

*Matthew Dale Kim*

## REPUTATION AND COMPLIANCE WITH INTERNATIONAL HUMAN RIGHTS LAW: EXPERIMENTAL EVIDENCE FROM THE US AND SOUTH KOREA

### **Abstract**

Past studies suggest that domestic public support for compliance with international human rights law can constrain governments to comply with human rights law. But the question remains: Why does the public care about compliance? Using a series of survey experiments in South Korea and the United States, this study finds that constituents are concerned about compliance in one issue area—such as human rights—because they believe it will affect the country’s reputation in other domains of international law. Cross-national survey experiments demonstrate that past noncompliance negatively affects the South Korean public’s second-order beliefs about the likelihood of future compliance across different issue areas. However, past noncompliance has a limited impact on the US public’s first-order beliefs across different domains.

### **Keywords**

international human rights law, public opinion, issue-linkage, reputation, compliance

### INTRODUCTION

Can international human rights law cause greater state compliance with human rights norms in the absence of centralized enforcement? Due to the decentralized nature of international human rights law, many have questioned whether international legal commitments can lead to greater compliance with human rights norms. Studies have looked to domestic politics to suggest several potential mechanisms through which international law encourages state compliance. Many argue that human rights treaties empower domestic actors, such as the courts and civil society, to advocate for greater state compliance (Simmons 2009; Hafner-Burton and Tsutsui 2005; Powell and Staton 2009; Neumayer 2005; Cardenas 2007; Smith-Cannoy 2012). Meanwhile, a handful of other studies examine whether international law can shape public perception and whether the public can ensure greater compliance through a constituency-driven domestic compliance mechanism (Chilton 2014a and 2014b; Wallace 2013 and 2014; Putnam and Shapiro 2017; Tomz 2008; Kreps and Wallace 2016; Dai 2005).

An important question remains, however: Why do domestic constituents care about compliance? This study finds that, at least in one particular bilateral, asymmetric treaty relationship, the public is concerned with the reputational consequences of

noncompliance. South Korean constituents fear that other states would no longer be willing to make future agreements if their government were to renege on an existing treaty and thereby earn a reputation as a noncompliant state.

This study's findings can be summarized in three parts. First, this study finds that in an experimental setting, a single instance of past noncompliance can negatively affect first-order and second-order beliefs about future compliance in a similar issue area. First-order beliefs are beliefs that one actor has about another actor's behavioral tendencies (i.e. what you think about me). Second-order beliefs are beliefs that one actor has about what others believe about the actor's behavioral tendencies (i.e. what I think you think about me).<sup>1</sup> A foreign public's first-order beliefs about a noncompliant state's future compliance are affected by the state's past compliance behavior. That is, if South Korea has not complied with a similar issue area in the past, Americans will believe that South Korea is less likely to comply in the future. Meanwhile, the domestic public's second-order beliefs about a foreign public's first-order beliefs are affected by past compliance behavior as well. That is, South Koreans are less likely to believe that Americans will expect South Korea to comply in the future, if South Korea has not complied with a similar issue area in the past. This result suggests a reputational concern motivating domestic public demand for compliance.

Second, the study finds that post-noncompliance, first-order beliefs about noncompliance are not very fungible, meaning one's reputation for noncompliance spills over to different areas of international law in decreasing increments as the issue area under consideration is further removed from the issue of the original violation. That is, American beliefs about future South Korean compliance increase as the issue area under consideration is further removed from the issue area of South Korea's past noncompliance. Americans are pessimistic about future South Korean compliance if South Korea has not complied in the past, but this pessimism about future South Korean compliance behavior does not easily extend to different areas of international law.

Third, unlike first-order beliefs, past noncompliance can negatively affect second-order beliefs about future compliance in different issue areas. South Korean citizens believe that their own state's noncompliance will create reputational spillovers to different areas of international law, even though an examination of American first-order beliefs suggests otherwise. If South Korea has not complied with a particular issue area in the past, South Koreans believe that Americans will be pessimistic about South Korean compliance in other issue areas, even though Americans are more optimistic about future South Korean compliance in other issue areas. In sum, if the issue area of noncompliance is sufficiently salient to capture the public's attention, this study suggests that the public's perceived reputational consequences of noncompliance can be one potential mechanism encouraging state compliance with international law.

#### THEORETICAL AND EMPIRICAL FOUNDATIONS

Due to the decentralized nature of the international legal order, many have questioned why states in anarchy comply with international human rights law (Downs, Rocke, and Barsoom 1996; Von Stein 2005). Some argue that human rights treaties change domestic perceptions of human rights norms and that domestic actors, such as the courts and NGOs, can advocate for greater state compliance post-ratification

(Simmons 2009; Hafner-Burton and Tsutsui 2005; Powell and Staton 2009; Neumayer 2005; Cardenas 2007; Smith-Cannoy 2012). On the other hand, some studies examine whether international human rights law can shift public perceptions and shape public discourse towards compliance (Chilton 2014a and 2014b; Wallace 2013 and 2014; Putnam and Shapiro 2017; Tomz 2008; Kreps and Wallace 2016; Dai 2005). Changed public perceptions could pressure policymakers and thereby ensure greater compliance through a constituency-driven domestic compliance mechanism. For both strands of research, a focus on public perception is warranted considering how courts, activists, and elected officials all rely on *some* public support in ensuring greater compliance with international law (Ron 2017). Not surprisingly, there is growing body of literature focusing on public perceptions of human rights and public support for compliance with human rights (Koo 2017; Valentino and Weinberg 2017; Barton, Hillebrecht, and Wals 2017; Gerber 2017; Crow 2017; Murdie and Purser 2017; Pandya and Ron 2017).

