Fighting Facts or Fighting Norms: Contestation Over International Law Violations

Julia C. Morse*and Tyler Pratt‡

September 9, 2022

Abstract

Governments that violate international law frequently craft public narratives designed to reduce backlash and minimize demand for punishment. Under what conditions do these strategies avoid the reputational consequences of non-compliance? This paper examines political competition over the interpretation of international law violations. Drawing on image repair theory, we theorize that two primary factors affect an audience's interpretation of transgressions: the level of certainty about the allegation and the degree to which the behavior contravenes an established norm. When non-compliance occurs, intergovernmental contestation over facts and norms affects foreign audiences' preferences for the severity and speed of punitive action. We test our theory in a survey experiment examining US citizens' willingness to punish foreign governments for violations of trade and humanitarian law. The results suggest that governments can successfully reduce reputation costs arising from violations, but strategies differ in their ability to reduce certainty and outrage. Our findings provide insight into how contestation may affect international law over the long term. While disinformation might induce uncertainty, our research suggests it also increases moral outrage, which may lead to more consequential punishments and reinforcement of regimes.

1 Introduction

On 24 February 2022, Russia launched a military offensive into Ukraine. But while the explosions may have started on a single day, Russia's informational efforts began much earlier. When videos of Russian troop and artillery movements flooded social media in April 2021, a Kremlin spokesman called the situation on the Ukrainian-Russian border "extremely unstable" and warned of "a humanitarian catastrophe similar to Srebrenica."[1] The implication was clear: Russia's actions were designed to protect defenseless civilians. One year later, Russian President Vladimir Putin would commemorate the Allied victory in World War II with similar rhetoric, claiming that his army's "special military operation" was defending against "executioners, punishers, and Nazis."[2]

When governments violate international law, they have strong incentives to minimize the reputational fallout from bad behavior. Flagrant violations may result in severe punishment or retaliatory action by other states, and even small violations carry stigma that may impede future cooperation. The Russian government has long understood this process. When Putin invaded Crimea in 2014, he crafted a narrative of plausible deniability, bolstered by the use of soldiers in unmarked military uniforms. This strategy increased uncertainty over whether the Russian military or pro-Russian Crimean separatists were responsible for instigating military action. It was effective, at least initially, partly because other governments failed to get ahead of the story. In 2022, however, the Biden administration had learned from previous US government mistakes and began to counter the Russian narrative before it even took root. This informational contestation undermined Russian claims of humanitarian intervention and self-defense, increasing political support for a strong response.

---

[1] https://www.nytimes.com/2021/04/09/world/europe/russia-ukraine-war-troops-intervention.html?searchResultPosition=10

[2] "Victory Day unfolds quietly in Ukraine as Putin defends invasion." *The Washington Post.* 17 May 2022.

This paper examines how inter-governmental contestation affects the public's willingness to punish foreign governments for violating international law. We begin with the assumption that the reputational consequences and concomitant costs of non-compliance depend fundamentally on how audiences interpret the violation (Bull 1977, Kratochwil and Ruggie 1986). Allegations of non-compliance trigger a strategic competition to shape public perceptions about the government's responsibility and the severity of the transgression – a process we call reputation management.

Our theory of reputation management identifies two primary means through which violating governments can minimize backlash. They can persuade audiences to delay judgment and decisions about retaliatory action until outrage over the violation has dissipated. Alternatively, they can attempt to directly shape the level of outrage by justifying their behavior or undermining the norm that has been transgressed. Both methods reshape the strategic context in which political actors make decisions about punishment, reducing the net reputation cost of violations.

How can non-compliant governments persuade audiences to delay judgment or excuse their behavior? Scholarship on crisis communication and image management suggests that actors make judgments about controversial behavior on the basis of both information and underlying values (Benoit, 2015; Fishbein & Ajzen, 1977). Building on such insights, we theorize that states seek to decrease certainty about the alleged behavior and/or shift perceptions of the behavior's offensiveness. In other words, states choose to fight over facts or fight over values. The former practice, which we refer to as *information engagement*, contests the informational environment. The most aggressive form of information engagement is the "fake news" approach, where a government overtly denies the alleged transgression; however, governments also frequently employ partial denials to obscure details or introduce plausible deniability. Information engagement may be combined with *norm engagement*, which attempts to limit the perceived offensiveness of behavior. Norm engagement often involves

undermining or delimiting the scope of a standard. Even strongly entrenched norms, such as sovereignty and territorial integrity, are subject to this type of contestation.

To test the effect of the four strategies, we implement an original survey experiment examining how US respondents perceive foreign governments that have violated international law. Our experiment examines reputation management across two issue areas (humanitarian law and trade law) and with two different perpetrators (India and China). To probe the effects of intergovernmental contestation, we randomize not only the foreign government's strategy but also the US government's response.

Our findings provide significant insight into how reputation management affects public opinion. We find that each strategy has the potential to reduce the public's willingness to punish non-compliant states, though they do so in different ways. Strategies that deny responsibility (obfuscate and obstruct) are more effective at limiting public certainty about the culpability of the violating government. However, these strategies also induce greater public outrage, which may translate into more costly punishment. Strategies that acknowledge the violation (atonement and attack) generate an opposite pattern of response. These results hold regardless of the violating government's normative frame, suggesting that information engagement is a more important determinant of public attitudes. We further find that the public's home government is constrained in its ability to shape views about the non-compliant state. Governments that support a foreign violator can reduce their own citizens' demands for retaliation, but doing so creates new reputational costs for the home government itself.

These tradeoffs provide important insight into how intergovernmental contestation may affect international law over the long term. International treaties help countries understand the boundaries of acceptable behavior by shaping norms and providing information. Successful reputation management must then push back at either the facts or the morality of an allegation. While disinformation campaigns and propaganda wars might induce uncertainty, our research suggests they also increase moral outrage among foreign publics, which may

be more willing to impose costly punishments in the wake of violations. To the extent that punishment reinforces the regime, this relationship may support cooperation in the long run.

2 Reputation and International Law

Scholars have long been interested in the role of reputation in international politics. Leaders, governments, and even citizens develop beliefs about the characteristics of foreign actors, based on social cues and previous behavior; these beliefs influence the conduct of foreign policy. Although elites often hold nuanced understandings of individual leaders and government officials, the public is more likely to assign a reputation to a particular country or regime. But even at the country level, reputations are complex and multifaceted. A country might have a general reputation for overarching qualities like consistency and resolve (Weisiger & Yarhi-Milo, 2015), and also an issue-specific reputation that applies to only a subset of topics or agreements (Downs & Jones, 2002). Moreover, different aspects of a country's reputation may matter more or less to different people, depending on their individual values (Brutger & Kertzer, 2018; Morse & Pratt, 2022).

Governments benefit from different reputational attributes depending on the specific area of foreign policy. A strong reputation for resolve allows a country to make more credible threats[3] and affects the credibility of commitments to deter adversaries (Schelling, 1966). A government with a reputation for neutrality, on the other hand, might be well-positioned to broker a bilateral agreement between adversaries. In more technocratic issue areas like financial regulation, a government might leverage its reputation for bureaucratic expertise to craft rules more in line with its values. A negative reputation in one issue area can trigger

---

[3]On the link between credibility and resolve, see Mercer (1996); Tang (2005); Weisiger & Yarhi-Milo (2015) among others.

backlash or consequences on unrelated issues.[4]

While reputations are often treated as fixed in the short term, they are mutable across time, particularly if conduct deviates from expectations. A generally uncooperative state that undertakes costly measures to comply with a treaty, for example, will improve its reputation to a greater degree than a country that always follows international law (Guzman, 2008). Similarly, when a government with a history of defaulting on debt pays its creditors (Tomz, 2007), this action is likely to result in greater reputational gains than when the United States or European Union undertakes a similar action. When a government acts contrary to type in a negative way, this is likely to trigger reputational losses and even the possibility of material penalties. Fear of shaming or stigmatization following negative updating may lead governments to be more cooperative.

