Compliance and Contestation: Reputation Management in Multilateral Institutions

Julia C. Morse*and Tyler Pratt‡

August 23, 2019

Abstract

How do governments respond when they are accused of violating international commitments? Allegations of non-compliance spur a wide range of reactions, including outright denial, attacks on monitoring bodies, and public apologies. The decision to challenge, diminish, or recommit to international rules has clear implications for the evolution of international cooperation, yet we know little about why states adopt these postures. We argue that states respond to alleged violations in ways that minimize reputation costs. Drawing from scholarship on corporate public relations and criminal defense strategy, we introduce a two-stage theory of international reputation management. In the first stage, an accused state deflects responsibility by denying the allegation or by framing the violation as accidental. In the second stage, the state shapes how audiences assess the likelihood of future violations. Second-stage tactics include diminishing the seriousness of the violation or recommitting to international rules. States' reputation management strategies are shaped by their investment in the regime in the first stage and by the relative importance of domestic and international audiences in the second stage. We illustrate our theory through case studies of Sweden and Israel, two democratic states that have been accused of violating international rules prohibiting torture. Our findings suggest that reputations for compliance result from a contested political process that can sometimes undermine international cooperation.

---

*Assistant Professor, UCSB. Email: jcmorse@polsci.ucsb.edu

†Assistant Professor, Yale University. Email: tyler.pratt@yale.edu

‡We are grateful to Robert Keohane, Lauren Peritz, Christina Schneider, Arthur Stein, and participants of the 2019 UC Conference on International Cooperation for feedback on this project; and to Guo Chen, Renaud Concordet, Rhiana Ferguson, Jever Mariwala, Nora Patterson, and Dana Sinclair for their research assistance. Portions of this project were presented at the American Political Science Association conference in 2017 and the International Political Economy Society conference in 2018.

1 Introduction

The International Convention on Civil and Political Rights (ICCPR) and the Convention Against Torture (CAT) prohibit countries from extraditing individuals when they face a real risk of torture. Like many treaties, however, the record of state compliance is imperfect. In December 2001, Sweden extradited two asylum seekers to Egypt for alleged terrorist activities. The men were detained without trial upon arrival in Egypt and repeatedly tortured (Mohammed Alzery v. Sweden, 2006). In 2006, Kyrgyzstan extradited a group of asylum seekers to Uzbekistan. The men, who had previously been involved in anti-government protests, were subsequently tortured by the Uzbek government (International Federation For Human Rights, 2009).

The UN Human Rights Committee found both Sweden and Kyrgyzstan to be in violation of their obligations under the ICCPR and the CAT. However, the two countries responded to this accusation in entirely different ways. Sweden largely accepted the Human Rights Committee's findings, but argued that the principal blame for the violation lay with Egypt. Swedish officials claimed they did their best to comply with the obligation of non-refoulement, including seeking assurances from Egypt and visiting the prisoners while they were in Egyptian custody. Egypt's decision to mislead Swedish authorities about their intentions meant that Egypt was primarily responsible for the violation (Mohammed Alzery v. Sweden, 2006).

Kyrgyzstan opted for a different strategy. Rather than claiming ignorance of any possibility of torture, the government argued that it was compelled to extradite the men by its other international commitments. Both Kyrgyzstan and Uzbekistan are members of the Shanghai Cooperation Organization (SCO), a regional security organization designed to combat the "three evils" of terrorism, separatism, and extremism. Kyrgyzstan acknowledged that its decision to extradite may have violated the CAT but argued that it was upholding its

obligations under the SCO (International Federation For Human Rights, 2009, 23).

These contrasting strategies highlight the diverse ways that states respond to accusations of non-compliance. While nearly all states offer justification for violating international rules (Stein, 2000), governments challenge alleged violations in different ways. Why would Sweden, a strong upholder of international human rights obligations, provide such a legally dubious defense of its behavior? And why would Kyrgyzstan, a country known for corruption and weak rule of law, offer an explanation for its actions based on principled adherence to its international commitments?

This paper examines the strategies that states use to manage the aftermath of non-compliance. We argue that these strategies are used to mitigate the reputation costs associated with violating international rules. International relations scholars frequently cite the logic of reputation to explain why states comply with international obligations. A state with short-term interests in violating international rules may nonetheless comply because it benefits from having a positive reputation for cooperation, and it is loath to incur the reputation costs associated with non-compliance. According to this logic, institutions can encourage compliance by broadcasting information to relevant actors, who may choose to punish non-compliant states or deny them the benefits of future cooperation.

For reputation to sustain compliance, however, actors must clearly attribute responsibility to the violating state and update their beliefs about its future behavior. This is often a highly contested process, with non-compliant states offering frames that minimize their expected reputational damage. Media organizations, civil society activists, and political leaders may offer competing frames that attribute greater responsibility to the violating state. Whether non-compliance results in significant reputation costs depends on how audiences interpret available information about the violation. This context allows states to strategically frame information in ways that preserve their reputation – a process we call "international reputation management."

Consistent with much of the existing literature, we argue that international rule violations pose a threat to the reputations of non-compliant states. The severity of reputation costs, however, depends on a two-stage process whereby audiences first decide whether a state is responsible for a particular violation and then develop expectations about its future behavior. Throughout these stages, non-compliant states adopt reputation management strategies that minimize reputation costs.

To construct our theory of international reputation management, we draw inspiration from other contexts where actors accused of misconduct confront an audience that can impose significant costs. In the first stage ("responsibility attribution"), we draw on scholarship examining corporate responses to public relations crises (Coombs, 2007). Like corporations facing a crisis, states accused of violating international rules risk a loss of reputation. They have strong incentives to frame rule violations in ways that limit their perceived responsibility. We identify two common strategies used in this stage. States with little investment in a particular regime may claim to be victims of biased monitoring institutions. Other states may frame the violation as unintentional, citing external forces that thwarted a good faith effort to comply. Audiences weigh these narratives against competing frames that portray violations as intentional and fully preventable.

After attributing responsibility, audiences must decide how to update their expectations about the future behavior of a violating state. Even if actors determine that the state violated rules intentionally, they may decide that the violation is minor or uninformative about the state's behavior going forward. In the second stage ("mitigation"), the violating state highlights mitigating factors associated with the violation. To characterize this stage, we draw on legal strategies presented during the sentencing phase of criminal trials. A non-compliant state may diminish a violation by highlighting extenuating circumstances, or it may show remorse by acknowledging the seriousness of the violation and committing to improved behavior. We theorize that a state's choice of strategy in this stage will depend on

whether domestic audiences are likely to punish the violating government. A state is more likely to apologize and recommit to international rules when violations have become salient domestically.

We illustrate this theory and reputation management strategies through two case studies examining non-compliance with international rules regarding torture. Strong normative and legal prohibitions against torture mean that states violating international rules can expect to incur significant reputational damage and potentially high costs. Within this issue area, however, there is significant variation as to whether states are invested in the regime and whether they confront a mobilized domestic audience. We examine two democratic states – Sweden and Israel – that were accused of violating international rules regarding torture. We show how varying levels of investment in the regime led Sweden and Israel to adopt different first-stage strategies aimed at deflecting responsibility. We then demonstrate how the accused states undertook different strategies to mitigate reputational damage among international and domestic audiences.

Our findings have several implications for research on international cooperation. First, we add to a body of scholarship underscoring the limitations of reputation as a mechanism for sustaining cooperation. While existing work emphasizes the potential for states to hold multiple, distinct reputations (Downs & Jones, 2002), we explain how strategic behavior by non-compliant states can deflect reputational damage away from the state and toward other actors or institutions. States differ, however, in their optimal reputation management strategies. States that are highly invested in an international regime are constrained to an accident frame since challenging the objectivity or legitimacy of a monitor might undermine future benefits. Violations that fail to mobilize domestic audiences are unlikely to spark expressions of remorse or public recommitments to international rules. As a result, states will respond to alleged violations in systematically different ways depending on such constraints.

Second, our argument has implications for how reputation management influences the

evolution of cooperation in an issue area. As the preceding paragraphs illustrate, reputation costs are not an automatic and constant feature of international cooperation. Reputational effects are subject to significant contestation between a monitoring institution, a non-compliant state, and stakeholders that must attribute responsibility and determine the seriousness of the transgression. The resolution of this process may support or impede continued cooperation. If non-compliant states successfully evade responsibility for violations, they create loopholes that others are likely to exploit in the future. Diminishing the seriousness of violations also reduces future cooperation as states recognize that certain transgressions go unpunished. On the other hand, if non-compliant states are unsuccessful in evading responsibility, or if they seek to minimize reputation costs by acknowledging the violation and committing to improved behavior, then contestation may encourage more compliance in the future. Over the long-term, the evolution of cooperation will be influenced by the reputation management strategies adopted by non-compliant states.

