									
Romania					X					X		
Russia				X					X			
Rwanda					X							
Singapore					X							X
Slovenia				X								
South Africa							X					
South Korea			X							X		
Spain				X								
Sweden				X								
Taiwan												
Tajikistan												X
Thailand							X			X		
Trinidad & Tobago			X									
Tunisia						X					X	
Turkey				X						X		
Ukraine				X								X
United States				X					X			
Uruguay				X								
Uzbekistan				X								
Vietnam												X
Yemen							X					
Zimbabwe					X							X