If governments generally value having positive reputations for following international law, they should also seek to defend themselves against allegations of wrongdoing. When information surfaces that a country has contravened an established international norm, this can create a reputational crisis for the accused country. Faced with a reputational crisis–that is, information that negatively threatens a state's reputation and could result in material penalties–a government will work to defend itself in order to mitigate reputational damage. Drawing on research on crisis communications, we develop a theory of reputation management and international law wherein governments work to shape foreign attitudes through informational and norm engagement.

3 A Theory of Reputation Management

For bad behavior to generate reputation costs, relevant audiences must interpret the actions as "wrong," negatively update their opinions of the perpetrator, and adjust their behavior

---

[4]For example, Woo & Murdie (2017) find that when human rights organizations publicize poor conditions in debtor countries, the International Monetary Fund is less likely to loan these countries money.

in a way that punishes the violating government. Thus, the consequences of a transgression depend not only the violation itself, but on how the behavior is understood by key audiences. We argue that the contested process of interpreting international law violations therefore provides an opportunity for non-compliant governments to mitigate reputation costs.

We assume that governments alleged to have violated international law will seek to minimize punishment from domestic and international actors. While not all international law violations are politically salient, salience is often linked to costly punishment. A foreign government has fewer incentives to impose sanctions or disrupt trade relations if the violating country's behavior is unimportant to the foreign public. For this reason, reputation management often requires managing public perceptions of a violation. The violating government may present its own, more favorable interpretation of the violation in order to reduce outrage and inspire leniency. It may also encourage other states to delay retaliatory action until outrage dissipates.

3.1 Shaping Reputation to Minimize Punishment

In figure 1, we visualize the strategic environment facing a government confronting accusations of non-compliance. The y-axis in each panel represents the willingness of political actors – most notably, foreign governments or domestic citizens – to punish the government for its behavior. The x-axis measures time since the violation was first revealed. The central challenge confronting governments in this environment is the high level of public outrage in the immediate aftermath of the revelation.[5]

Public outrage is generally decreasing over time, reflected in the downward trend of the line in the left panel of figure 1. We include several small spikes to illustrate how outrage may temporarily increase in response to new information; on average, however, outrage will decay

---

[5]This is an essential scope condition of the theory; violations must be salient enough to generate potential costs for the government.
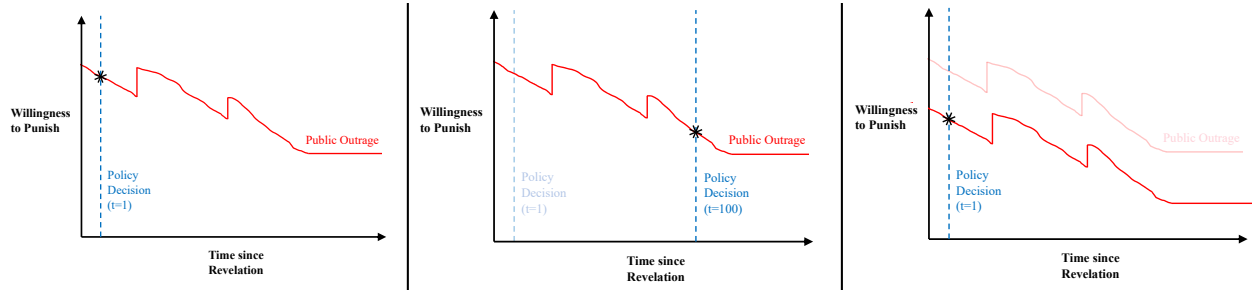
Figure 1: *Punishment in the wake of International Controversies*: The figure visualizes a theoretical model in which public outrage dissipates over time, reducing the amount of costly punishment that governments can sustain (left panel). Violating governments can reduce punishment by delaying policy action (middle panel) or decreasing outrage (right panel).

over time. Both psychological and political processes contribute to this decay. Individually, outrage is difficult to sustain; like other emotional states, it often involves a surge of intense activation that subsequently abates (Gross, 2015; Goldenberg *et al.*, 2020). Like other forms of recency bias, anger over transgressions fades as "memory of past atrocities mingles with daily experience" (Petersen & Zukerman, 2010).

Outrage is also less likely to translate into a meaningful political response as time lapses. An initial, concentrated surge of citizen anger can sometimes push governments to drastically change policies to punish a violating country. Over time, however, opposing interests are able to mobilize and take action to reinforce the status quo. Eventually, the level of public outrage stabilizes into a new equilibrium that is higher than before the violation but substantially lower than the initial revelation.

The decay of public outrage means that the level of punishment imposed on violating governments depends, in part, on the timing of policy decisions. In the left panel of figure 1, a foreign government crafts a policy response to the international law violation at $t = 1$, soon after the behavior is revealed. The foreign government's policy choice is informed by the high level of public outrage, yielding a relatively costly response (represented by the black star). One clear way for the violating government to reduce punishment is to induce delay in the policy decision (middle panel). Since time is on the violator's side, delay pushes

the policy decision into a more advantageous environment. The net effect is a lower level of punishment.

A second way to limit punishment by political actors is to reduce the level of public outrage. Governments accomplish this by articulating public justifications that persuade audiences to interpret their behavior in a more favorable light. Communications research emphasizes two judgments that shape how harshly audiences view a scandal-plagued actor: blame and offensiveness (Pomerantz, 1978). Blame reflects the level of perceived responsibility. When an actor is clearly responsible for a transgression, the behavior will generate more backlash than when the facts are more ambiguous. Similarly, when the behavior contravenes a widely accepted norm, the level of outrage will be larger than less offensive actions. If violating governments can successfully reduce perceptions of blame or offensiveness, they shift the level of public outrage downward (right panel). As above, the result is a decrease in the expected level of punishment.

These examples show two distinct mechanisms through which violators shape how their behavior is judged and acted upon. We call this process "reputation management." We now turn to the set of rhetorical strategies states use when managing the reputational fallout from international law violations.

3.2 A Typology of Reputation Management Strategies

Counteracting a reputational threat depends on reducing either perceived responsibility or offensiveness (Benoit, 2015, 11). An individual working to reduce perceived responsibility might challenge the underlying facts of a violation. This approach could include everything from an outright denial to introducing new facts that help limit responsibility. During the North Atlantic Treaty Organization's (NATO) 1999 bombing campaign in Serbia, for example, NATO missiles killed civilians and struck the Chinese embassy. In response to these events, NATO officials minimized perceived responsibility by arguing these targets

were hit unintentionally as "a weapon went astray and hit civilian buildings."[6]

Governments can also attempt to reduce the moral outrage associated with their behavior.[7] A government might bolster its image by citing previous good behavior – an approach common among Western democracies when accused of human rights violations. An actor might also try to reduce the magnitude of negative sentiment by citing additional context or suggesting observers are categorizing the behavior in the wrong way. Contestation over international law in particular is prone to such renegotiations; even the boundary of something as longstanding as "sovereignty" has shifted over time as states have redefined the norm to incorporate the "Responsibility to Protect" doctrine. Other ways of reducing offensiveness include suggesting the behavior aligns with alternative values or attempting to compensate for bad behavior.

In the context of international law violations, we argue that reputation management strategies vary along two primary dimensions: information and norms. Informational engagement is designed to manipulate the level of certainty about a government's behavior.[8] At one end of the spectrum, a government can acknowledge an allegation as completely true. At the other, a government can completely deny the incident. While governments do sometimes employ complete acceptance or full denial, many reputation management strategies fall between these extremes as governments contextualize their actions.

In addition to informational engagement, a government can choose to accept or reject the norm, rule, or law that has been transgressed. We refer to this process as norm engagement.[9]

---

[6]Daniel Williams, "Missiles Hit Chinese Embassy," 8 May 1999, *The Washington Post*, retrieved from: https://www.washingtonpost.com/wp-srv/inatl/longterm/balkans/stories/belgrade050899.htm.