2 Reputation Costs and International Cooperation

Scholars of international relations have long recognized the potential for reputational effects to sustain international cooperation. International politics occurs amidst great uncertainty about the capability and intentions of other actors. In the presence of such uncertainty, states and non-state actors make decisions about how to interact with others based on reputation – that is, a belief or judgment about an actor's trait, type, or behavioral tendency (Dafoe *et al.* , 2014, 365). A key insight from this definition is that beliefs, not facts, constitute the basis of reputation. Information about a state's behavior may influence its reputation, but ultimately, reputation depends on observers integrating new information in a way that ascribes a state's behavior to underlying traits or behavioral tendencies (Mercer, 1996, 6).

2.1 Reputation and International Institutions

Reputation is commonly cited as a powerful force for sustaining compliance. If states interact over multiple time periods, the reputation that they accrue in earlier periods will influence their success later on. As Keohane argues, "a good reputation makes it easier for a government to enter into advantageous agreements; tarnishing that reputation imposes costs by making agreements more difficult to reach" (1984, 105-106). Reputation is one way that states can help resolve information asymmetries that might otherwise preclude mutually beneficial cooperation among states. When a state is observed violating an international agreement, other actors update their beliefs about the state's likelihood of complying in the future. They can then deny the violator the benefits of future cooperation by screening them out of future agreements.

Scholars have found empirical support for the logic of reputational effects in a variety of contexts. Axelrod (1984) uses a series of simulated cooperative interactions to show how actors can maximize their own payoffs by conditioning on the past behavior of cooperative partners. Milgrom *et al.* (1990) demonstrate the importance of reputation, as well as the institutions that bolster it, in enabling trade among merchants in medieval Europe. Other scholarship highlights the role of reputation in facilitating compliance with sovereign loan agreements (Tomz, 2007), monetary rules (Simmons, 2000), trade agreements (Kono, 2007), and the laws of war (Morrow, 2007).

International institutions heighten reputational effects in at least two ways. First, institutions often legalize interstate agreements. By embedding their commitments in international law, states increase expectations of compliance and risk greater punishment from violating rules. Simmons & Hopkins (2005) argue that formal agreements like international treaties "enhance the reputational effects" of policy declarations by linking compliance to the broader principles of international law (623). Guzman (2008) similarly argues that states use legalization to calibrate the reputational consequences of non-compliance. Legal agreements allow

states to make a "maximal pledge of reputation," increasing the credibility of commitments by raising the expected costs of non-compliance (59).

Second, international institutions often include monitoring schemes that provide information on patterns of compliance. Institutionalized monitoring and assessment capabilities can range from voluntary state reporting to investigating and publicizing violations (Bradley & Kelley, 2008). The United Nations Security Council, for example, monitors the implementation of targeted sanctions against Al-Qaida, but relies on states to self-report on specific laws and procedures for freezing terrorist assets. In contrast, the Financial Action Task Force (FATF), a technocratic body that also focuses on combating terrorist financing, collects detailed information and lengthy assessment visits in every country to monitor compliance with FATF standards. Whereas the Security Council discusses only general trends in behavior, the FATF publishes a non-complier list that publicly identifies countries with deficient policies (Morse, Forthcoming). These monitoring systems facilitate the construction of reputations by broadcasting signals of compliance to a broader audience.

2.2 Reputation Costs

For reputation to sustain international cooperation, it is not enough that actors update their beliefs about a state based on its compliance with international institutions. These actors must also change their behavior after observing instances of non-compliance. A reputation cost occurs when states or non-state actors respond to new information by adjusting their behavior toward a non-compliant state, reducing its current or future welfare. This change in behavior can occur for two reasons. First and most directly, observers may adjust their behavior because an informational signal, such as a monitoring report from an international organization, provides information about a state's underlying traits or characteristics in a way that signals something about future behavior. Second, observers may adjust behavior due to information about how other observers are likely to view and behave toward the

non-compliant state.[1]

Multiple actors can impose costs on non-compliant states. States may punish other states directly, decreasing foreign aid, straining bilateral ties, or even abandoning alliance commitments. States may also work through third-party international organizations to impose reputation costs, including the expulsion of states from IOs. Alternatively, domestic actors in non-compliant states may impose costs on leaders, either through lobbying or directly at the polls. Finally, third parties such as non-governmental organizations, transnational activists, and markets may impose reputation costs, reallocating resources, spurning future cooperation, and publicly condemning the non-compliant behavior.

3 Managing Reputation Costs

The logic of reputation costs is attractive to scholars of cooperation for several reasons. It provides a coherent account of why states would make costly adjustments to comply with international institutions. In doing so, it offers a rebuttal to the realist critique that international institutions have no independent effect on state behavior (Mearsheimer, 1994). Finally, it provides a mechanism through which institutions that lack formal enforcement tools might still impose costs on violators. Simply through monitoring and the dissemination of information about non-compliance, institutions can trigger reputational effects.

We argue, however, that the strategic setting assumed by many proponents of reputation costs is incomplete. In particular, it precludes the opportunity for non-compliant states to contest allegations of non-compliance. Contestation over compliance is pervasive in international politics. States accused of violating international rules do not passively accept

[1]Dafoe *et al.* (2014, 374) term this distinction "first-order and "second-order" beliefs. First-order beliefs reflect an actor's observations about another actor's characteristics or behavioral tendencies. Second-order beliefs are an actor's beliefs about what a larger group of observers believes. While most work on reputation focuses on first-order beliefs, the decision to censure or punish rule violators often depends significantly on second-order beliefs. For example, permanent members of the United Nations Security Council may prefer to avoid vetoing a decision punishing an instance of non-compliance if their position is isolated.

a diminished reputation. Instead, they contextualize the allegations of non-compliance or deny them entirely, attacking the credibility of the monitoring institution. Contestation is consequential for understanding the ways in which reputation sustains cooperation because states can strategically manipulate the way that actors interpret observed violations.

To see how contestation affects reputation, consider the chain of events that must occur for states or non-state actors to impose reputation costs on non-compliant states. First, IOs or other actors must identify a violation of international rules. Once an IO determines a state is failing to fulfill its obligations, it must disseminate information about non-compliance to actors that can impose costs. Third, audiences must interpret this information in a way that attributes responsibility to the violating state.[2] Fourth, audiences must adjust their expectations about the non-compliant state's future behavior. Other governments may conclude that the non-compliant state has a broad tendency to violate international rules and exclude it from future cooperation. Alternatively, they may decide that the state's tendency to violate rules is constrained to a specific issue area, or they may make no inferences at all about the likelihood of future rule violations.[3] Finally, once beliefs are updated, actors adjust their behavior in ways that generate costs for the non-compliant state.

While some links in this chain have received significant scholarly attention,[4] the political process of interpreting violations deserves further scrutiny. This process, in which actors

[2]Mercer (1996) highlights the important role of interpretation in shaping reputational effects. For information to change a state's reputation for compliance, observers must decide the state's behavior is attributable to dispositional rather than situational characteristics. Despite the focus in the existing literature on dispositional traits, observers often consider situational factors when drawing inferences about non-compliance.

[3]Downs & Jones (2002) argue that the shift in beliefs arising from signals of non-compliance are smaller than commonly believed. States typically have an abundance of information about the past behavior of potential partners, limiting the importance of singular acts of noncompliance. States also cultivate and maintain multiple reputations with respect to different issue areas, casting doubt on whether noncompliance in one agreement confers a meaningful signal about future behavior in another.

[4]Dai (2007), for example, develops a typology of IO monitoring schemes and highlights how victims of non-compliance can facilitate monitoring. Recent scholarship has also focused on the mode of information transmission, arguing that information spread through ratings, rankings, indices, or blacklists may increase reputation costs because it facilitates comparisons across states and reduces uncertainty (Kelley & Simmons, 2015; Cooley & Snyder, 2015; Kelley, 2017; Morse, Forthcoming).

attribute responsibility for non-compliance and draw inferences about future behavior, provides ample space for strategic contestation. Both rationalist and constructivist accounts of belief formation suggest contestation can meaningfully affect how audiences interpret non-compliant behavior. Formal models of "cheap talk" communication indicate that the reputational effects of IO monitoring will depend on the perceived bias of the institution (Crawford & Sobel, 1982). States can contest allegations of non-compliance by challenging the impartiality of the institution, or by issuing their own "jamming" messages that conflict with the IO's account (Minozzi, 2011; Minozzi & Woon, 2016). Constructivist accounts focus on the contested social process that gives meaning to violations.[5] They argue that the consequences of rule violations depend on how the non-compliant behavior is interpreted by other states and non-state actors (Bull, 1977; Kratochwil & Ruggie, 1986).

We focus our analysis on the period after a state has been accused of non-compliant behavior.[6] We assume that states benefit from having a positive reputation for international cooperation (Keohane, 1984; Guzman, 2008), and that the alleged violation threatens to harm the reputation of the accused state. These conditions create incentives for the state to adopt strategies that minimize reputational damage.

We argue that international reputation management is usefully analyzed as a two-stage process. The stages correspond to two judgments that observers must make when non-compliant behavior is revealed: how to attribute responsibility for the violation, and how to update expectations about the state's future behavior. Throughout these "attribution" and "mitigation" stages, accused states attempt to shape the beliefs of observers in ways that limit reputational damage.