However, despite the increasing scholarly focus on public support for compliance, it is still unclear why the public cares about this issue. In other words, while it is becoming more evident that the public supports compliance with international human rights treaties, it is unclear why the public cares about compliance. One possibility is that the public cares about the substantive commitment made in the treaty. That is, the public cares about the norms codified in the international human rights treaties and supports compliance because those norms are aligned with their values. Another possibility is that the public supports compliance because it believes noncompliance will have negative reputational effects.

#### REPUTATION

Rationalists and constructivists both elaborate on reputation as a reason for why the public cares about compliance. Rationalists argue that international law can operate as a commitment device in a system of anarchy through beliefs about reputation for future compliance (Morrow 2014). If a state reneges on its treaty commitments, then foreign publics may form beliefs about the reneging state. More specifically, foreign publics may assign the reneging state a reputation for noncompliance based on past behavior, which may reduce their state's willingness to cooperate with the reneging state in the future. When a state's past behavior causes others to view the state negatively in such a manner, that state is said to incur a reputation cost (Brutger and Kertzer 2018). If the noncompliant state's constituents believe that their state will incur a reputation cost for noncompliance, they may demand compliance. Meanwhile, constructivists argue that the normative belief that agreements should be complied with, *pacta sunt servanda*, creates an incentive for states to comply with agreements and avoid developing a reputation as a noncompliant state (Simmons 2010). This study therefore examines whether post-noncompliance reputational concerns are driving public demand for compliance. Does past noncompliance affect public beliefs about future compliance?

Before continuing further, a precise definition for reputation is warranted. Reputation is a broad term that traditionally refers to any belief about an actor or his or her behavioral tendencies based on the actor's past actions (Dafoe, Renshon, and Huth 2014). A state may assign other states a reputation to predict their future behavior (Mercer 1996). States can get assigned a reputation for many different traits or behavior such as a

reputation for resolve, for fulfilling threats, for hostility, and for greed (Dafoe, Renshon, and Huth 2014). Much of the literature on reputation focuses on reputation for resolve, rather than reputation for treaty compliance (Schelling 1966; Tang 2005; O'Neill 1996 and 2006; Kagan 1996; Lebow 2008; Walt 2015; Huth 1997; Kertzer 2016; Weisiger and Yarhi-Milo 2015). Some research on reputation for resolve suggests that political elites care deeply about their own state's reputation and believe their actions will have far-reaching reputational consequences (Schelling 1966; Tang 2005; O'Neill 1996 and 2006; Kagan 1996; Lebow 2008; Walt 2015). However, others argue that when policy-makers predict the credibility of other states' threats, they rely on present day behavior rather than reputation based on the other states' past behavior (Press 2005; Hopf 1994). Snyder and Borghard (2011) do not find strong evidence of reputation for resolve having a significant influence on elite decision-making, partly because elites rarely issue threats that are clear or commit the state's reputation in the first place.

While drawing on the theoretical implications of the divergent findings of past studies, I focus on reputation for treaty compliance rather than reputation for resolve, and whether the public, rather than elites, hold beliefs about reputation costs. This research focus is pertinent not only because of the aforementioned domestic compliance mechanisms that begin with public perceptions about international law, but also because of the nature of international legal commitments. Unlike reputation for resolve, where threats are not always clear, states should theoretically pay a high reputation cost for noncompliance with international treaties because when states sign a treaty or make a legal commitment, they do so in plain sight for all to see. Both the domestic public and the international community observe the commitment. International legal commitments also leave a paper trail that reduces the opportunity for deniability. They are essentially "the most solemn pledge a state can make" and a "maximal pledge of reputation" (Guzman 2008, 59; Lipson 2003; Simmons and Hopkins 2005). As such, the public should also be theoretically more likely to assign states with a reputation for treaty compliance based on post-ratification compliance behavior.

#### *QUESTIONS POSED*

In order for reputational concerns to be motivating public demand for compliance and thereby fueling the aforementioned domestic compliance mechanisms, a few key aspects of the reputational logic must be addressed. First, domestic constituents' beliefs about how other foreign publics perceive the constituency's own state must be affected by the state's past noncompliance. In other words, reputational logic requires that second-order beliefs—beliefs one actor has about what others believe about the actor's behavioral tendencies (i.e. what a noncompliant state's public believes about a foreign public's beliefs about the noncompliant state's future compliance)—are influenced by the past compliance behavior of one's own state. However, we have yet to systematically test whether the public believes that foreign publics will draw reputational inferences from the state's noncompliance with international law. As mentioned above, past studies have focused on elites and reputation for resolve to arrive at differing conclusions about the impact of reputation. It is also unclear whether first-order beliefs (i.e. what a foreign public believes about the noncompliant state's future compliance) are affected by past noncompliance. This study addresses these gaps by first exploring

whether noncompliance affects the public's first-order and second-order beliefs about future compliance.