[7]The strategies described in this paragraph map onto the discussion of how to reduce offensiveness in Benoit (2015, 28).

[8]This corresponds to what Pomerantz (1978) terms "blame" and what Fishbein & Ajzen (2010) describe as "beliefs".

[9]Norm engagement corresponds with Pomerantz's discussion of offensiveness and Fishbein and Ajzen's discussion of values.

If a government wants to reduce the offensiveness of its behavior, challenging or redefining a norm may be an effective strategy. As with informational engagement, norm engagement falls along a continuum from acceptance to rejection, but most observed behavior falls somewhere in between. Many instances of norm engagement implicitly undermine a specific norm by arguing transgressions are frequent and inconsequential. Norm engagement may also be more outward-focused, as when one country cites another's bad behavior. When the US government criticized China for failing to show up to the COP26 climate summit in November 2021, for example, a Chinese foreign ministry retorted "Actions speak louder than words. What we need...is concrete action rather than empty words. China's actions in response to climate change are real."[10]

A government's reputation management strategy must contend with both the informational and normative environment. In an idealized form, this suggests there are four main response strategies, presented in the 2×2 diagram in figure 2. Governments that acknowledge responsibility for the behavior can either accept the norm in question (atonement strategy) or reject it (attack strategy). Alternatively, a government may deny its bad behavior while accepting the underlying norm (obfuscation) or both deny and reject the norm (obstruction). Each of these strategies has the potential to minimize punishment by reducing outrage or inducing delay in the policy choices of key actors. Below, we describe each of them in greater detail.

Atonement

An atonement strategy occurs when a government works to improve its reputation by making amends for bad behavior. At its core, this response includes a government accepting responsibility for the violation and thus acknowledging the facts of the case. Informational

---

[10] "China fires back after Biden slams Xi's COP26 no-show," *Al Jazeera*, 4 November 2021, retrieved from: https://www.aljazeera.com/economy/2021/11/4/china-fires-back-after-biden-slams-xis-cop26-no-show.

Informational Engagement

	Acknowledgment	*Denial*
Endorsement	Atonement	Obfuscation
Rejection	Attack	Obstruction

Norm Engagement appears to the left, labeling the rows.

Figure 2: *Typology of Reputation Management Strategies.*

engagement may be explicit or implicit. A government may express regret that an action occurred without explicitly accepting blame, as corporations often do when confronting large scandals. In practice, though, even implicit atonement still signals responsibility and informational acknowledgement.[11]

Atonement may also engage with established norms in implicit and explicit ways. A government may ask for forgiveness, an act of formal apology that implicitly endorses a norm.[12] While this strategy implies that a government supports better behavior, it is a less significant type of norm endorsement than an explicit commitment to follow the rules in the future. In 2008, for example, the Canadian Prime Minister formally apologized to and asked forgiveness of indigenous Canadians who were removed from their families and forced to attend residential schools. While his apology was considered a victory for indigenous communities, he was also criticized for failing to endorse the United Nations Declaration on the Rights of Indigenous Peoples.[13]

Atonement has several strategic advantages. Apologies are more effective at protecting reputation than many other strategies, including no comment, denial, excuse, or justification

---

[11]See the discussion of sympathy, compensation, and accepting responsibility in Coombs & Holladay (2008).

[12]Benoit and Drew (1997) and Fuchs-Burnett (2002) suggest an apology is marked by both responsibility and a request for forgiveness .

[13]"Government apologizes for residential schools in 2008," CBC Archives, 25 June 2018, retrieved from: https://www.cbc.ca/archives/government-apologizes-for-residential-schools-in-2008-1.4666041.

(Bradford and Garrett 1995, Dean 2004). Other atonement stratements, such as compensating victims or expressing sympathy, perform similarly in experimental tests (Coombs & Holladay, 2008). A government may also prefer atonement when it violates a particularly important strand of international law. If non-compliance poses a threat to regime longevity, then self-accountability and norm reinforcement are ways to counteract these negative effects.

Atonement also carries risks that may outweigh potential benefits. By admitting that a violation occurred, a government removes any uncertainty about its responsibility. Acknowledging the violation risks damaging the perceived strength of relevant international law, even if it also tries to reinforce the norm. Moreover, while apologies, sympathy, and compensation may be appealing to foreign audiences, a government's domestic public may punish a government for pursuing this approach. Morse & Pratt (2022) find that while apologies have a positive impact on a government's image, they are less effective than attack strategies at signaling that a government looks out for its citizens. Contrition can even trigger a domestic backlash, which may alarm former adversaries and potentially increase tensions (Lind, 2008).

Attack

An attack strategy occurs when a government acknowledges that it broke international law but contests the underlying norm. The core of this approach is a moral justification or defense – rather than challenging information, a government works to reduce outrage by recasting its behavior as less offensive. Direct norm engagement might include calling a particular law or treaty unfair or unjust, or attacking the other side as illegitimate or evil. Indirect norm engagement might include providing additional context or details that justify or explain the offensive action.

In the context of international law, attack strategies often involve attempts to redraw the boundaries of norms rather than directly challenging them. Even overtly anti-globalist leaders may avoid norm dismantlement, preferring instead to cite competing principles or

standards. When Trump imposed tariffs on steel and alumnium in 2018, for example, he justified the move under a Cold War-era law that allows the US president to adjust imports that threaten national security.[14] Rather than openly reject the norm of free trade, Trump's general rhetoric with respect to the World Trade Organization focused on accusing other countries of violating norms of fairness and reciprocity (Carnegie & Carson, 2019).

Attack strategies that attempt to redraw the boundaries of norms are similar to what Alter (2022) terms "extra-ordinary contestation" over international law – contestation that takes place outside the formal boundaries of traditional rule of law. Such practices include pitting competing international laws against each other, using domestic law to supplant international authority, and directly attacking the legitimacy of international law. Alter argues that extra-ordinary contestation occurs when international law is domestically invasive or politically constraining, or when actors seek to change the international system itself.

Obfuscation

Obfuscation occurs when a government's reputation management strategy focuses primarily on denying the facts of a violation. Denials might include calling information about an allegation "fake news," attacking an information source or media outlet as biased, or suggesting an incident was staged to make a country look bad. Russia's invasion of Ukraine is filled with examples of reputation management via obfuscation. When gruesome reports surfaced suggesting Russian troops executed civilians in Bucha, the Russian defense ministry claimed the scenes were faked and a provocation. Russian Foreign Minister Sergey Lavrov even went so far as to demand a meeting of the Security Council because "we see such provocations as

---

[14]David Lawder, "US Court upholds Trump's national security tariffs on steel imports," 4 Feb 2021, *Reuters*, retrieved from: https://www.reuters.com/article/us-usa-trade-steel/u-s-court-upholds-trumps-national-security-tariffs-on-steel-imports-idUSKBN2A42MN.

a direct threat to international peace and security."[15]

Denial can also be direct rejections of responsibility rather than the facts themselves. When the Iranian military accidentally downed an airliner in January 2020, for example, Iranian government officials denied initially that missiles could have downed the aircraft and blamed a mechanical malfunction.[16] As western intelligence agencies began to contradict the report, Iran was eventually forced to change its approach and acknowledge the mistake.

We assume that when a government engages in an informational denial without engaging with the relevant norm, this strategy implicitly accepts the norm itself. For this reason, obfuscation can be a relatively constraining strategy depending on context. If significant ambiguity exists, a government may easily deny an allegation of wrongdoing or call a story fake news. In these cases, obfuscation is likely to be particularly effective at inducing delay while more information is gathered. For high stakes issues or those subject to outside monitoring, however, other parties may be able to prove that a denial is a lie. Faced with proof, a government may continue to espouse the same obfuscation rhetoric or it may be forced to change strategies. Because of this potential downside, governments sometimes pair informational denials with norm rejection – a strategy that we call "obstruction."