The initial attribution stage determines whether states will be deemed responsible for

[5]As Simmons (1998) summarizes, "the thrust of this literature is that perceived legitimacy of a legal rule or authority heightens the sense of obligation to bring behavior into compliance with the rule" (87).

[6]Recent work by Carnegie & Carson (2018) and Terman & Voeten (2018) examines strategic incentives in the prior period, when states decide whether to disclose violations discovered via monitoring.

rule violations. To analyze state strategy in this stage, we draw on scholarship examining organizational responses to corporate public relations crises. We identify two primary reputation management strategies that states use to evade responsibility. A state may adopt a victim posture – arguing that allegations are false, unfair, or malicious – or it may frame the violating behavior as accidental and related to factors beyond its control. Both strategies reduce reputational damage by shifting responsibility to other actors or circumstances. We theorize that the primary determinant of state strategy in the attribution stage is its level of investment in the regime. A victim frame directly challenges an IO's authority, while an accident frame implicitly acknowledges the legitimacy of the regime. For this reason, a violating state's reputation management strategy in the first stage is constrained by the state's level of investment in the *status quo* regime.

In most cases, the process of reputation management will extend into a second stage. Assuming the state is unable to completely evade responsibility for the violation, it shifts its focus to shaping expectations about future behavior. If international and domestic audiences expect the accused state to continue violating rules, they will be more likely to punish the state or exclude it from future cooperative endeavors. To characterize strategies in the mitigation stage, we draw inspiration from the presentation of mitigating information by defense attorneys during criminal sentencing. We theorize that non-compliant states, like convicted individuals seeking lenient sentences, will claim that violations are poor indicators of future behavior. As in the first stage, states have two basic strategies to mitigate reputational damage. They may diminish the violation by minimizing its seriousness or pointing to extenuating circumstances. Alternatively, a state may recommit to the regime by strengthening internal compliance procedures or compensating victims of the violation. While these strategies are not mutually exclusive, recommitment is significantly more costly, and therefore we anticipate that governments are more likely to adopt this strategy when international violations become sufficiently salient in domestic politics that their political

survival is threatened.

3.1 Attribution Stage: Victim or Accident

When an IO begins to investigate a purported violation or when a public report surfaces, the accused state faces a threat to its reputation that resembles a corporate public relations crisis. For both states and corporations, a public relations crisis that reveals bad behavior threatens to impose reputational damage and affect how stakeholders interact with the state or company in the future. To characterize state strategy in this stage, we draw on Situational Crisis Communication Theory (SCCT) (Coombs, 2007). SCCT is a framework for analyzing post-crisis communication strategies that maximize reputational protection for corporations. SCCT's core insight is that reputational damage depends on how stakeholders attribute responsibility for a crisis. A significant body of scholarship has shown that reputational threat correlates with crisis responsibility: as stakeholders attribute greater responsibility for a crisis to organizations, they view the organization less favorably (Coombs, 1998; Coombs & Holladay, 1996, 2002). For this reason, understanding how a crisis is framed is crucial for understanding how reputation costs emerge.

Insights from political science suggest that attribution of responsibility may depend in large part on how information is presented to relevant stakeholders. Scholarship on framing effects emphasizes that individuals form political opinions based in part on what considerations are made available, accessible, and relevant at the time of opinion formation (Chong & Druckman, 2007). When a description emphasizes a subset of potentially important considerations, it can lead recipients to focus on these factors and ignore others (Druckman, 2001, 230). Extending this line of research to televised political discourse, Mutz (2007) finds that even something as innocuous as close-up camera angles can make viewers more negative toward issue arguments that they are already prone to dislike. Individuals are more receptive to information that aligns with their preexisting views or when they feel positively toward

the source of information (Slothuus & de Vreese, 2010).

A state accused of non-compliance will have incentives to present information in a way that minimizes its responsibility for the violation. Building on Coombs (2007), we argue that accused states typically adopt either a victim frame or an accident frame.[7] These strategies are described in more detail below.

Attribution Strategy 1: Victim Frame

As originally articulated in SCCT, a victim frame assigns the lowest responsibility for the crisis to an organization. Examples of corporate crises amenable to victim frames are workplace violence, product tampering, natural disasters, or false rumors. In international relations, victim status is more likely to be politically negotiated. While IOs may sometimes produce false reports, more often states claim victim status by challenging the legitimacy of an IO, its rules, or its monitoring process. A non-compliant state argue that an IO is biased against the state and therefore not to be believed, or may suggest that the state itself is a victim of rules that favor the powerful over the weak. The goal of such an approach is to convince audiences to react to signals of non-compliance by drawing negative inferences about the regime, not the accused state.

The logic of a victim frame relies on the assumption that more legitimate institutions generate higher reputation costs. A significant body of scholarship links legitimacy with better compliance behavior. International legal scholars argue that the legitimacy of norms and rules affects the probability that states will comply with them (Franck, 1990).Rules created by IOs are more likely to be perceived as legitimate compared to rules created by states or non-state actors, because IOs embody a type of "rational-legal authority that

[7]Coombs (2007) also identifies a third frame ("preventable") that portrays a corporation as knowingly taking action that placed people at risk or violated a law. In the international context, we assume states accused of non-compliance have no incentive to portray themselves as fully responsible for violations of international law. This frame may be advanced by others, such as monitoring institutions and victims of non-compliance, and will compete with the accident and/or victim frames offered by the non-compliant state.

modernity views as particularly legitimate and good" (Barnett & Finnemore, 1999, 707). Buchanan & Keohane (2006) argue that IOs are perceived as legitimate when they are representative, knowledgeable, and accountable. States seeking to challenge the legitimacy of IO monitoring are likely to attack one or more of these three IO characteristics.

A state can employ a victim frame in one of two ways. First, it may challenge the content of monitoring directly, claiming the information itself is wrong. For such a strategy to succeed, the non-compliant state usually needs to attack the competence of the IO to explain why the monitoring report is incorrect. For example, in 2015 the World Anti-Doping Agency (WADA) accused the Russian government of perpetrating a systematic, state-sponsored doping program that delivered performance-enhancing drugs to the country's athletes. Russian officials strongly denied the accusations, calling them "baseless" and "fictional" (Holdsworth, 2015). They simultaneously attacked the integrity of institutions decision-making process, arguing that the WADA chairman overstepped his authority to single out Russia at the behest of Western states ((Stinson, 2016).

Alternatively, a state may target bias in an IO by highlighting its unrepresentative nature or its lack of accountability to stakeholders. In this approach, a state does not necessarily reject the content of monitoring itself, but rather criticizes the institution and the obligations it places on states. When Ecuador, for example, found itself on a "blacklist" of states due its weak oversight of money laundering and terrorist financing, the government critiqued the institution, not the information. Ecuadorian Foreign Minister Ricardo Patino stated that the Ecuadorian government did not recognize the right of the monitor (in this case, an intergovernmental body to which Ecuador did not belong) to "dictate policy" (The Economist, 2010).

Regardless of approach, when a violating state adopts a victim frame, it seeks to avoid potential reputation costs by shifting blame onto the regime. If successful, a victim frame damages the credibility or legitimacy of an IO and may affect future benefits from coop-

eration. As a result, states that gain from ongoing cooperation with a particular regime should be reluctant to deploy this approach. Instead, they will look for a way of deflecting responsibility without damaging the relevant IO.

Attribution Strategy 2: Accident Frame

SCCT also highlights a second frame where a crisis is described as an accident. In an accident frame, a corporation has some responsibility for a crisis but its responsibility is limited. For corporations, accident frames are common when a problem results from a technical error or when norms have gradually shifted and a company has not yet adjusted its policies. The key aspect of an accident frame is that the country accepts the facts of the crisis but underscores a lack of intent to violate. In international politics, a state deflecting responsibility through an accident frame might cite a lack of government capacity, highlight competing international standards, or shift blame to other actors. Each approach has its own strengths and weaknesses, and is appropriate for different contexts.

When a state deflects responsibility through capacity arguments, it signals that it has made a good faith effort to comply but failed because of its own internal constraints. In their canonical work on compliance, Chayes & Chayes (1993) highlight lack of capacity as a primary reason that states fail to comply with international law. The authors highlight how many international regulatory treaties require states to adopt new types of domestic legislation and develop provisions for enforcing such laws that may be quite challenging for less developed countries. Recent work by Borzel *et al.* (2010) suggests that a lack of administrative capacity may also affect compliance in more developed economies. This latter group of countries, however, may have more trouble employing such an approach. The goal of an accident frame is to try to deflect responsibility for a violation, and stakeholders are much less likely to accept such an argument when it comes from a developed country.

A second way that a country might implement an accident frame is by claiming that

the violating behavior is consistent with a competing or overlapping rule. In this approach, the accused state argues that its violation does not imply an intent to ignore its international commitments. Instead, the state encountered multiple rules that placed contradictory demands on its behavior. In a growing literature on "international regime complexity," scholars argue that many policy domains are now characterized by such an environment. As multilateral bodies have proliferated, specific issue areas are increasingly governed by a dense network of institutions that lack a formal mechanism for adjudicating between conflicting rules (Raustiala & Victor, 2004; Alter & Meunier, 2009).[8] Depending on the issue area, a state may be able to strategically exploit the presence of multiple rules to deflect responsibility for the non-compliant behavior onto the broader institutional environment.