Second, reputation costs should be fungible for reputational concerns to motivate public demand for compliance more broadly. The domestic public must believe that foreign publics will view reputation for compliance to be linked across different issue areas and therefore predictive of compliance behavior more broadly. If members of the public do not think that a foreign state's compliance with issue  $x$  at time  $t$  tells them anything about the same state's compliance with  $y$  at time  $t+1$ , then beliefs about reputation are less likely to result in public demand for compliance. That is, if the public does not consider reputation for issue  $x$  as linked with a different issue  $y$ , then the public's beliefs should only impact the assessment of future compliance behavior regarding that specific issue area  $x$ , reducing the public's perceived need and demand for compliance. Whether issues are linked and the extent to which reputation is fungible are a source of active scholarly debate (Weisiger and Yarhi-Milo 2015; Hopf 1994; Snyder and Diesing 1977). However, the focus of past studies has been on elites and these studies have arrived at different conclusions as to whether reputational inferences spill over to other issue areas. This study is the first, at least to my knowledge, to examine whether the public believes reputation for compliance in one particular issue area is linked to other issue areas of international law.

Third, this study explores the possibility of divergence between first-order and second-order beliefs. There is no reason, after all, for first-order and second-order beliefs to perfectly match all the time (Dafoe, Renshon, and Huth 2014). Past research on reputation, specifically the reputation for resolve, suggests that political elites believe that other states will draw reputational inferences while they themselves do not draw reputational inferences of others (Tang 2005). Tang (2005) refers to this paradoxical elite behavior as the "cult of reputation," where leaders are consistently preoccupied with the reputational consequences of their actions, even though they do not draw reputational inferences from other states' behavior. Tang writes, "[A]lthough a state constantly fears that others may assign reputation to it based on its past behavior, the state never assigns reputation to other states based on their past behavior" (Tang 2005, 42). Such studies suggest a potential for divergence between first-order and second-order beliefs in regards to issue-linkage, where the public's second-order beliefs are affected by past noncompliance in different issue areas, due to the public's constant concern about their state's own reputation, while the public's first-order beliefs are not affected by past noncompliance in different issue areas. That is, reputational concerns may not be sufficiently salient to permeate across issue areas for first-order beliefs, but they may be sufficiently salient to permeate across issue areas for second-order beliefs.<sup>2</sup>

Studies on motivated reasoning also suggest the possibility of divergence. According to Mercer's (1996) attribution-based framework for reputational inferences, leaders make reputational inferences differently depending on whether the foreign state is in the in-group or the out-group. Leaders have varying expectations for allies and adversaries; they are more forgiving of states in the in-group or allies, but less forgiving of states in the out-group or adversaries. Drawing on past studies of elites, given that the noncompliant state is a treaty partner, and thus a member of the in-group, foreign publics may try to reconcile their beliefs about a state willing to ratify a treaty and the state's subsequent noncompliance by attributing the instance of noncompliance to issue-specific

circumstances. That is, in order to resolve the dissonance between a treaty partner who has failed to follow through, foreign publics may be more willing to “explain away” the noncompliance as a momentary lapse and not link future compliance in different issue areas to the instance of past noncompliance in a particular issue area. In contrast, the domestic public of a noncompliant state may assume the worst, along the lines of Tang’s (2005) cult of reputation, and believe that foreign publics will draw reputational inferences across issue areas. This study thus explores whether foreign publics consider another state’s past noncompliance as less indicative of the state’s future compliance in different issue areas and whether the domestic public of the noncompliant state nonetheless believes that foreign publics will link different issues areas.

## METHODS

I conducted cross-national surveys in South Korea and the United States that served as mirrors of each other to explore the impact of past noncompliance on the formation of first-order and second-order beliefs about future compliance. Survey experiments were ideal for this study because reputation costs are considered the “dark matter of international relations” (Dafoe and Caughey 2016, 372) and difficult to study in an uncontrolled environment in which leaders avoid incurring reputation costs (Brutger and Kertzer 2018). The experimental setting allowed for the manipulation of the key variables—compliance or noncompliance behavior and the issue area being considered for future cooperation—while holding other variables constant.

## *SURVEY TEXT AND HYPOTHESES*

To elaborate further, I first estimated the impact of noncompliance versus compliance on first-order beliefs and second-order beliefs about the likelihood of a noncompliant state’s future compliance. This first step allowed me to estimate the effect of noncompliance against the compliance treatment condition. The exact procedure is as follows. All US respondents read a hypothetical vignette that South Korea had signed a security treaty with the United States, which included a stipulation about the humane treatment of prisoners of war (POWs). Respondents were then told of South Korean noncompliance (i.e. noncompliance treatment group) or compliance (i.e. compliance treatment group) with the treaty. Respondents in the noncompliance treatment group were told that although South Korea had ratified the treaty, there was conclusive evidence that South Korea violated the agreement by torturing North Korean covert operatives who were captured after infiltrating the border into South Korean territory.<sup>3</sup> Respondents in both groups were then told of the possibility of signing a new treaty in a different issue area within the domain of the laws of war—the preservation of artifacts during war. All respondents were then asked whether they believed South Korea would abide by the new agreement if signed. More precisely, U.S. respondents read one of the following versions.

Our ally South Korea previously signed a security treaty with our country. In the security treaty, South Korea made a formal commitment not to mistreat POWs for a military advantage. The security agreement created an independent organization to observe the treatment of POWs.