Obstruction

Obstruction is the most antagonistic reputation management strategy. When a government relies on obstruction, it challenges both facts and norms. Informational engagement and norm engagement are often linked in a way that reinforces each other. A government may reject an allegation through a combination of shifting blame and citing a competing norm.

---

[15] "Russian claims Bucha civilians massacre faked as "provocation" as outrage builds over Ukraine war atrocities," 4 April 2022, *CBS News*, retrieved from: `https://www.cbsnews.com/news/bucha-massacre-ukraine-russia-putin-provocation-war-crimes-atrocities/`.

[16] Bethan McKernan, 11 Jan 2020, "Iran admits unintentionally shooting down Ukrainian airliner," *The Guardian*, retrieved from: `https://www.theguardian.com/world/2020/jan/11/iran-admits-shooting-down-ukrainian-airliner-unintentionally`.

Brazilian President Jair Bolsonaro, for example, was criticized for his environmental record and his desire to open the Amazon to business interests. When record fires burned the rainforest in 2019, he claimed that non-governmental organizations (NGO), rather than farmers, were starting the fires to hurt his government because his administration had cut NGO funding. In a Facebook Live broadcast, he told reporters "Crime exists. These people are missing the money."[17]

As with attack strategies, which also challenge norms, obstruction approaches work to lessen the offensiveness of behavior in different ways. A government might talk about how a particular rule is unfair or ought to be changed in an effort to convince observers that, even if the behavior did occur, it wouldn't be so bad. More likely, however, a government will pair an informational denial with redrawing the boundaries of a norm as these two approaches can reinforce each other. When the Bush administration wanted to claim its enhanced interrogation techniques weren't torture and also weren't violations of the Geneva Conventions, for example, US government officials redefined both the meaning of torture and the scope of who constituted a prisoner of war. If norms are redrawn, then a government can claim the violation itself never occurred.

3.3 Reputation Management, Contestation, and Foreign Publics

As the preceding discussion underscores, mitigating punishment requires countering negative attitudes among core audiences. The set of relevant audiences for governments engaged in reputation management is quite varied. If an issue is normatively important or tied to core domestic interests, domestic actors like foreign policy elites, NGOs, and citizens may all be

---

[17]Anthony Boadle and Gabriel Stargardter, 21 August 2019, "Igniting global outrage, Brazil's Bolsonaro baselessly blames NGOs for Amazon fires," *Reuters*, retrieved from: https://www.reuters.com/article/us-brazil-politics-idUSKCN1VB1BY.

willing to punish a government for violating international law.[18] NGOs, for example, highlight non-compliance with environmental agreements (Raustiala, 1997; Von Stein, 2008; Dai, 2007) and human rights treaties (Keck & Sikkink, 1998; Simmons, 2009; Hafner-Burton & Tsutsui, 2005; Risse & Sikkink, 1999) in ways that make the public more aware of violations, shape the process of legal contestation, and increase political costs. At a more general level, research on audience cost theory suggests citizens may punish leaders who renege on foreign policy commitments.[19]

Governments also have strong incentives to punish other governments for violating international law. International regimes provide standards of behavior against which another country's performance can be measured (Keohane, 1984); such standards make it easier to identify instances of non-compliance. When countries form an agreement, they expect that all sides will generally follow through on their commitments. In such contexts, punishing violations is not just retaliatory action, but also reflects an effort to prevent future non-compliance.

Governments have a variety of tools at their disposal to punish countries that violate international agreements. Foreign policy elites may engage with each other, reaching across geographic boundaries to deliver words of condemnation or expectations of better behavior. Governments may also opt to "name and shame" a country that violates international law, particularly if the violation is egregious and publicly salient. Stronger forms of retaliation include referring a situation to an international organization, limiting or terminating financial assistance, deploying sanctions, or even (in rare contexts) using military force to punish violators.

As governments consider the policy response to international law violations, they pay

---

[18]On how international legal commitments shape public support for policies, see Chaudoin (2014); Powers (2022); Strezhnev *et al.* (2019); Kreps & Wallace (2016); Chu (2019); Chilton (2014); Lupu & Wallace (2019); Wallace (2013) among others.

[19]For work on audience costs, see Fearon (1994); Smith (1998); Trager & Vavreck (2011) among others.

close attention to the attitudes of their own citizens. Most of the policy options available to punish non-compliant states are costly, and citizens bear the primary burden of punishment. A government can criticize and censure via backchannels with few domestic consequences, but as retaliatory action becomes more public, a government will want its citizens' support. This is particularly true for retaliatory action that imposes economic costs on the public or counters public values in key ways. The public is thus a key constraint in determining possible avenues of punishment, and an important target for reputation management.

Public opinion is likely to matter most when violations are particularly high stakes. If a government is considering punishing a country for a violation in a significant way, its leadership is likely to consider how the public views the crisis at hand. Existing research suggests that how the public perceives a foreign country affects citizens' policy views toward that country. Goldsmith & Horiuchi (2012) find that other countries' public attitudes toward US foreign policy affect the willingness of the public to support troop commitments to Iraq.

For high-stakes policy issues, other governments will have their own strategic priorities, and they are likely to intervene in the reputation management process by contradicting or supporting the violating government's narrative. Translating our theory of reputation management to this context of intergovernmental contestation, we expect that the violating government will use informational and normative engagement to reduce the foreign public's willingness to punish and to delay action, but the strength of impact will depend on how the public's own government responds to the violation. For a government to opt for costly punishment, the public must view the alleged violation as egregious and most likely true. The government's response is likely to shape both perceptions, and may translate into support for punishment.

4   Testing Reputation Management

We test how the four reputation management strategies affect citizens' willingness to punish foreign governments in an original survey experiment. The survey presents respondents with a hypothetical international law violation by a foreign government. An experimental design allows us to randomly assign the rhetorical justification offered by the government, avoiding the selection bias that plagues observational studies. In adopting this approach, we build on a growing body of work leveraging experimental methods to assess the effect of international law on public attitudes (e.g., Chaudoin (2014); Chilton (2014); Zvobgo (2019); Powers (2022); Brutger & Strezhnev (2022); Morse & Pratt (2022)).

4.1   Hypotheses

Our theory predicts that citizen perceptions of international law violations — and thus their preferences for punishing violating governments — are shaped by the framing strategies adopted by key political actors. We have four primary theoretical expectations regarding the effects of reputation management.

First, we hypothesize that each of the response strategies described in section 3 will decrease respondent willingness to punish the violator, compared to no justification (*H1*). Although the four strategies vary in the underlying logic they cite, each is designed to minimize blame in the wake of the alleged violation. We utilize three primary measures of respondents' preferences for punishment: support for imposing sanctions on the foreign government, willingness to bear economic costs arising from sanctions, and support for delaying retaliatory action until more facts are known. Response strategies should reduce these outcome measures compared to the control condition.

Second, we expect that strategies will operate by shaping two important perceptions among respondents: *certainty* that the violation occurred and *outrage* about the alleged

behavior. Government denials of the allegation — represented in our experiment by the obfuscation and obstruction strategies — should reduce respondent certainty ($H2$). Strategies that contest the underlying norm involved in the violation (attack and obstruction strategies) should reduce respondent outrage ($H3$).

Third, both certainty and outrage shape respondents' willingness to punish the foreign government. These perceptions are key mechanisms through which reputation management strategies shape preferences for punishment. We expect that greater certainty about the violation and higher levels of outrage will be associated with more punitive attitudes toward the foreign government ($H4$).

Finally, citizens' support for punishment are also shaped by cues from their home government. We hypothesize that condemnatory statements from the US government will increase US respondents' willingness to punish foreign violators, while supportive statements by the US government will reduce support for punishment ($H5$).