A third way that a country can implement an accident frame is by shifting blame to another state or to a non-state actor. This framing strategy is similar to a *mens rea* defense in criminal law, whereby an accused individual cites a mistaken understanding of facts as a defense for illegal behavior. For example, an accused individual might claim that she unknowingly sold equities in a pyramid scheme or distributed a substance without realizing that it contained illegal drugs. The core claim is that the rule violation was unintentional and therefore should lead to a lesser penalty. In international relations, a state may shift blame by highlighting non-compliant actions by a second state. The rhetoric surrounding the escalating trade war between the US and China, where each country highlights how the other country has violated its international commitments, is one such example.

A country's ability to shift blame to another state may be linked to its preexisting reputation within an issue area. Just as a judge is likely to ignore a repeat offender arguing he did not intend to commit a crime, audiences are unlikely to accept an intent-based defense

[8]Existing scholarship highlights how regime complexity may enable states to challenge or escape existing obligations by choosing among alternative institutions (Helfer, 2004, 2009; Busch, 2007; Davis, 2009) or by creating new institutions (Morse & Keohane, 2014). We complement this work by demonstrating how the presence of overlapping rules creates opportunities for non-compliant states to deflect responsibility for rule violations.

from a government with a poor reputation in a particular issue area. European countries, for example, frequently defend their decisions to extradite individuals to countries with a risk of torture by citing country-specific improvements in human rights. Because they have strong reputations in this issue area, they can draw on specific evidence to support extradition decisions and give their governments plausible deniability if torture actually does occur. In contrast, when Russia and Morocco extradite individuals to countries with a risk of torture, they never employ this argument.[9]

The shared aspect of all of these accident frames is that a state acknowledges that a violation occurred but attempts to deflect responsibility for the behavior. Context-specific factors are likely to influence a state's choice of strategy. Some issue areas have a great degree of overlap while others do not. Some violations involve other actors who can be held responsible; other violations involve only a single government. Even the type and detail of monitoring might affect the initial strategy; if early reports on a violation describe intentional negligence, it will be much more difficult for a state to claim a violation is related to a lack of capacity. Regardless of approach, however, all accident frames tacitly accept the legitimacy of the regime, and this quality makes them quite distinct from victim frames.

Selecting a First-Stage Strategy

We assume that states are rational actors who consider both current and future gains from cooperation. In choosing a strategy to deflect responsibility for a rule violation, a state will opt for whatever approach offers the potential for reducing reputation costs in a way that minimizes undesirable regime effects. States that support an international regime in its current configuration have an incentive to avoid undermining its effectiveness; as a result, they are most likely constrained to the accident frame. Alternatively, states that do not

[9]These statements are based on the authors' current data-gathering exercise on cases brought under Article 22 of the Convention Against Torture.

favor the regime *status quo* would prefer a strategy that undermines the current institutional order. Such states should actively seek to avoid an accident frame as this approach tacitly acknowledges the legitimacy of the regime.

- *Attribution Stage Hypothesis: When a state is accused of violating international law, its level of investment in the relevant regime will determine how the state attempts to attribute responsibility. A state with low investment in the regime will rely on a victim frame, while a state with medium-to-high investment in the regime will adopt an accident frame.*

3.2 Mitigation Stage: Diminishment or Recommitment

After responsibility for the violation has been attributed, stakeholders must decide how the violation will inform their expectations about the future behavior of the non-compliant state. The link between responsibility attribution and subsequent punishment of the offender is not automatic. For the non-compliant state to incur a reputation cost, audiences must come to believe that the violation reflects an underlying tendency to breach commitments. This process presents a second opportunity for contestation between the violating state, monitoring institution, and relevant audiences. We call this the mitigation stage of the reputation management process.

The mitigation stage determines whether audiences view the violation as a serious breach that is likely to be repeated or a minor infraction limited to a specific set of circumstances. This stage is most salient when the initial stage of responsibility attribution has determined that a state is at least partially culpable for a breach of international rules. As before, the goal of the non-compliant government is to secure a set of beliefs that minimize reputational damage. Other stakeholders, including the monitoring institution that initially disclosed the violation, often propose alternative frames that underscore the serious nature of the

non-compliant act.

To characterize strategies used in the mitigation stage, we draw on scholarship examining the presentation of mitigating information by defense attorneys during criminal sentencing. This phase of a criminal trial closely mirrors the mitigation stage of reputation costs we seek to unpack: after an offending actor has been found culpable for a violation, an audience must determine the appropriate level of punishment. That determination depends on whether the offender is viewed as a continued threat to society.

In capital cases, the presentation of mitigating information is standard practice. Many defense teams hire "mitigation specialists" to assist with the development and presentation of mitigating factors. Like a non-compliant state attempting to preserve its reputation, the defense team presents "a persuasive description of the forces that shape human behavior and that, in some instances, explain behavior that violates law" (Gohara, 2013, 4). In the criminal sentencing process, these are countered by aggravating factors typically presented by the prosecution. Judges and juries weigh both mitigating and aggravating factors to determine whether the convicted individual is likely to repeat the violation.

Criminal defense attorneys can select from a wide variety of mitigating factors. A common approach is the discussion of extenuating circumstances that contributed to the violation, such as poverty, trauma, or intellectual disability (Gohara, 2013). Defense teams may also present evidence of general good character positive attributes of the offender that are weighed against the specific violation under consideration (Steiker & Steiker, 1992). Other strategies include stoking residual doubt about the defendant's guilt (Garvey, 1998), the elicitation of mercy (Trahan, 2011), the expression of remorse, or minimization of the underlying crime.

Not all mitigation strategies in capital trails are relevant in the context of international legal violations. We identify two strategies that are linked to the logic of reputation and are often adopted by states accused of violating international commitments. These strategies, diminishment and recommitment, are described below.

Mitigation Strategy 1: Diminishment

In a diminishment strategy, a non-compliant state portrays a violation as a "special case" that is uninformative about its subsequent behavior. The goal of this strategy is to diminish the importance of a particular violation in the minds of observers as they predict whether non-compliance will recur in the future. The strategy involves highlighting unique circumstances of the violation that set it apart from a state's typical pattern of compliance. If successful, it reduces reputation costs by limiting the scope of reputational inferences drawn by audiences.

There are several manifestations of the diminishment strategy. Some focus on the specific rule that has been violated. States who are deemed responsible for non-compliance often portray the breach as producing little or no harm to other states or individuals. The message implicit in this claim is that the violated rule is minor and unworthy of significant censure. Future violations will be limited to similarly trivial matters, and the state will continue to honor its important commitments.

Other examples of diminishment highlight extenuating circumstances surrounding the violation. Here, non-compliant states portray the violation as an unusual breach that stemmed from particularly difficult circumstances. Domestic or international security threats are commonly cited as extenuating circumstances. States will cite these threats, often by declaring a state of emergency, to justify the suspension of domestic freedoms or the violation of human rights commitments. These diminishment claims attempt to minimize reputational damage by arguing that violations reflect a unique and temporary set of circumstances, rather than the underlying disposition of the state. After the extreme circumstances have passed, the state will return to its normal pattern of compliance with international commitments.

Mitigation Strategy 2: Recommitment

The second mitigation strategy is apology and recommitment. In this approach, a govern-

ment accepts its own wrong doing and indicates that it is strongly committed to avoiding such behavior in the future. A recommitment strategy is similar to expressions of remorse during criminal sentencing: it induces leniency by demonstrating that the violator acknowledges guilt and has internalized the norms underlying the relevant rule. If successful, it convinces observers that a state is reformed and therefore less likely to violate rules in the future.

Recommitment strategies confront an inherent credibility problem and are therefore most effective when accompanied by costly reforms. A state might adopt new domestic procedures that limit the government's ability to commit future violations. For example, domestic leaders might announce an independent task force to investigate the violation and hold relevant perpetrators accountable. A leader might implement regulatory changes or reorganize the bureaucracy as part of a larger effort to ensure future violations do not occur. Recommitment could include public signals like press conferences or public hearings, and more private signals such as meeting with families of victims (where relevant) or paying compensation to those who have suffered as a result of government wrong doing. Finally, a government might opt to demonstrate recommitment by joining additional international treaties or protocols that allow for greater international oversight and monitoring.

In some cases, recommitment may coincide with governmental change, as a way of reassuring the public that the new government is different from the old. When Chile transitioned from a dictatorship to an elected democracy, for example, President Patricio Awlwin established a truth and reconciliation body known as "the Rettig Commission." Composed of eight commissioners, the Rettig Commission was responsible for documenting human rights abuses that resulted in death or disappearances under the Pinochet regime.[10] In the United States, newly-elected President Barack Obama also employed a recommitment strategy to

[10]For more on the Rettig Commission, see "Truth Commission: Chile 90," *United States Institute of Peace*, retrieved from: `https://www.usip.org/publications/1990/05/truth-commission-chile-90`.

improve the US government's human rights reputation. In addition to releasing memos detailing the use of torture by the Bush administration, President Obama announced a ban on waterboarding and other harsh interrogation techniques that were widely considered to be torture.