*[Noncompliance treatment: The organization recently reported that South Korea violated the security treaty when dealing with captured North Korean special forces (spies) who had infiltrated the border in the past few years.]*

*[Compliance treatment: The organization recently reported that South Korea faithfully followed through with the treaty. Whenever South Korea captured North Korean special forces (spies) who had infiltrated the border, they were treated in accordance with the international treaty.]*

The US is now considering a new security agreement with South Korea that would prohibit South Korea from destroying historical artifacts and monuments during war. The new security agreement does not deal with POWs. Instead, the new security agreement deals with the preservation of historical artifacts during war. The new security agreement will allow an independent organization to monitor historical sites. The monitoring organization has been successful in detecting the destruction of historical artifacts and monuments in the past.

How likely do you think South Korea will stick to the agreement if signed?

Next, the South Korean survey posed the same vignette and asked Korean respondents about the extent to which they thought Americans would trust South Korea to abide by a new agreement regarding the preservation of artifacts, given South Korean noncompliance or compliance. The South Korean survey thus measured second-order beliefs.

[Translated from Korean; the original Korean survey text is included in the Supplementary Materials to this article]

The United States previously signed a security treaty with our country. In the security treaty, our country made a formal commitment not to mistreat POWs for a military advantage. The security agreement created an independent organization to observe the treatment of POWs.

*[Noncompliance treatment: The organization recently reported that our country violated the security treaty when dealing with captured North Korean special forces (spies) who had infiltrated the border in the past few years.]*

*[Compliance treatment: The organization recently reported that our country faithfully followed through with the treaty. Whenever our country captured North Korean special forces (spies) who had infiltrated the border, they were treated in accordance with the international treaty.]*

The US is now considering a new security agreement with our country that would prohibit our country from destroying historical artifacts and monuments during war. The new security agreement does not deal with POWs. Instead, the new security agreement deals with the preservation of historical artifacts during war. The new security agreement will allow an independent organization to monitor historical sites. The monitoring organization has been successful in detecting the destruction of historical artifacts and monuments in the past.

How likely do you think Americans will believe our country to stick to the new agreement if signed?

Given the literature on reputation for resolve suggesting that states incur a reputation for past behavior (Schelling 1966; Tang 2005; O'Neill 1996 and 2006; Kagan 1996; Lebow 2008; Walt 2015), I made the following hypotheses regarding reputation for treaty compliance. I expected US first-order beliefs about future South Korean compliance to be negatively affected by South Korea's past noncompliance. I also expected South Korean second-order beliefs about US beliefs about future South Korean compliance to be negatively affected by South Korea's past noncompliance.

*H1 (first-order beliefs post-noncompliance): US respondents will be less likely to believe that South Korea will comply with a new treaty post-noncompliance than post-compliance.*

*H2 (second-order beliefs post-noncompliance): South Korean respondents will be less likely to believe that Americans will believe South Korea will comply with a new treaty post-noncompliance than post-compliance.*

Second, in a separate stage of the experiment, I explored post-noncompliance issue-linkage by adding two noncompliance treatment conditions. Respondents were told, post-noncompliance, that the new agreement was in the same exact issue area, a different issue area within the domain of the laws of war (i.e. the noncompliance treatment condition from above), or an issue area outside the domain of the laws of war, all post-noncompliance. That is, the issue area of the new treaty moved progressively further away from the issue of the original treaty to test whether noncompliance with one issue was linked with similar and dissimilar issues. Respondents were told that the United States and South Korea were considering signing a new treaty on POWs, a new treaty on the preservation of artifacts during war, or a new treaty on the prohibition on trade tariffs on select goods. By holding South Korean noncompliance constant, and instead varying the issue area of the future treaty, I could test the extent of issue-linkage for both first-order and second-order beliefs. All of the treatment conditions specified whether the issue area of the new treaty was within or beyond the domain of the original treaty. [Figure 1](#) summarizes all of the treatment conditions.

The respondents read one of the following versions.

For the US:

Our ally South Korea previously signed a security treaty with our country. In the security treaty, South Korea made a formal commitment not to mistreat POWs for a military advantage. The security agreement created an independent organization to observe the treatment of POWs. The organization recently reported that South Korea violated the security treaty when dealing with captured North Korean special forces (spies) who had infiltrated the border in the past few years.

*[POW treatment: The US is now considering a new security agreement with South Korea that would prohibit South Korea from mistreating POWs. The new security agreement will allow the independent organization to continue observing the treatment of POWs.]*

*[Artifact treatment (noncompliance treatment from above): The US is now considering a new security agreement with South Korea that would prohibit South Korea from destroying historical artifacts and monuments during war. The new security agreement does not deal with POWs. Instead, the new security agreement deals with the preservation of historical artifacts during war. The new security agreement will allow an independent organization to monitor historical sites. The monitoring organization has been successful in detecting the destruction of historical artifacts and monuments in the past.]*

*[Tariff treatment: The US is now considering a new trade agreement with South Korea. The trade agreement would prohibit South Korea from imposing harmful trade tariffs on select US exports entering South Korea. The new trade agreement does not deal with POWs or war. Instead, the agreement deals with the separate area of international trade and commerce. The trade agreement will allow an independent organization to monitor trade between the two countries. The monitoring organization has been successful in detecting attempts to impose tariffs in the past.]*

How likely do you think South Korea will stick to the agreement if signed?