4.2 Survey Methodology

We administered the survey in July 2022 to a nationally representative sample of 2,883 US-based respondents.[20] The experiment consists of two hypothetical scenarios involving violations of international law. Respondents are presented with the first scenario, then view one or more rhetorical messages from political actors, and provide their opinion on retaliatory punishment for the violating government. They then proceed to the second scenario, featuring a different government and a different violation, and again provide their views on punishment. We randomly assign the order of the scenarios and the identity of the foreign government (China or India).

The two scenarios depict violations of humanitarian and trade law. We include both to assess whether the effects of reputation management strategies vary by the strategic setting

---

[20]Respondents were recruited via Lucid Theorem. See the appendix for sample characteristics.

or underlying norms of different policy domains. To minimize the effect of respondent views on current US or foreign political leadership, we set each scenario ten years in the future. The humanitarian law scenario describes alleged war crimes:

> International law prohibits targeting civilians during military conflict. Governments are supposed to avoid attacking non-combatants, civilian infrastructure, and medical facilities.
>
> Imagine it is ten years in the future, and [China/India] is fighting a war in another country. Media sources report that a barrage of missiles that appeared to originate from [Chinese/Indian]-controlled territory struck several civilian targets. The attack destroyed a large hospital and an apartment complex, killing hundreds of civilians.

The trade law scenario describes the imposition of tariffs:

> International law prohibits imposing trade restrictions under certain conditions. Governments that sign on to trade agreements are supposed to avoid placing unauthorized taxes on foreign imports.
>
> Imagine it is ten years in the future and a trade dispute emerges. Media sources report that [China/India] has imposed a tax on imported products, making foreign goods more expensive than similar products from [China/India]. Multinational firms claim that the tax is putting them at a competitive disadvantage.

Respondents in the control condition receive no justification from the foreign government. Those in the treatment conditions are presented with a rhetorical justification from the foreign government. The justification takes one of four general forms. We present the language from the humanitarian law scenario here and include the full survey text in the appendix.

- **Atone (Acknowledgement, Norm Acceptance)**: The [Chinese/Indian] government acknowledged the attack, calling it a terrible mistake and pledging to protect civilians going forward.

20

- **Attack (Acknowledgement, Norm Challenge)**: The [Chinese/Indian] government acknowledged the attack but argued that it should not be held responsible for civilian deaths since its enemies use civilians as human shields.

- **Obfuscate (Denial, Norm Acceptance)**: The [Chinese/Indian] government denied the attack, calling the story a lie designed to cast blame on [China/India].

- **Obstruct (Denial, Norm Challenge)**: The [Chinese/Indian] government denied the attack, calling the story a lie designed to cast blame on [China/India]. The [Chinese/Indian] government further argued that it should not be held responsible for civilian deaths since its enemies use civilians as human shields.

Respondents are then randomly assigned to a second treatment condition describing a response from the US government. The response may be neutral, supportive, or condemnatory of the violator government.

- **Neutral**: The United States government said it was closely following the situation.

- **Support**: The United States government expressed support for [China/India] and criticized efforts to politicize the current crisis.

- **Condemn**: The United States government condemned [China/India], calling its behavior a serious violation of international law.

After viewing the scenario, respondents proceed to an outcome questionnaire. Our primary outcomes measure a respondant's support for economic sanctions and their willingness to delay retaliation. We inform respondents that the US government is considering its policy response to the violation, which include taking no action, imposing economic sanctions, conducting drone attacks on the violator, and other responses. Respondents select which options they support. Our first punishment measure (`sanctions`) records whether they select

the sanctions option. Separately, we asses their willingness to bear direct costs associated with sanctions. For that measure (`costly sanctions`), we suggest that sanctioning will increase US gas prices and ask respondents the maximum price increase per gallon they are willing accept.[21] Finally, we assess willingness to postpone punishment (`delay`) by asking whether respondents support delaying retaliatory action until an international investigation can confirm more facts.[22]

We measure perceived certainty and outrage about the violation by measuring level of agreement with two statements:"The [Chinese/Indian] government definitely conducted the attack" and "If the [Chinese/Indian] government conducted the attack, it was morally wrong," respectively. Finally, although our primary interest is in respondent views about the foreign government, we include a question gauging support for the position of the home (US) government in the scenario.

5 Results

We begin by testing how each reputation management strategy affects citizens' preferences for punishment. Figure 3 presents the treatment effects and 95% confidence interval on support for sanctions, willingness to bear costs associated with sanctions, and support for delayed retaliation. Estimates reflect the difference in respondent views between the control group (no foreign government response) and the four response types (*Atone*, *Attack*, *Obfuscate*, and *Obstruct*). For ease of comparison, we standardize the outcome measures.

The results are mixed for hypothesis *H1*, which predicts that all strategies should reduce punitive attitudes about the foreign government. We find that each strategy significantly decreases respondents' willingness to impose economic sanctions. *Atone* has the largest substantive effect, reducing the likelihood of backing sanctions by 0.2 standard deviations

---

[21]Response options are 0 (no increase), an increase of 10 cents per gallon, an increase of 20 cents per gallon, and an increase of 50 cents per gallon.

[22]Respondents select among a 5-point scale from "strongly oppose" to "strongly support."

(10 percentage points) compared to the control group. The remaining strategies diminish the likelihood of supporting sanctions by roughly six percentage points. When we underscore the direct economic cost of sanctions, however, only *Atone* has a statistically significant effect. It reduces tolerance for a sanctions-related increase in gas prices by 0.13 standard deviations (2.1 cents). No response strategy clearly affects willingness to delay retaliation, though the effect of *obfuscate*, a strategy where a government denies the allegation, approaches conventional levels of statistical significance ($p = 0.06$).

These results suggest that atonement strategies are the most effective at minimizing short-term demand for retaliation against international law among foreign publics. This finding complements existing work demonstrating that atonement postures can reduce backlash among dovish domestic audiences but not among hawks (Morse & Pratt, 2022). There is also some evidence that strategies differ in their ability to shape short-term preferences for punishment vs. their ability to encourage a delay in judgment.

The next set of tests interrogate the effect of each strategy on perceptions of certainty and outrage surrounding the violation. We hypothesized that strategies involving denials should decrease respondent certainty that the violation occurred as reported, while strategies that contest the underlying norm should reduce outrage about the alleged behavior.

Figure 4 presents the treatment effects of each strategy on these dimensions. The top panel displays the estimated effect on respondents' belief that the violating government "definitely" committed the violation. The results do not strictly align with expectations: the two response strategies that deny the behavior (*Obfuscate* and *Obstruct*) have no significant effect on respondent certainty. The limited impact of these informational challenges on certainty, compared to the control condition, may be explained by the ambiguity of the control condition itself. In contrast, the two strategies that explicitly acknowledge the violation (*Atone* and *Attack*) substantially *increase* certainty that the violation occurred.[23]

---

[23]*Atone* increases perceptions of certainty by approximately 0.4, and *Attack* by 0.5, on the 5-point scale.
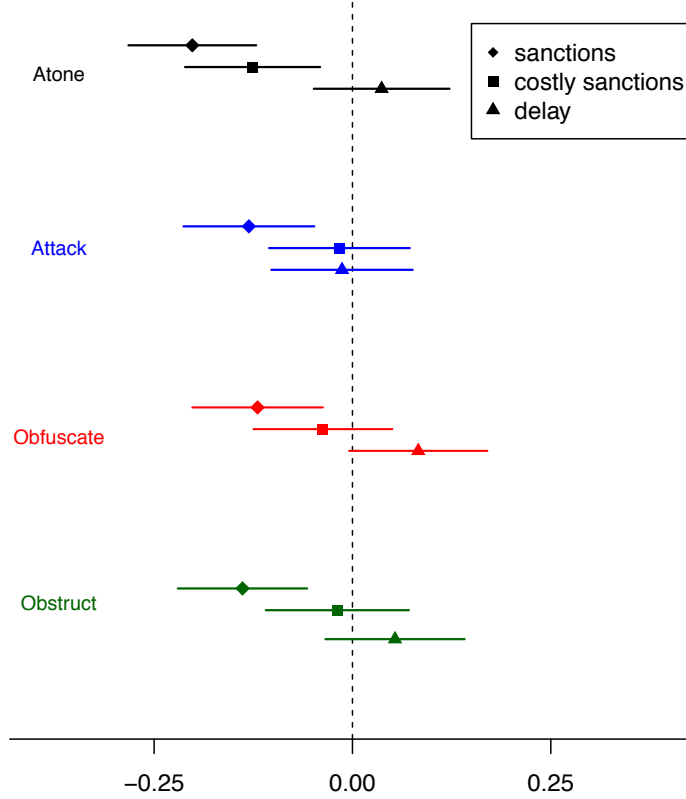
Figure 3: *Effect of Response Strategies on Respondent Willingness to Punish*: The figure shows the treatment effect of each reputation management strategy on respondent willingness to sanction, willingness to bear sanction-related costs, and willingness to delay punishment. All effects are relative to the control condition (no government response). Point estimates and 95% confidence intervals are displayed for each treatment.