Selecting a Second Stage Strategy

Because the recommitment strategy often entails costly action and constraints on future behavior, we argue that states prefer to use the diminishment strategy in the second stage. States will escalate to the recommitment strategy only when forced to do so in order to defuse a potent reputational threat. In particular, we argue that states are most likely to adopt a recommitment strategy when international rule violations produce serious domestic political costs, such as threatening the political survival of the leader or ruling party.[11] This occurs when rule violations become sufficiently salient that domestic actors agitate for political change. If a government is primarily concerned about an international audience, we expect them to stick to the diminishment strategy.

- *Mitigation Stage Hypothesis: When a state is accused of violating international law, its primary audience will determine how the state attempts to mitigate the consequences of the violation. A state confronting an engaged domestic audience that is opposed to the violation will adopt a recommitment strategy, while a state confronting an international audience will adopt a diminishment strategy.*

4 Reputation Management in the Torture Regime

In this section, we illustrate our theory of reputation management through case studies of Sweden and Israel in the aftermath of alleged violations of international law prohibiting

[11]Not all violations become salient among domestic audiences. Recent work demonstrates that domestic audiences often respond to non-compliance in unexpected ways (Chaudoin, 2016; Lupu & Wallace, 2018).

torture. In each case, several IOs declared the state to be in violation of international human rights law. As both crises developed, governments responded by attempting to alleviate the reputation costs associated with the violation, first by rejecting responsibility for the crisis and then by shaping expectations about compliance in the future. The strategies selected in each stage, however, differed across the two cases. The Swedish government advanced an accidental frame to shift responsibility elsewhere and later publicly recommitted to international rules. The Israeli government used a victim frame with one human rights body and an accident frame with a second IO, and in the latter context, also attempted to diminish the importance of its violations by citing security threats.

Our analysis of Sweden focuses on the extradition case discussed in the introduction. We describe how in ongoing communications with the UN Human Rights Committee and the Committee Against Torture, Sweden repeatedly emphasized its efforts to follow international rules and laid blame on Egypt for misleading Swedish officials. Later, Sweden issued a public apology, compensated victims, and recommitted to international rules. We also discuss the case of Israel, a democratic country that has come under repeated criticism for human rights violations in the Palestinian territories. We describe how two UN bodies – the Human Rights Council and the Committee Against Torture – deemed Israeli detainment policies and violence against Palestinian protesters as not in compliance with international law, but Israel engaged quite differently with each IO. While Israel repeatedly criticized the Human Rights Council as biased and illegitimate, Israel argued in a brief to the Committee Against Torture that the laws of war rather than human rights law apply to the Palestinian territories. It also attempted to diminish the importance of the violation by citing the security context that led to the alleged violations.

4.1 International Rules Regarding Torture

International rules prohibiting the use of torture originate in numerous formal and informal institutions. Legal and normative sources for the prohibition include multilateral agreements, customary international law, and non-binding documents like the Universal Declaration of Human Rights. Because torture may occur in either wartime or peace, rules in this issue area draw on both international humanitarian law and international human rights law. Since 1949, the Geneva Conventions have prohibited the use of torture against civilians, noncombatants, and capture soldiers during war.[12] Extending the ban to times of peace, however, occurred more slowly. The non-binding Universal Declaration of Human Rights established a strong norm against torture in any circumstance as early as 1948.[13] Subsequent treaties like the European Convention on Human Rights and the International Covenant on Civil and Political Rights codified this norm into law.[14]

International rules against torture expanded significantly with the UN General Assembly's adoption of the Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (CAT) in 1984. The CAT includes a definition of torture (Article 1),[15] a requirement that states ensure torture does not occur in any territory under their

[12]Article 3 of the First Geneva Convention forbids "mutilation, cruel treatment and torture" of these protected persons, and includes a broader prohibition on "humiliating and degrading treatment" (Gen, 1949b). Article 31 of the Fourth Convention explicitly bans the use of "physical or moral coercionagainst protected persons, in particular to obtain information from them or from third parties" (Gen, 1949a).

[13]The Declaration states that "no one shall be subjected to torture or to cruel, inhuman or degrading treatment or punishment" (UNH, 1948)

[14]Article 3 of the European Convention on Human Rights stipulates that "no one shall be subjected to torture or to inhuman or degrading treatment or punishment" (ECH, 1950). Article 7 of the ICCPR explicitly forbids the use of torture and cruel, inhuman, or degrading punishment.

[15]The CAT defines torture as "any act by which severe pain or suffering, whether physical or mental, is intentionally inflicted on a person for such purposes as obtaining from him or a third person information or a confession, punishing him for an act he or a third person has committed or is suspected of having committed, or intimidating or coercing him or a third person, or for any reason based on discrimination of any kind, when such pain or suffering is inflicted by or at the instigation of or with the consent or acquiescence of a public official or other person acting in an official capacity. It does not include pain or suffering arising only from, inherent in or incidental to lawful sanctions."

jurisdiction (Article 2), and a prohibition on deporting, extraditing, or refouling individuals to countries where they are likely to face torture (Article 3). Subsequent human rights treaties and regional agreements reinforced these rules and reiterated their applicability in specific circumstances. The Convention on the Rights of the Child, for example, states that "no child shall be subjected to torture or other cruel, inhuman or degrading treatment of punishment" (Article 37 a). The prohibition against torture is also considered to have the highest standing under customary international law (*jus cogens*).[16]

UN agencies are the primary multilateral monitors of compliance with the prohibition against torture, and these monitors include both technical and intergovernmental bodies. The UN Committee Against Torture, which is composed of ten independent experts elected by states to renewable four-year terms, monitors compliance with the CAT. Under the CAT, signatories are required to submit compliance reports every four years to this committee. The committee's experts examine country reports and consult with NGOs and other interested parties. Following each review, the experts issue a set of "Concluding Observations" that identifies specific instances of non-compliance and recommends changes in state policy.[17] A similar panel of independent experts (the UN Human Rights Committee) monitors compliance with the International Covenant on Civil and Political Rights. Both committees work closely with the UN High Commissioner for Human Rights, who is charged with promoting rights guaranteed in the Universal Declaration of Human Rights, and the UN Special Rapporteur on Torture.

In addition to these expert bodies, the UN Human Rights Council (formerly the UN Commission on Human Rights) is an intergovernmental body tasked with reviewing and protecting human rights in all UN member states, to include the prohibition against torture.

[16]Human Rights Watch. The Legal Prohibition Against Torture. Available at: https://www.hrw.org/news/2003/03/11/legal-prohibition-against-torture.

[17]For an example, see the Committees recent Concluding Observations regarding Pakistan, released in June 2017 (Committee Against Torture, 2017).

The Council is composed of 47 member states, which are elected to the Council by the UN General Assembly. Because of this institutional design, states exercise direct control over the Council agenda and the identification of human rights violations.

Some states are further subject to supra-national courts that have been empowered to interpret international rules regarding torture. For example, the 1998 Rome Statute lists torture as a crime against humanity, granting the International Criminal Court jurisdiction over acts of torture (ICC, 1989). The European Court of Human Rights reviews allegations that Council of Europe member states violated the European Convention on Human Rights (ECHR), including its rules prohibiting torture.[18] The International Court of Justice (ICJ) may also hear cases related to the use of torture, provided that a disagreement arises between states that agree to submit to the ruling of the court. In 2012, for example, the ICJ ruled in a dispute between Belgium and Senegal that the CAT obligates states to prosecute alleged perpetrators of torture or extradite them to another country for prosecution.[19]

4.2 Sweden and the War on Terror

Our first case study highlights Sweden's reputation management strategy vis-a-vis allegations that it violated the Convention Against Torture. Sweden has historically portrayed itself as a champion of human rights, and its reputation reflects a pattern of advocacy and respect for human rights issues. In a 2017 global survey, respondents ranked Sweden first among all countries in concern for human rights.[20] Both the Swedish government and Swedish society are highly invested in UN human rights organizations, and therefore when Sweden was accused of violating the Convention Against Torture, the government attempted to frame

[18]For example, the European Court of Human Rights ruled in the case of Shishkin vs. Russia (2011) that the Russian government violated the ECHRs prohibition against torture.

[19]See https://www.asil.org/insights/volume/16/issue/29/belgium-v-senegal-international-court-justice-affirms-obligation.

[20]https://www.usnews.com/news/best-countries/articles/2017-09-18/the-10-countries-that-care-the-most-about-human-rights-according-to-perception.

its violation as accidental. In the second stage, Sweden engaged in a recommitment strategy to mitigate the domestic reputation costs associated with the violation.