**FIGURE 1 Treatment conditions**

| Issue Area of Future Treaty                            | United States   |                    | South Korea     |                    |
|--------------------------------------------------------|-----------------|--------------------|-----------------|--------------------|
|                                                        | Post-compliance | Post-noncompliance | Post-compliance | Post-noncompliance |
| Same issue area (POW treaty)                           |                 | ✓                  |                 | ✓                  |
| Different issue area, same domain (artifact treaty)    | ✓               | ✓                  | ✓               | ✓                  |
| Different issue area, different domain (tariff treaty) |                 | ✓                  |                 | ✓                  |

For South Koreans:

[Translated from Korean; the original Korean survey text is included in the Supplementary Materials]

The United States previously signed a security treaty with our country. In the security treaty, our country made a formal commitment not to mistreat POWs for a military advantage. The security agreement created an independent organization to observe the treatment of POWs. The organization recently reported that our country violated the security treaty when dealing with captured North Korean special forces (spies) who had infiltrated the border in the past few years.

*[POW treatment: The US is now considering a new security agreement with our country that would prohibit our country from mistreating POWs. The new security agreement will allow the independent organization to continue observing the treatment of POWs.]*

*[Artifact treatment (noncompliance treatment from above): The US is now considering a new security agreement with our country that would prohibit our country from destroying historical artifacts and monuments during war. The new security agreement does not deal with POWs. Instead, the new security agreement deals with the preservation of historical artifacts during war. The new security agreement will allow an independent organization to monitor historical sites. The monitoring organization has been successful in detecting the destruction of historical artifacts and monuments in the past.]*

*[Tariff treatment: The US is now considering a new trade agreement with our country. The trade agreement would prohibit our country from imposing harmful trade tariffs on select US exports entering our country. The new trade agreement does not deal with POWs or war. Instead, the agreement deals with the separate area of international trade and commerce. The trade agreement will allow an independent organization to monitor trade between the two countries. The monitoring organization has been successful in detecting attempts to impose tariffs in the past.]*

How likely do you think Americans will believe our country to stick to the new agreement if signed?

Based on the literature suggesting that leaders do not draw reputational inferences across domains (Weisiger and Yarhi-Milo 2015; Hopf 1994; Snyder and Diesing 1977), as well as the literature on motivated reasoning (Mercer 1996), I made the following hypotheses regarding first-order beliefs and issue-linkage. I expected US first-order beliefs to be less affected by past South Korean noncompliance in different issue areas.

That is, I expected US beliefs about future South Korean compliance in a particular issue area to be less affected by past South Korean noncompliance in other issue areas.

*H3 (first-order beliefs and issue linkage): US respondents will be less likely to believe that South Korea will comply with a new treaty on the same issue as the original violation (POW treaty) than a new treaty on a separate issue within the same domain as the original violation (Artifact treaty).*

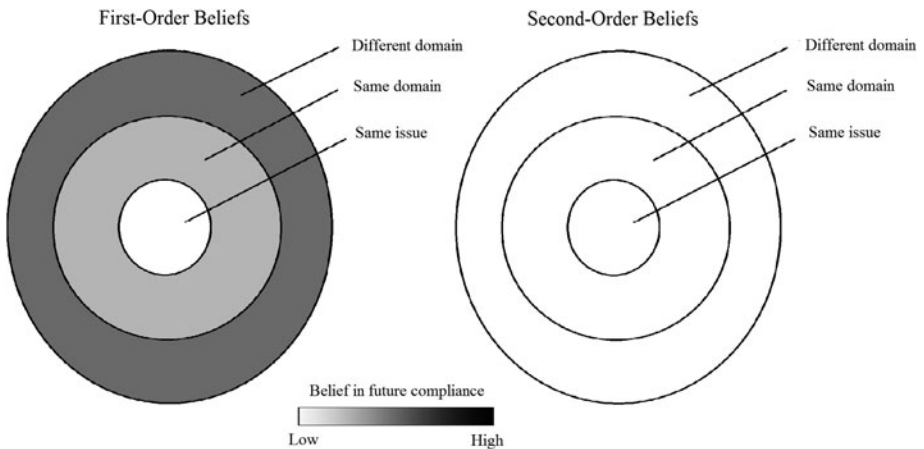
*H4 (first-order beliefs and issue linkage): US respondents will be less likely to believe that South Korea will comply with a new treaty on a separate issue within the same domain as the original violation (Artifact treaty) than a new treaty on a separate issue outside the domain (Tariff treaty).*

Meanwhile, based on the literature suggesting that political elites believe that other states will draw reputational inferences while they themselves do not draw reputational inferences of others (Tang 2005; Dafoe, Renshon, and Huth 2014), I expected second-order beliefs to be strongly affected by noncompliance in different issue areas. I expected that, due to the public’s concern about their own state’s reputation, the public’s second-order beliefs would be strongly affected by past noncompliance in different issue areas. In other words, I expected South Korean beliefs about US beliefs about future South Korean compliance to be affected by past South Korean noncompliance in other issue areas. The hypotheses are also depicted in Figure 2.