In the bottom panel, the effects are reversed. Measures of moral outrage are unaffected by *Atone* and *Attack*, while *Obfuscate* and *Obstruct* increase respondent outrage about the violation. Together, these results indicate a tradeoff between strategies that acknowledge the violation and those that deny it. Acknowledgment strategies increase certainty about the alleged behavior but do not deepen outrage, while denials have the opposite effect.

In Table 1, we investigate the relationship between certainty, outrage, and willingness to punish the foreign government. *H4* predicts that greater certainty and higher levels of outrage will be associated with more punitive attitudes. We test this hypothesis in a linear model regressing the punishment measures on perceptions of certainty and outrage. While
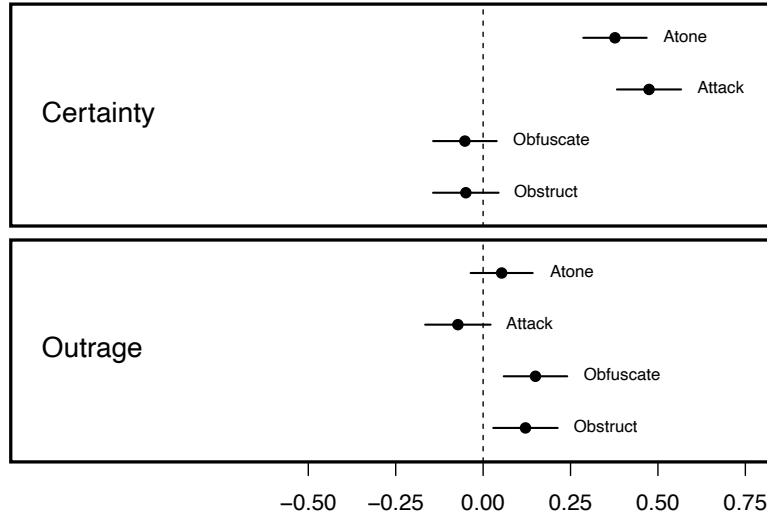
Figure 4: *Effect of Response Strategies on Respondent Certainty, Outrage*: The figure shows the treatment effect of each reputation management strategy on perceptions of certainty and outrage about the violation. All effects are relative to the control condition (no government response). Point estimates and 95% confidence intervals are displayed for each treatment.

these results are not causally identified (we do not randomly assign certainty and outrage), they provide insight into the mechanisms driving preferences for punishment.

	Sanctions	Costly Sanctions	Delay
	(1)	(2)	(3)
Certainty	0.056***	1.080***	−0.033*
	(0.007)	(0.247)	(0.017)
Outrage	0.070***	1.929***	0.018
	(0.006)	(0.236)	(0.017)
Observations	5,024	4,942	4,955
Adjusted R^2	0.047	0.023	0.001

Table 1: *Effect of Respondent Certainty, Outrage on Preferences for Punishment.* Results of linear models predicting willingness to punish on the basis of respondent perceptions of certainty and outrage. Standard errors are clustered by respondent. Statistical significance is denoted by $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

The results are largely consistent with expectations. Perceptions of certainty and outrage are strongly linked to support for sanctions and tolerance for sanctions-related costs. In terms of effect size, a one-unit increase in the 5-point certainty scale is associated with a 5 percentage point increase in the likelihood of supporting sanctions and a 1 cent increase in willingness to bear gas price increases. The effect of outrage is slightly larger, increasing support for sanctions by 7 percentage points and tolerance for gas prices by 1.9 cents. Only perceptions of certainty shape respondents' willingness to delay retaliation until more facts are confirmed.

Finally, we assess whether the home government's response can undermine or reinforce reputation management by the foreign government. Figure 5 displays the estimated effect of US government condemnation and support for the foreign government on punitive attitudes. These effects are relative to a neutral US government statement indicating that it is "closely following the situation."

We find that condemnation (top panel) has a null effect on respondents' preferences for punishment. When the US government criticizes the foreign government, it does not meaningfully shape calls for sanctions or willingness to delay punishment. Supportive statements substantially reduce support for economic sanctions ($p < 0.01$) but do not affect willingness to bear sanction-related costs or tolerance for delayed action. The effect of the US government response does not interact with the identity of the violating government (India or China) or the reputation management strategy adopted by the foreign government.

Although our central focus is on public support for punishing foreign violators, we also examined whether the home government has its own political incentives to condemn or support international law violations. Table 2 displays the effects of the US government posture on respondent approval of the US government response. Column 1 is a baseline model showing the treatment effect of US statements of condemnation and support. While condemning foreign violators has no effect on the US government's domestic approval, statements of sup-
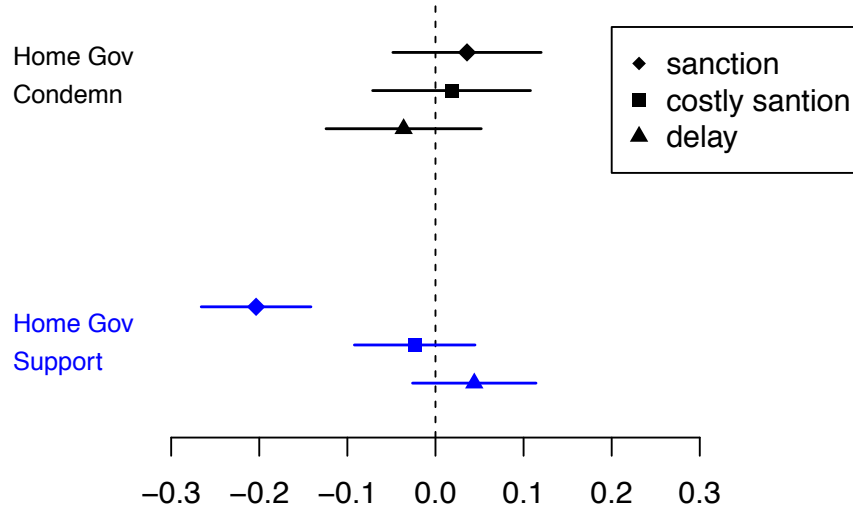
26

Figure 5: *Effect of Home Government Condemnation and Support on Willingness to Punish*: The figure shows the effect of statements of support from the home government on respondent willingness to punish the violating government. Effects are relative to a neutral home government response. Point estimates and 95% confidence intervals are displayed for each treatment.

port significantly undermine the government's standing among its own citizens. In Column 2, we test whether this effect is conditioned by the identity of the foreign government. We find that US statements of support are particularly damaging to its domestic standing if the violator is China. Finally, Column 3 examines whether foreign government *Atonement* strategies can mitigate the costs of supportive statements. The results suggest that they do, potentially pointing to one additional advantage of this strategy by foreign governments.

| | DV: Home Government Approval | | |
	(1)	(2)	(3)
Condemn	0.033	0.019	0.045
	(0.046)	(0.068)	(0.053)
Support	−0.539***	−0.429***	−0.619***
	(0.040)	(0.056)	(0.047)
China		−0.014	
		(0.039)	
Condemn × China		0.027	
		(0.094)	
Support × China		−0.223***	
		(0.079)	
Atone			−0.108*
			(0.063)
Condemn × Atone			0.011
			(0.109)
Support × Atone			0.349***
			(0.094)
Observations	4,267	4,267	4,267
Adjusted R^2	0.049	0.051	0.052

Table 2: *Effect of Home Government Response on Approval of Home Government.* Estimated treatment effects of home government statements of support and condemnation on respondent approval of the home government. Standard errors are clustered by respondent. Statistical significance is denoted by *p<0.1; **p<0.05; ***p<0.01.