Sweden's violation of the Convention Against Torture occurred in the context of the "war on terror." In 2001, the Swedish government denied the asylum claims of two Egyptian nationals, Mohammed Alzery and Ahmed Agiza, and expelled them from the country. Egypt alleged that both men were linked to terrorist organizations that sought to overthrow the Egyptian government, a charge they denied. Upon arrival in Egypt, the men were imprisoned and repeatedly tortured.

The extradition of Alzery and Agiza was part of the "extraordinary rendition" operation carried out by US intelligence services after the September 11 terrorist attacks. In December 2001, Swedish officials handed over the men to CIA operatives for transfer from Stockholm to Cairo. Cognizant of the prohibition against returning individuals to countries where they face a risk of torture, Sweden sought and received a commitment from Egyptian authorities that the men would be treated humanely. These assurances were quickly violated. The men were imprisoned, tortured, and held in inhumane conditions for several years. Alzery was released without charges in October 2003, and Agiza served a ten-year sentence in Egyptian prison.

In the years following the extradition, many international organizations condemned the behavior of the Swedish government. In 2005, the Committee Against Torture determined that the extradition violated Sweden's commitments under the CAT and ICCPR (Mohammed Alzery v. Sweden, 2006). The Committee Against Torture concluded that the expulsion was a violation due to the likelihood that Egyptian authorities would torture the men. The Human Rights Committee reached a similar conclusion. Both bodies specifically noted that diplomatic assurances were an inadequate safeguard against torture (Mohammed Alzery v. Sweden, 2006; Agiza v. Sweden, 2005). These institutions were joined by several European monitoring bodies, as Swedish behavior became a focal point for European oppo-

sition to the American rendition program. The Parliamentary Assembly of the Council of Europe and the European Parliament both severely criticized the violation (Human Rights Watch, 2006).

The Swedish government recognized the reputational threat posed by the extradition[21] and attempted first to limit its own responsibility for the violation through an accident frame. The government characterized its behavior as a good faith effort to comply with international rules. It did not challenge the facts as presented by monitoring institutions, but insisted that the primary responsibility for the violation laid with Egypt, which misrepresented its intentions. Swedish officials also emphasized their efforts to ensure compliance, including sending officials to monitor the treatment of the men in Egyptian prison (Mohammed Alzery v. Sweden, 2006). They made it clear that the decision to extradite would not have been made without credible assurances from the Egyptian government that the men would be treated humanely.[22] Sweden's framing of the violation was clear: its participation in the affair, while unfortunate, was unintentional.[23]

Sweden's use of an accident frame is consistent with its high level of investment in the regime. Human rights promotion is central to Swedish identity. A recent UN review noted "the broad consensus within Swedish society and across the political spectrum on the impera-

[21]In 2005, for example, a Swedish Parliamentarian acknowledged that the violation was "damaging for Sweden in the short term, particularly in the eyes of international organizations." `https://www.thelocal.se/20050415/1285`.

[22]The Swedish government made this argument directly to the Committe Against Torture: "After careful consideration of the option to obtain assurances from the Egyptian authorities with respect to future treatment, the [Swedish] government concluded it was both possible and meaningful to inquire whether guarantees could be obtained to the effect that the complainant and his family would be treated in accordance with international law upon return to Egypt. Without such guarantees, return to Egypt would not be an alternative. On 13 December 2002, requisite guarantees were provided" (Agiza v. Sweden, 2005, 4.12).

[23]Berndt Ekstrom, a member of the foreign affairs committee in the Swedish Parliament, emphasized that the government "did what it thought was correct at the time." `https://www.thelocal.se/20050415/1285` .

tive of practising human rights-based international solidarity."[24] As a result, the government was unlikely to portray itself as a victim of a biased regime when managing its reputation in the aftermath of the Alzery and Agiza extraditions. Instead, its initial response was to limit its responsibility for the violation by portraying non-compliant behavior as unintentional.

The government's first-stage strategy was not entirely successful. While it may have ameliorated some reputational damage, it could not prevent judgments by international monitoring bodies that held Sweden responsible for violating its commitments (Mohammed Alzery v. Sweden, 2006; Agiza v. Sweden, 2005). As backlash against U.S. counteterrorism measures grew in Europe, the government was forced to confront international and domestic actors who deemed the country complicit in a serious violation of international law. It responded by adopting many features of the recommitment strategy.

In 2007, Sweden began taking public steps to make amends for the violation. The government formally overturned the decision to forcibly repatriate both men. In 2008, the Swedish Ministry of Justice reached an agreement to compensate both Alzery and Agiza with a sum of three million kronor (approximately $400,000).[25] As part of the agreemeent, the Chancellor of Justice acknowledged the state's responsibility for the violation, admitting that the men were exradited "under circumstances that were not acceptable."[26] In 2012, Agiza was granted permanent residency in Sweden.[27]

Sweden's recommitment strategy was driven in large part by the salience of this issue for domestic audiences. In May 2004, a Swedish television station aired a news special, "Kalla Fakta" (The Broken Promise) examining the extradition. The program revealed,

[24] "Report of the Independent Expert on Human Rights and International Solidarity," https://documents-dds-ny.un.org/doc/UNDOC/GEN/G19/108/15/PDF/G1910815.pdf.

[25] https://www.jpost.com/International/Sweden-pays-502000-in-compensation-to-exonerated-terror-suspect https://www.thelocal.se/20080919/14456.

[26] https://www.jpost.com/International/Sweden-pays-502000-in-compensation-to-exonerated-terror-suspect

[27] https://www.thelocal.se/20120704/41832.

in painful detail, how Swedish police physically assaulted the men before handing them to U.S. operatives, who blindfolded, hooded, and drugged them prior to transfer to Cairo.[28] Domestic opposition to the violation grew so severe that a 2012 survey named the extradition affair the country's worst political scandal in the past 20 years.[29] Faced with this level of domestic backlash, the Swedish government had little choice but to apologize for the violation and signal its recommitment to liberal human rights practices.

4.3 Israeli Detainment and Interrogation in the Palestinian Territories

Our second case study examines how Israel has responded to alleged violations of human rights law in the Palestinian territories. Israel has a complicated history with human rights monitoring bodies and provides an opportunity to showcase how a state's level of investment in a regime influences its reputation management strategy. We highlight how Israel engages quite differently with the UN Human Rights Council (UNHRC), an institution with which it has low investment, compared to the UN Committee Against Torture. Although it contests accusations of non-compliance from both bodies, the government adopts different strategies in each case. While Israel challenges UNHRC allegations with a victim frame that emphasizes institutional bias, it responds to Committee Against Torture allegations with an accident frame that emphasizes a conflict between the laws of war and human rights law. Israel's engagement with the Committee Against Torture also shows its second-stage reputation management strategy of diminishment, where it highlights the extenuating circumstances that might lead to any possible violations.

More than any other democratic state, Israel has complex and varied relationships with UN human rights bodies. Israel has signed onto many international human rights treaties,

[28] `https://www.hrw.org/legacy/backgrounder/eca/eu0107/7.htm`.

[29] `https://www.thelocal.se/20120704/41832`.

including the Convention Against Torture, the Convention on the Rights of the Child, and the Convention on the Elimination of Racial Discrimination. It participates in the monitoring exercises for these treaties and is described by Freedom House as a country with strong and independent institutions that guarantee politics rights and civil liberties for most of the population.[30] This positive image exists alongside a second image of Israel as an occupying state that has committed human rights abuses in the course of achieving national security objectives. International bodies like the UN General Assembly and the UNHRC frequently criticize Israel for violating human rights law in the Palestinian territories.[31] As a result, the Israeli government views parts of the UN as biased and antagonistic, reducing its perceived investment in the regime.[32] This is particularly true of the UNHRC.

Over the last five years, Israel has been repeatedly criticized by UN human rights bodies for human rights violations in the Palestinian territories. In response to an increase in violence in late 2015, Israeli authorities began to detain a large number of Palestinian minors for acts like throwing stones. The Israeli government also adopted new legislation with up to a 20-year prison sentence for throwing stones and custodial sentences for children as young as 12.[33] The Committee Against Torture's 2016 review of Israel highlighted problems with torture and ill-treatment of detainees, particularly those held as unlawful combatants and minors accused of security-related offenses (Committee Against Torture, 2016). The report

[30] `https://freedomhouse.org/report/freedom-world/2019/israel`

[31] While the Israeli government has varied interactions with different UN human rights bodies, Israel has had a particularly negative relationship with the UNHRC. Since its establishment in 2006, the US government and the Israeli government have criticized it for disproportionately targeting Israeli. Israel is the only permanent standing item on the Council's agenda and while the Council frequently scrutinizes Israeli actions, it fails to take actions on many other human rights crises due to political protection by Council members.

[32] See, for example, Israel and the U.N., AIPAC, retrieved from: `https://www.aipac.org/learn/issues/issue-display/israel-and-the-un`.

[33] "Concern over conditions and violence against Palestinian children in detention, *UN Office for the Coordination of Humanitarian Affairs*, 9 February 2016, retrieved from: `https://www.ochaopt.org/content/concern-over-conditions-and-violence-against-palestinian-children-detention`.

also expressed concern about allegations of "excessive use of force, including lethal force, by security forces, mostly against Palestinians...in the context of demonstrations" (Committee Against Torture, 2016, 7-8). The UNHRC has also criticized Israel for similar violations, highlighting concern about the excessive use of force and advanced interrogation techniques.