*H5 (second-order beliefs and issue linkage): The South Korean public’s second-order beliefs about future South Korean compliance with a new treaty on the same issue as the original violation (POW treaty) will be similar to South Korean second-order beliefs about future South Korean compliance with a new treaty on a separate issue within the same domain as the original violation (Artifact treaty).*

*H6 (second-order beliefs and issue linkage): The South Korean public’s second-order beliefs about future South Korean compliance with a new treaty on a separate issue within the same domain as the original violation (Artifact treaty) will be similar to South Korean second-order beliefs about future South Korean compliance with a new treaty on a separate issue outside the domain (Tariff treaty).*

**FIGURE 2 Hypothesized Spillover Effects of Noncompliance**



*SAMPLE DEMOGRAPHICS*

Relevant demographic characteristics were collected for later analysis. These included standard demographic characteristics widely used in the literature such as age, gender, education level, employment status, income, and race (Moravcsik 2005). In addition, I measured political ideology, knowledge level, and political activism using standard questions from the American National Election Studies [ANES].<sup>4</sup> Other relevant predispositions also drawn from existing studies included levels of cooperative internationalism, militant internationalism, and isolationism (Kertzer, Powers, Rathbun, and Iyer 2014).<sup>5</sup> To be clear, due to the randomization of the treatment assignment, statistical models involving covariates are not necessary for estimating unbiased treatment effects (Wallace 2013; Mutz and Permantle 2015). Therefore, the main models in the following analyses do not include the covariates, and models controlling for covariates are included only as robustness checks.

The surveys involved a sample of 1,098 American citizens and 1,000 South Korean citizens. The survey of Americans was conducted through Amazon Mechanical Turk [MTurk] on October 22 and 23, 2016, and the survey of South Koreans was conducted by a professional polling firm Macromill Embrain from November 16 to November 22, 2016. Sample characteristics for US respondents and Korean respondents used in this study have been provided in the Supplementary Materials. It should be noted that typical American MTurk samples are typically not nationally representative, and they tend to over-represent younger, liberal, lower income individuals while under-representing Hispanics and African Americans. This is the case here: the study's US sample slightly over-represents liberals while under-representing Hispanics and African Americans. However, all of the following results using the US MTurk sample can be replicated using post-stratification weights to account for deviations from a nationally representative sample. These results have also been reproduced in the Supplementary Materials as robustness checks.<sup>6</sup> Meanwhile, Macromill Embrain ensured that the South Korean sample was representative on key demographic characteristics such as gender, political ideology, age, and education. The outcome measures—whether US respondents believed South Korea would comply and whether South Korean respondents believed that Americans would believe South Korea would comply—were measured on a 7-point Likert scale but converted to percentage points for analysis.<sup>7</sup>

*CASE SELECTION*

The selection of an actual state, South Korea, as the counterpart to the US had several methodological advantages. Using a real state, as opposed to a hypothetical state, allowed respondents to bring in their prior beliefs about the dyad. As such, this experimental setup mimicked a more realistic setting in which respondents update their prior beliefs about a specific state based on new information to guide their final decision. Using a hypothetical state for which respondents do not have any priors about compliance behavior would have exaggerated the effect of any single instance of noncompliance. It should be made clear, though, that using actual states did not mirror reality perfectly since in reality people are often bombarded with new information that can drown out the effect of any single instance of compliance or noncompliance.

However, survey experiments are rarely, if ever, able to mimic reality precisely. Using a dyad grounded in reality mitigated concerns about validity somewhat by allowing prior beliefs to play a role. At the very least, this experimental setup allowed for findings about the effect of noncompliance on the margins.

Viewed more reasonably, the selection of these two particular states also allowed for a deeper exploration of US–South Korean relations, which is a worthwhile endeavor in itself. The US and South Korean conceptions of reputation costs as it pertains to their bilateral relationship have far-reaching geopolitical implications, given their close military alliance and the numerous bilateral agreements to which the two states are party. In addition, the asymmetric nature of the US–South Korean relationship allowed for potential insights about how power discrepancies may affect each polity's perception of the reputation costs in an unequal bilateral relationship, if not in bilateral relationships more generally.

Moreover, in designing the experiment, it was essential that respondents from both states accepted the premise of the treatment and found the hypothetical to be plausible. The torture of POWs is an issue that receives much media attention in reality, and the hypothetical case had to concern a state that concealed its treatment of POWs and simultaneously faced a sufficiently dire security threat to have a strong incentive to engage in torture. Fortunately for the purposes of this study, the actual treatment of captured North Korean covert operatives by South Korean officials is confidential. While South Korean security forces often capture North Korean operatives in South Korea, it is unclear what type of treatment these operatives are subjected to. Moreover, because of the unique national security threat South Korea faces from its North Korean neighbors, South Korea has an incentive to engage in enhanced interrogation techniques. Such factors gave the vignette a degree of plausibility. The reverse case in which the United States was the noncompliant state was precluded due to similar concerns about plausibility.

Other premises of the vignette, namely as the existence of a security treaty, were also plausible given the military alliance the two states have shared for decades. The particular treatment condition that stated that the two states were considering a new trade treaty also reflected reality. There were lengthy negotiations about trade practices between the two states, followed by the eventual ratification of the US–South Korea Free Trade Agreement, which was being hotly debated in South Korea at the time of the survey. In short, ensuring the plausibility of the vignette in order to mitigate the risk of respondents rejecting the premise of the treatment dictated the choice of South Korea as the counterpart to the United States.