6 Conclusion

This paper examines contestation over non-compliance with international law. We argue that the choice to punish violating governments is inherently linked to how audiences interpret the alleged transgression. Acts of non-compliance trigger a political process in which the violator, foreign governments, and other actors attempt to strategically shape perceptions of the violation to their advantage.

We articulate a theory of reputation management in which non-compliant states seek to minimize punishment by others. They have two primary mechanisms for accomplishing this goal. First, they can "run out the clock," inducing delay in retaliation until the initial outrage has subsided. Second, they can persuade others in the near term that their behavior does not warrant a strong response.

We introduce a typology of reputation management strategies that governments use to induce delay and reduce willingness to punish. These strategies differ along two dimensions: whether the government acknowledges or denies the allegation (information engagement), and whether it embraces or attacks the underlying norm (norm engagement). This categorization scheme identifies four distinct strategies that states use in the wake of international law violations.

The four strategies are tested in a survey experiment on nearly 3,000 American respondents. While all four strategies reduce respondents' willingness to punish a violating government, they affect public attitudes in different ways. When a government acknowledges a violation (atonement and attack), its rhetoric increases certainty but has no significant effect on moral outrage. When a government denies a violation (obfuscate and obstruct), its rhetoric has no effect on certainty but it generates significant increases in moral outrage, regardless of its normative frame. These findings suggest that informational engagement is a more important determinant of public attitudes than normative engagement, perhaps

because norms are more fixed in the short-term. Interestingly, however, moral outrage has the strongest association with costly retaliation, suggesting that morality and norms are important for understanding the public's willingness to punish a perpetrator.

We further establish that both certainty and outrage are linked to preferences about punishment. When the public is more certain that an event took place, citizens are more willing to deploy sanctions and less supportive of delay. A violating government thus faces a tradeoff: if it denies its bad behavior, it will increase moral outrage, which is strongly associated with a willingess to support costly sanctions, but the public may support delaying action until further inquiry. But if it acknowledges the violation, a portion of the public still supports punishment and delay is less likely. In other words, accept a violation and face weak immediate punishment, or deny a violation and face the possibility of stronger action in the medium term.

Our results provide new insights on the link between international law, domestic politics, and enforcement. While scholars have long cited reputational mechanisms as a positive force for inducing compliance, our work suggests these mechanisms are not automatic. The political process of justification, persuasion, and contestation that accompanies violations is key to shaping the severity of reputation costs. This is especially true for costly punishment like economic sanctions, which require maintaining public support.

The theory and empirical results we present suggest new avenues for additional research. Future work should examine the conditions under which states draw upon particular reputation management strategies. The frequency with which states employ each strategy is important for understanding reputation costs, but also for public attitudes regarding international law more generally. If states increasingly respond to alleged violations by attacking international norms, for example, this could erode the legitimacy of multilateral institutions and the existing international order. We also encourage additional theoretical development on how public opinion may shape punishment and retaliation related to international law.

References

Benoit, William L. 2015. *Accounts, Excuses, and Apologies (2nd Edition): Image Repair Theory and Research.* SUNY Press.

Brutger, Ryan, & Kertzer, Josh. 2018. A Dispositional Theory of Reputation Costs. *International Organization*, **72**(3), 693–724.

Brutger, Ryan, & Strezhnev, Anton. 2022. International Investment Disputes, Media Coverage, and Backlash Against International Law. *Journal of Conflict Resolution*, 00220027221081925.

Carnegie, Allison, & Carson, Austin. 2019. Reckless Rhetoric? Compliance Pessimism and International Order in the Age of Trump. *The Journal of Politics*, **81**(2), 739–746.

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

Chilton, Adam S. 2014. The influence of international human rights agreements on public opinion: An experimental study. *Chi. J. Int'l L.*, **15**, 110.

Chu, Jonathan A. 2019. A clash of norms? How reciprocity and international humanitarian law affect American opinion on the treatment of POWs. *Journal of conflict resolution*, **63**(5), 1140–1164.

Coombs, W. Timothy, & Holladay, Sherry J. 2008. Comparing apology to equivalent crisis response strategies: Clarifying apology's role and value in crisis communication. *Public Relations Review*, **34**(3), 252–57.

Dai, Xinyuan. 2007. *International Institutions and National Policies.* Cambridge: Cambridge University Press.

Downs, George W, & Jones, Michael A. 2002. Reputation, Compliance, and International Law. *The Journal of Legal Studies*, **31**(Jan.), S95–S114.

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

Fishbein, M., & Ajzen, I. 1977. Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological bulletin.*

Fishbein, Martin, & Ajzen, Icek. 2010. *Predicting and Changing Behavior: The Reasoned Action Approach.* Psychology Press.

Goldenberg, Amit, Garcia, David, Halperin, Eran, & Gross, James J. 2020. Collective emotions. *Current Directions in Psychological Science*, **29**(2), 154–160.

Goldsmith, Benjamin E., & Horiuchi, Yusaku. 2012. In Search of Soft Power: Does Foreign Public Opinion Matter for US Foreign Policy. *World Politics*, **64**(3), 555–85.

Gross, James J. 2015. Emotion regulation: Current status and future prospects. *Psychological inquiry*, **26**(1), 1–26.

Guzman, Andrew T. 2008. *How International Law works: A Rational Choice Theory.* Oxford University Press.

Hafner-Burton, Emilie M., & Tsutsui, Kiyoteru. 2005. Human Rights in a Globalizing World: The Paradox of Empty Promises. *American Journal of Sociology*, **110**(5).

Keck, Margaret E, & Sikkink, Kathryn. 1998. *Activists beyond Borders: Advocacy Networks in International Politics.* Ithaca and London: Cornell University Press.

Keohane, Robert O. 1984. *After Hegemony.* Cooperation and Discord in the World Political Economy. Princeton University Press.

Kreps, Sarah E, & Wallace, Geoffrey PR. 2016. International law, military effectiveness, and public support for drone strikes. *Journal of Peace Research*, **53**(6), 830–844.

Lind, Jennifer. 2008. *Sorry States: Apologies in International Politics*. Cornell University Press.

Lupu, Yonatan, & Wallace, Geoffrey PR. 2019. Violence, nonviolence, and the effects of international human rights law. *American Journal of Political Science*, **63**(2), 411–426.

Mercer, Jonathan. 1996. *Reputation and international politics*. Cornell University Press.

Morse, Julia C, & Pratt, Tyler. 2022. Strategies of Contestation: International Law, Domestic Audiences, and Image Management. *The Journal of Politics*, **84**(4), 000–000.

Petersen, Roger, & Zukerman, Sarah. 2010. Anger, violence, and political science. *Pages 561–581 of: International handbook of anger*. Springer.

Pomerantz, Anita. 1978. Attributions of responsibility: Blamings. *Sociology*, **12**, 115–121.

Powers, Ryan. 2022. Is Context Pretext? Domestic Audiences and the Situational Politics of Non-Compliance. *Working Paper*.

Raustiala, Kal. 1997. States, NGOs, and International Environmental Institutions. *International Studies Quarterly*, **41**, 719–740.