While both the Committee Against Torture and the UNHRC have accused Israel of violating human rights law and the Convention Against Torture, Israel has responded quite differently to allegations from each institution. With the UNHRC, the Israeli government has adopted exclusively a victim frame. In 2013, Israel became the first country to opt not to appear before the UNHRC for its mandatory universal periodic review.[34] Since that time, the Council has continued to criticize Israel for its actions in the West Bank and Gaza, and Israel has responded to such challenges by framing itself as a victim of a biased and illegitimate international body. In its statement to the UNHRC in June 2016, the Israeli government criticized the Council for its "politicized debates, biased resolutions, preposterous reports, discriminatory conduct and unfounded accusations" and claimed that the Council needed "a moral compass.[35] Most recently in February 2019, Israel's ambassador to the UN accused the Council of being "blinded by hatred of Israel and the Israeli Defense Forces." Israel's Foreign Minister issued a similarly dismissive statement, "The Human Rights Council's Theatre of the Absurd has once again produced a report that is hostile, mendacious and biased against Israel."[36]

In contrast to its strategy toward the UNHRC, the Israeli government has attempted to

[34] "Israel boycotts UN rights council in unprecedented move, 29 January 2013, *BBC*, retrieved from: `https://www.bbc.com/news/world-middle-east-21249431`.

[35] "Israel statement at 32nd session of UN Human Rights Council, Israel Ministry of Foreign Affairs, 14 June 2016, retrieved from: `https://mfa.gov.il/MFA/InternatlOrgs/Pages/Israel-statement-at-32nd-session-of-UN-Human-Rights-Council-14-June-20160615-1504.aspx`.

[36] Both quotes are drawn from: Ben Evansky, "Israel condemns UN Report on Gaza protests, says Human Rights Council blinded by hatred, 28 February 2019, *Fox News*, retrieved from: `https://www.foxnews.com/world/israel-condemns-un-report-on-gaza-protests-says-human-rights-council-blinded-by-hatred`.

deflect reputational fallout with the Committee Against Torture by employing an accident frame and a diminishment strategy. During the Committee Against Torture's 2016 review of Israel,[37] the Israeli Permanent Representative to the UN in Geneva expressed respect and appreciation for the work of this Committee and of other UN treaty monitoring bodies" before noting that Israel's Supreme Court is highly regarded by scholars of international law for its "balancing of security and the protection of individuals."[38] Throughout the Israeli government's testimony before the Committee, government officials noted their respect for the Committee's mission and for the Convention's principles but argued that the Convention did not apply to the Palestinian territories. Instead, Israel opted to limit responsibility for any violations by citing an alternative set of legal standards, namely the law of armed conflict. While Israel acknowledged a link between human rights law and the law of armed conflict, it argued in its submission to the committee that "these two systems-of-law...remain distinct and apply in different circumstances."[39]

Israel also engaged in a second-stage reputation management strategy during its interactions with the Committee Against Torture in 2016. In his opening remarks to the Committee, the Israeli Permanent Representative highlighted that Israel faced a challenging security situation, citing a new wave of terrorist attacks and violence since September 2015.[40] In

[37]Israel is a signatory of the Convention Against Torture, which mandates a periodic review of all member states every four years. The Committee reviewed the Israeli governments activities and compliance with the Convention most recently in 2016.

[38]Quoted from H.E. Eviatar Manor, Permanent Representative of Israel to the United Nations, Geneva, Statement Before the Committee Against Torture and other Cruel, Inhuman, or Degrading Treatment or Punishment, 3-4 May 2016, retrieved from: https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/SessionDetails1.aspx?SessionID=1011&Lang=en.

[39]Fifth periodic reports of States parties due in 2013: Israel, CAT/C/ISR/5/, 16 February 2015, retrieved from: https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/SessionDetails1.aspx?SessionID=1011&Lang=en.

[40]Quoted from H.E. Eviatar Manor, Permanent Representative of Israel to the United Nations, Geneva, Statement Before the Committee Against Torture and other Cruel, Inhuman, or Degrading Treatment or Punishment," 3-4 May 2016, retrieved from: https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/SessionDetails1.aspx?SessionID=1011&Lang=en.

testimony to the Committee, the Israeli government also attempted to diminish any ongoing violations, such as the detention of minors, by highlighting the steps that it had taken to implement the Committee's 2009 recommendations (from the previous periodic review). Israel's Deputy Attorney General of International Law highlighted how many reforms, including a bill making confessions procured under torture inadmissible, the establishment of a military juvenile justice system in the West Bank, and the prosecution of ideologically-motivated offenses against Palestinians in the West Bank, reflected Israel's attempts to implement the Committee's previous concluding observations against Israel. By highlighting the country's larger commitment to the Committee's work and the Convention Against Torture, Israel was attempting to signal that any violations reflected extenuating circumstances rather than a lack of underlying commitment. The country's choice of a diminishment strategy is consistent with the lack of a significant domestic political threat stemming from the violation.

5 Conclusion

Reputation is a core causal mechanism for explaining how international institutions promote cooperate between states. But if reputational concerns drive states to comply with international rules, then states are likely to seek out strategies to impede this process. Existing scholarship has highlighted how an IO's ability to identify and publicize non-compliant behavior can damage the reputations of non-compliant states. But there is an unexplored step in this causal relationship: states and non-state actors must interpret the IO's signal and decide whether to alter their behavior toward the non-compliant state. We argue that non-compliant states have an opportunity to influence this process, changing how audiences understand the violation.

We have laid out a two-stage theory of reputation management whereby a state first seeks to limit how others attribute responsibility for a violation and then attempts to signal

its future intent to comply. In the attribution stage, a state may adopt either a victim frame or an accident frame to limit its responsibility for a violation. In a victim frame, a state rejects the violation entirely, framing the IO or its monitoring as biased or illegitimate. In an accident frame, a state accepts that the violation occurs but deflects responsibility by citing a lack of capacity, competing international rules, or bad behavior from another state. When choosing between these strategies, a state is constrained by its level of investment in the regime. A state that anticipates few future benefits from cooperating with an IO will frame its behavior as a victim; it will want to avoid an accident frame since it tacitly acknowledges the legitimacy of the regime. In contrast, a state with high regime investment will only opt for an accident frame since a victim frame could undermine the IO.

Most states will also opt for a second-stage reputation management strategy. In the mitigation stage, a violating state may attempt to signal its future intent to comply by either diminishing the violation or by recommitting to the regime. In a diminishment approach, a state portrays its violation as a special case that is uninformative about future behavior. In a recommitment strategy, a state accepts responsibility for any wrongdoing but seeks to demonstrate through internal changes or compensatory measures that it plans to avoid similar actions in the future. Because recommitment entails more costly action, we theorize that it is most likely to occur when a government is seeking to avoid reputation costs from a domestic audience, whereas diminishment may be enough to appease international actors that care primarily about the regime itself.

Our two-stage theory of reputation management suggests reputational mechanisms operate in a highly complex and uncertain environment. The effect of institutional monitoring on compliance depends not only on the credibility and capability of monitoring bodies but also on the political process of interpreting non-compliance. This process is important to understand because of its relevance to questions of institutional design and the evolution of international regimes. Although we do not probe it in this paper, one implication of

our argument is that certain types of monitoring schemes may be more or less amenable to contestation. Regularized universal monitoring by technical experts, for example, may be more likely to furnish benefits to a broad cross-section of states, generating a higher level of investment in the regime. Regime legitimacy may also matter for how states engage with allegations of non-compliance. While states may still challenge monitoring from IOs that are widely perceived as legitimate, the content of such challenges is unlikely to undermine the regime. Indeed, if the IO's information can generate domestic reputation costs and lead to recommitment, contestation may ultimately help reinforce the monitoring institution. In contrast, an IO that faces repeated victim frames and legitimacy challenges may find itself spiraling toward oblivion. Future work should directly examine the effects of these strategies on the efficacy and stability of cooperation.

References

1948. *Universal Declaration of Human Rights*. adopted on 10 December 1948. UN General Assembly resolution 217 A.

1949a. *Geneva Convention Relative to the Protection of Civilian Persons in Time of War*. adopted and opened for signature, ratification and accession 12 August 1949. 75 U.N.T.S. 287.

1949b. *Geneva Convention Relative to the Treatment of Prisoners of War*. adopted and opened for signature, ratification and accession 12 August 1949. 75 U.N.T.S. 135.

1950. *Convention for the Protection of Human Rights and Fundamental Freedoms*. adopted and opened for signature, ratification and accession 11 April 1950. U.N.T.S. 213.

1989. *Rome Statute of the International Criminal Court*. Adopted and opened for signature, ratification and accession on 17 July 1998. U.N.T.S. 2187.

Agiza v. Sweden. 2005 (May). *Committee Against Torture*.