While the selection of actual states helped the experiment mimic reality by allowing priors to play a role and to probe deeper into the US–South Korean relationship, selecting these two states for this survey also had its disadvantages, specifically in terms of generalizability. This study's findings may be limited to asymmetric dyads. Yet a large number of the treaty commitments made by the US involve asymmetric dyads anyway. The prevalence of US treaty commitments alone likely justifies the study's relevance despite its limited generalizability. One could argue that the resulting treatment effects are also dyad-specific. That is, the case selection may limit the generalizability of the findings to that of the US–South Korea dyad and at most an asymmetric bilateral relationship. Reactions to the treatments could have been tempered by dyad-specific biases, such as feelings of mutual trust that have developed between the two countries

over the past few decades. On the other hand, some Korean respondents, influenced by rising tensions between the two states and strong anti-American sentiments, could have instead allowed their preconceived notions to color their perception of how Americans may respond to South Korean noncompliance. Some US respondents could have been abnormally influenced by the perception that their allies are not doing their fair share in securing their own national security. Within segments of the American populace, there is a widely held belief that American allies such as South Korea are “free-loading”—benefitting from American military commitments while giving nothing in return. During the 2016 presidential election, then-candidate Donald Trump noted how the United States gets “practically nothing” in return for protecting South Korea (Pruitt 2016). Such concerns may have negatively affected American first-order beliefs about future South Korean compliance behavior in ways that are not generalizable to other states.

It is also possible that differences in post-noncompliance beliefs about future compliance across different issue areas were due to differing baseline expectations about compliance in certain issue areas, rather than decreasing issue-linkage. For instance, Americans may have believed that South Korea was more likely to comply with a tariff treaty than an artifact treaty due to general expectations about compliance with tariffs and therefore have a greater belief in South Korea’s future compliance with the tariff treaty. Such an explanation would suggest that issue-linkage and past noncompliance do not play a significant role. To mitigate some of the concerns, I asked opened-ended questions afterwards, which are discussed below and in the Supplementary Materials.

Finally, this study’s case selection cannot account for the possibility that cross-national differences between first-order and second-order beliefs were due to baseline differences between Americans and South Koreans about issue-linkage. It is possible that Americans do not believe in issue-linkage while South Koreans do. However, there is no obvious theoretical reason to expect such baseline differences between Americans and South Koreans. Rather, one would theoretically expect the publics of all countries to hold similar attitudes towards issue-linkage, and the results are more likely a reflection of the difference between first-order and second-order beliefs. Nonetheless, to address these concerns, a future study could hold constant the state and measure first-order and second-order beliefs from people of the same state. For instance, one could imagine a survey of American first-order beliefs post-South Korean noncompliance and another separate survey of American second-order beliefs about South Korean beliefs post-American noncompliance. Such a research design, of course, would have to resolve plausibility concerns that were carefully addressed in this study’s research design.

Overall, the benefits of mimicking reality and reducing the risk of respondents rejecting the premise of the treatment outweighed concerns about generalizability and ultimately motivated the case selection of South Korea and the United States. The more skeptical reader may interpret the following results as applying only to one specific dyad or asymmetrical bilateral relationships, as opposed to bilateral relationships more generally.

## RESULTS

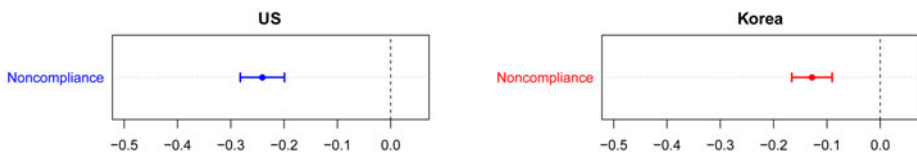
First, I estimated the effect of noncompliance on first-order beliefs using an OLS model. Comparing the compliance treatment to the noncompliance treatment, both of which

considered future compliance regarding the same issue area (i.e. preservation of artifacts), noncompliance decreased first-order beliefs about future South Korean compliance by 23.3 percentage points. Noncompliance also decreased South Koreans' second-order beliefs about future compliance by 12.6 percentage points. The results are summarized in [Figure 3](#) and [Table 1](#). In short, I found support for my first hypothesis regarding first-order beliefs about compliance and for my second hypothesis regarding second-order beliefs about compliance. Past noncompliance decreases first-order and second-order beliefs about future compliance.

All results are robust to models that control for pre-treatment covariates. Proportional odds models often used for Likert scale responses resulted in similar findings have been reproduced in the Supplementary Materials for these models and all subsequent models. It should be noted that a comparison of the magnitude of the treatment effects between the two countries (23.3 versus 12.6 percentage points) is misleading because cross-national differences in cell means could be the result of differences in the translation of the survey questions or a variety of other confounding factors. Rather, the takeaway should be that noncompliance, as opposed to compliance, leads to reduced first-order and second-order beliefs about future compliance.