Risse, Thomas, & Sikkink, Kathryn. 1999. The socialization of international human rights norms into domestic practices: introduction. *In: The Power of Human Rights International Norms and Domestic Change.*

Schelling, Thomas. 1966. *Arms and Influence*. New Haven: Yale University Press.

Simmons, Beth A. 2009. *Mobilizing for Human Rights*. International Law in Domestic Politics. New York: Cambridge University Press.

Smith, Alastair. 1998. International crises and domestic politics. *American Political Science Review*, 623–638.

Strezhnev, Anton, Simmons, Beth A, & Kim, Matthew D. 2019. Rulers or Rules? International Law, Elite Cues and Public Opinion. *European Journal of International Law*, **30**(4), 1281–1302.

Tang, Shiping. 2005. Reputation, Cult of Reputation, and International Conflict. *Security Studies*, **14**(1), 34–62.

Tomz, Michael. 2007. *Reputation and International Cooperation: Sovereign Debt Across Three Centuries*. Princeton: Princeton University Press.

Trager, Robert F., & Vavreck, Lynn. 2011. The Political Costs of Crisis Bargaining: Presidential Rhetoric and the Role of Party. *American Journal of Political Science*, **55**(3), 526–545.

Von Stein, Jana. 2008. The International Law and Politics of Climate Change: Ratification of the United Nations Framework Convention and the Kyoto Protocol. *Journal of Conflict Resolution*, **52**(2), 243–268.

Wallace, Geoffrey PR. 2013. International law and public attitudes toward torture: An experimental study. *International Organization*, 105–140.

Weisiger, Alex, & Yarhi-Milo, Keren. 2015. Revisiting Reputation: How Past Actions Matter in International Politics. *International Organization*, **69**(2), 473–495.

Woo, Byungwon, & Murdie, Amanda. 2017. International Organizations and Naming and Shaming: Does the International Monetary Fund Care about the Human Rights Reputation of Its Client. *Political Studies*, **65**(4), 767–85.

Zvobgo, Kelebogile. 2019. Human rights versus national interests: Shifting US public attitudes on the international criminal court. *International Studies Quarterly*, **63**(4), 1065–1078.

Reputation Management and International Law Violations

Survey Instrument

The experiment consists of two hypothetical scenarios involving a violation of international law. Respondents are presented with the first scenario, then view one or more rhetorical messages from political actors, and provide their opinion on the violating government. They then proceed to the second scenario, featuring a different government and a different violation, and again provide their opinion on the violating government. We randomly assign the issue area and the identity of the foreign government (China or India) in the first scenario, and then assign respondents a different government and issue area in the second scenario.

The two scenarios depict violations of humanitarian and trade law. The humanitarian law scenario, below, describes alleged war crimes as reported by media sources:

> International law prohibits targeting civilians during military conflict. Governments are supposed to avoid attacking non-combatants, civilian infrastructure, and medical facilities.

> Imagine it is ten years in the future, and [VIOLATOR] is fighting a war in another country. Media sources report that a barrage of missiles that appeared to originate from [VIOLATOR]-controlled territory struck several civilian targets. The attack destroyed a large hospital and an apartment complex, killing hundreds of civilians.

The trade law scenario describes the imposition of tariffs:

> International law prohibits imposing trade restrictions under certain conditions. Governments that sign on to trade agreements are supposed to avoid placing unauthorized taxes on foreign imports.

> Imagine it is ten years in the future and a trade dispute emerges. Media sources report that [VIOLATOR] has imposed a tax on imported products, making foreign goods more expensive than similar products from India. Multinational firms claim that the tax is putting them at a competitive disadvantage.

Respondents in the control condition are further told that "The United States government said it was closely following the situation" and then shown the post-scenario questionnaire. Those in the treatment conditions receive a rhetorical justification from the violator government. The justification takes one of four forms:

- **Atonement**:
    - o War Scenario: The [VIOLATOR] government acknowledged the attack, calling it a terrible mistake and pledging to protect civilians going forward.

- o Trade Scenario: The [VIOLATOR] government acknowledged the new tax, calling it an unfortunate mistake and pledging to preserve free trade going forward.

- **Obfuscation**:
  - o War Scenario: The [VIOLATOR] government denied the attack, calling the story a lie designed to cast blame on [VIOLATOR].
  - o Trade Scenario: The [VIOLATOR] government denied the attack, calling the story a lie designed to cast blame on [VIOLATOR].

- **Attack**:
  - o War Scenario: The [VIOLATOR] government acknowledged the attack but argued that it should not be held responsible for civilian deaths since its enemies use civilians as human shields.
  - o Trade Scenario: The [VIOLATOR] government acknowledged the new tax but argued that free trade unfairly benefits the richest countries.

- **Obstruction**:
  - o War Scenario: The [VIOLATOR] government denied the attack, calling the story a lie designed to cast blame on [VIOLATOR]. The [VIOLATOR] government further argued that it should not be held responsible for civilian deaths since its enemies use civilians as human shields.
  - o Trade Scenario: The [VIOLATOR] government denied the attack, calling the story a lie designed to cast blame on [VIOLATOR]. The [VIOLATOR] government further argued that free trade unfairly benefits the richest countries.

Respondents are then randomly assigned a second treatment condition describing a response from the US government. The response may be neutral, supportive, or condemnatory of the violator government.

- **Neutral**: The United States government said it was closely following the situation.

- **Support**: The United States government expressed support for [VIOLATOR] and criticized efforts to politicize the current crisis.

- **Condemn**: The United States government condemned [VIOLATOR], calling its behavior a serious violation of international law.

After the US government response, respondents are randomly assigned to a final treatment condition describing a possible response from other political actors.

- **Control**: no further information

- **IO investigation**:
  - o War Scenario: The International Criminal Court underscored the importance of protecting civilians during war and opened an investigation into the allegation.

o Trade Scenario: The World Trade Organization underscored the importance of complying with trade commitments and opened an investigation into the allegation.

- **Domestic Actors** (only when US government condemns the violator government): However, several prominent media figures and politicians in the United States expressed

Respondents then complete a manipulation check to measure treatment uptake and then proceed to the following outcome questionnaire.

1. If the situation happened just as we described, would you agree or disagree with the following statements? (Select from "strongly disagree" to "strongly agree")
   o *Certainty about violation*
     - War: The [VIOLATOR] government definitely conducted the attack.
     - Trade: The [VIOLATOR] government definitely imposed the tax.
   o *Certainty about norm*
     - War: If the [VIOLATOR] government conducted the attack, its behavior was morally wrong.
     - Trade: If the [VIOLATOR] government imposed a tax, its behavior was morally wrong.

2. *Cooperation with Violator Government*
   Based on what you read, should the US government decrease cooperation with the [VIOLATOR], increase cooperation with the [VIOLATOR], or leave things about the same?

3. *Punishment of Violator Government*
   Suppose the US government was considering an additional response to the scenario that you just read and government officials were choosing from a variety of options. How should the US government respond to the actions of [VIOLATOR]? You may select multiple options.
   o Take no action
   o Rally other countries to support [VIOLATOR]
   o Issue a formal diplomatic censure of [VIOLATOR]
   o Impose economic sanctions on [VIOLATOR]
   o Work to expel the [VIOLATOR] from the UN Human Rights Council/World Trade Organization
   o Conduct drone attacks on select [VIOLATOR] military targets

4. *Costly Punishment*
   Suppose the US government announced a ban on imports from [VIOALTOR] in response to the events described above, and the move is expected to increase US gas prices. What is the **maximum** increase in gas prices that you would be willing to accept to punish [VIOLATOR]?
   o 0 (no increase)
   o An increase of 10 cents per gallon

- An increase of 20 cents per gallon
- An increase of 50 cents per gallon

5. *Delay*

Suppose that some countries have called for deferring a response to the events until more details come to light. Would you support the US government delaying retaliatory action until an international investigation can confirm more facts? (Select from "strongly support" to "strongly oppose")