Alter, Karen J., & Meunier, Sophie. 2009. The politics of international regime complexity. *Perspectives on politics*, **7**(01), 13–24.

Axelrod, Robert. 1984. *The evolution of cooperation*. Vol. 5145. Basic Books (AZ).

Barnett, Michael N, & Finnemore, Martha. 1999. The politics, power, and pathologies of international organizations. *International organization*, **53**(4), 699–732.

Borzel, Tanja A., Hofman, Tobias, Panke, Diana, & Sprungk, Carina. 2010. Obstinate and Inefficient: Why Member States Do Not Comply with European Law. *Comparative Political Studies*, **43**(11), 1363–1390.

Bradley, Curtis A, & Kelley, Judith G. 2008. The Concept of International Delegation. *Law and Contemporary Problems*, **71**(1), 1–36.

Buchanan, Allen, & Keohane, Robert O. 2006. The legitimacy of global governance institutions. *Ethics & international affairs*, **20**(4), 405–437.

Bull, Hedley. 1977. *The Anarchical Society: A Study of Order in World Politics*. Macmillan International Higher Education.

Busch, Marc L. 2007. Overlapping institutions, forum shopping, and dispute settlement in international trade. *International Organization*, **61**(4), 735–761.

Carnegie, Allison, & Carson, Austin. 2018. The Spotlight's Harsh Glare: Rethinking Publicity and International Order. *International Organization*, **72**(3), 627–657.

Chaudoin, Stephen. 2016. How Contestation Moderates the Effects of International Institutions: The International Criminal Court and Kenya. *The Journal of Politics*, **78**(2), 557–571.

Chayes, Abram, & Chayes, Antonia Handler. 1993. On compliance. *International Organization*, **47**(02), 175–205.

Chong, Dennis, & Druckman, James N. 2007. Framing Public Opinion in Competitive Democracies. *American Political Science Review*, **101**(4), 637–655.

Committee Against Torture. 2016. Concluding Observations on the Fifth Periodic Report of Israel. June.

Committee Against Torture. 2017. *Concluding observations on the initial report of Pakistan*. CAT/C/PAK/CO/1, 1 June 2017.

Cooley, Alexander, & Snyder, Jack. 2015. *Ranking the World: Grading States as a Tool of Global Governance*. New York: Cambridge University Press.

Coombs, W Timothy. 1998. An analytic framework for crisis situations: Better responses from a better understanding of the situation. *Journal of public relations research*, **10**(3), 177–191.

Coombs, W Timothy. 2007. Protecting Organization Reputations During a Crisis: The Development and Application of Situational Crisis Communication Theory. *Corporate reputation review*, **10**(3), 163–176.

Coombs, W Timothy, & Holladay, Sherry J. 1996. Communication and attributions in a crisis: An experimental study in crisis communication. *Journal of public relations research*, **8**(4), 279–295.

Coombs, W Timothy, & Holladay, Sherry J. 2002. Helping crisis managers protect reputational assets: Initial tests of the situational crisis communication theory. *Management Communication Quarterly*, **16**(2), 165–186.

Crawford, Vincent P, & Sobel, Joel. 1982. Strategic Information Transmission. *Econometrica: Journal of the Econometric Society*, 1431–1451.

Dafoe, Allan, Renshon, Jonathan, & Huth, Paul. 2014. Reputation and status as motives for war. *Annual Review of Political Science*, **17**, 371–393.

Dai, Xinyuan. 2007. *International institutions and national policies*. Cambridge University Press.

Davis, Christina L. 2009. Overlapping institutions in trade policy. *Perspectives on politics*, **7**(01), 25–31.

Downs, George W, & Jones, Michael A. 2002. Reputation, compliance, and international law. *The Journal of Legal Studies*, **31**(S1), S95–S114.

Druckman, James N. 2001. The Implications of Framing Effects for Citizen Competence. *Political Behavior*, **23**(3), 225–256.

Franck, Thomas M. 1990. *The Power of Legitimacy Among Nations*. New York: Oxford University Press.

Garvey, Stephen P. 1998. Aggravation and Mitigation in Capital Cases: What do Jurors Think? *Columbia Law Review*, 1538–1576.

Gohara, Miriam S. 2013. Grace Notes: A Case for Making Mitigation the Heart of Noncapital Sentencing. *American Journal of Criminal Law*, **41**, 41.

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

Helfer, Laurence R. 2004. Regime shifting: the TRIPS agreement and new dynamics of international intellectual property lawmaking. *Yale J.Int'l L.*, **29**, 1.

Helfer, Laurence R. 2009. Regime shifting in the international intellectual property system. *Perspectives on Politics*, **7**(01), 39–44.

Holdsworth, Nick. 2015. Kremlin, Russian Media Slam Allegations of State-Sponsored Doping. *hollywoodreporter.com*, November 10.

Human Rights Watch. 2006 (November). *Sweden Violated Torture Ban in CIA Rendition*.

International Federation For Human Rights. 2009. *Kazakhstan/Kyrgyzstan: Exploitation of Migrant Workers, Protection Denied to Asylum Seekers and Refugees*.

Kelley, Judith G. 2017. *Scorecard Diplomacy: Grading States to Influence their Reputation and Behavior*. New York: Cambridge University Press.

Kelley, Judith G, & Simmons, Beth A. 2015. Politics by Number: Indicators as Social Pressure in International Relations. *American Journal of Political Science*, **59**(1), 55–70.

Keohane, Robert. 1984. 0.(1984) After Hegemony. *Cooperation and Discord in the World Political Economy, Princeton*.

Kono, Daniel Y. 2007. Making anarchy work: International legal institutions and trade cooperation. *Journal of Politics*, **69**(3), 746–759.

Kratochwil, Friedrich, & Ruggie, John Gerard. 1986. International Organization: A State of the Art on an Art of the State. *International organization*, **40**(4), 753–775.

Lupu, Yonatan, & Wallace, Geoffrey PR. 2018. Violence, Non-Violence, and the Effects of International Human Rights Law. *unpublished manuscript*.

Mearsheimer, John J. 1994. The false promise of international institutions. *International Security*, **19**(3), 5–49.

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

Milgrom, Paul R., North, Douglass C., & Weingast, Barry R. 1990. The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs. *Economics & Politics*, **2**(1), 1–23.

Minozzi, William. 2011. A Jamming Theory of Politics. *The Journal of Politics*, **73**(2), 301–315.

Minozzi, William, & Woon, Jonathan. 2016. Competition, Preference Uncertainty, and Jamming: A Strategic Communication Experiment. *Games and Economic Behavior*, **96**, 97–114.

Mohammed Alzery v. Sweden. 2006 (Oct). *Human Rights Committee*.

Morrow, James D. 2007. When Do States Follow the Laws of War? *American Political Science Review*, **101**(03), 559–572.

Morse, Julia C. Forthcoming. Blacklists, Market Enforcement, and the Global Regime to Combat Terrorist Financing. *International Organization*.

Morse, Julia C., & Keohane, Robert O. 2014. Contested multilateralism. *The Review of International Organizations*, 1–28.

Mutz, Diana. 2007. Effects of "In-Your-Face" Television Discourse on Perceptions of a Legitimate Opposition. *American Political Science Review*, **101**(4), 621–635.

Raustiala, Kal, & Victor, David G. 2004. The regime complex for plant genetic resources. *International organization*, 277–309.

Simmons, Beth A. 1998. Compliance with International Agreements. *Annual Review of Political Science*, **1**(1), 75–93.

Simmons, Beth A. 2000. International law and state behavior: Commitment and compliance in international monetary affairs. *American Political Science Review*, **94**(4), 819–835.

Simmons, Beth A, & Hopkins, Daniel J. 2005. The Constraining Power of International Treaties: Theory and Methods. *American Political Science Review*, **99**(4), 623–631.

Slothuus, Rune, & de Vreese, Claes H. 2010. Political Parties, Motivated Reasoning, and Issue Framing Effects. *Journal of Politics*, **72**(3), 630–645.

Steiker, Carol S, & Steiker, Jordan M. 1992. Let God Sort Them Out? Refining the Individualization Requirement in Capital Sentencing. *Yale Law Journal*, **102**, 835–870.

Stein, Arthur A. 2000. The Justifying State: Why Anarchy Doesnt Mean No Excuses. *Peace, Prosperity, and Politics*, 235–55.

Stinson, Scott. 2016. Red-Faced Russia Lashes Out; Conspiracy theories run amok despite overwhelming and damning evidence. *Vancouver Sun*, December 12.

Terman, Rochelle, & Voeten, Erik. 2018. The Relational Politics of Shame: Evidence from the Universal Periodic Review. *The Review of International Organizations*, **13**(1), 1–23.

The Economist. 2010 (March). *The Andean Laundry*.

Tomz, Michael. 2007. *Reputation and international cooperation: sovereign debt across three centuries*. Princeton University Press.

Trahan, Adam. 2011. In Their Own Words: Capital Jurors' Reactions to Mitigation Strategies. *International Journal of Punishment and Sentencing*, **7**, 1–16.