Next, I compared the post-noncompliance treatment outcomes to each other in order to estimate the extent of issue-linkage, holding constant past noncompliance behavior. First, the following analyses using an OLS model reveal low issue-linkage for first-order beliefs in the US. While noncompliance will negatively affect first-order beliefs about a noncompliant state's future compliance behavior, respondents are more likely to believe that South Korea will comply in the future as one moves further away from the issue of the original noncompliance. Treatment conditions (i.e. issue areas) that were hypothesized to be comparably less affected by past noncompliance were in fact less affected. A new treaty in a different issue area in the same domain as the original noncompliance—a new treaty about the preservation of artifacts during war which is in the domain of a security treaty—increased US first-order beliefs regarding future South Korean compliance, compared to a new treaty in the same issue area as the original violation. The difference was a statistically significant 4.7 percentage points. A new treaty in a different issue area in a separate domain—a new treaty on trade tariffs, which is in the domain of trade not security—increased US first-order beliefs in future South Korean compliance, compared to a new treaty in a different issue area in the same domain as the original violation. The difference was a statistically significant 10.0 percentage points. In sum, first-order beliefs about future compliance are less affected by past noncompliance in different issue areas.

**FIGURE 3** Estimated Effects of Noncompliance (OLS)



The effect of noncompliance compared to compliance on first-order beliefs and second-order beliefs about future compliance is shown above. The figure depicts average treatment effects from OLS models that do not include pre-treatment covariates

**TABLE 1** Estimated Effects of Noncompliance on First-Order and Second-Order Beliefs (OLS)

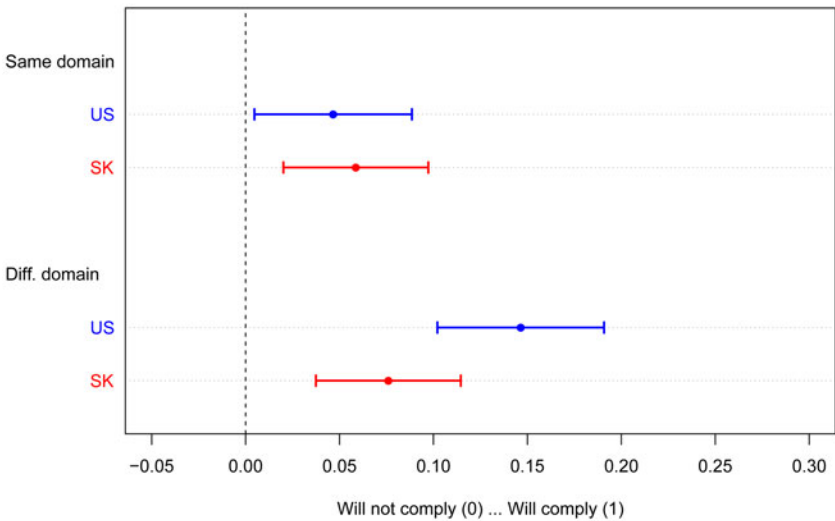
|                              | US (Model 1)             | US (Model 2)             | South Korea (Model 1)   | South Korea (Model 2)   |
|------------------------------|--------------------------|--------------------------|-------------------------|-------------------------|
| Noncompliance                | -0.241***<br>(0.021)     | -0.233***<br>(0.021)     | -0.128***<br>(0.019)    | -0.126***<br>(0.020)    |
| Gender                       |                          | 0.057***<br>(0.022)      |                         | 0.010<br>(0.020)        |
| Political ideology           |                          | -0.006<br>(0.007)        |                         | -0.001<br>(0.010)       |
| Knowledge                    |                          | 0.152**<br>(0.066)       |                         | 0.048<br>(0.047)        |
| Cooperative internationalism |                          | 0.109*<br>(0.058)        |                         | -0.0001<br>(0.082)      |
| Military internationalism    |                          | 0.001<br>(0.052)         |                         | 0.146**<br>(0.058)      |
| Isolationism                 |                          | -0.060<br>(0.053)        |                         | 0.001<br>(0.029)        |
| Constant                     | 0.707***<br>(0.015)      | 0.620***<br>(0.095)      | 0.579***<br>(0.014)     | 0.478***<br>(0.078)     |
| Observations                 | 552                      | 552                      | 500                     | 500                     |
| R <sup>2</sup>               | 0.191                    | 0.241                    | 0.081                   | 0.114                   |
| Adjusted R <sup>2</sup>      | 0.189                    | 0.219                    | 0.079                   | 0.092                   |
| Residual Std. Error          | 0.248 (df = 550)         | 0.243 (df = 535)         | 0.216 (df = 498)        | 0.214 (df = 487)        |
| F Statistic                  | 129.788*** (df = 1; 550) | 10.634*** (df = 16; 535) | 43.917*** (df = 1; 498) | 5.211*** (df = 12; 487) |

\*p < 0.1; \*\*p < 0.05; \*\*\*p < 0.01

Note: The reference group is the compliance treatment group. Not all pre-treatment covariates that were used in estimating each state's respective Model 2 have been shown above. Additional pre-treatment covariates not shown above are age, education, employment status, income, race (for US only), and political involvement. Political ideology represents individual respondents' political leaning with 0 being liberal and 1 being conservative. The dependent variable has been converted to percentage points. For first-order beliefs (US), 1 signifies a 100 percentage point increase in US respondents' beliefs regarding the likelihood of South Korea's future compliance. For second-order beliefs (South Korea), 1 signifies a 100 percentage point increase in South Korean respondents' beliefs regarding U.S. beliefs about the likelihood of future South Korean compliance.

Unlike first-order beliefs, second-order beliefs have higher issue-linkage. Second-order beliefs—what South Koreans believed about US beliefs about future South Korean compliance—seemed to be affected by past noncompliance in related and unrelated issue areas, even though US first-order beliefs about future South Korean compliance were not affected by past compliance in different issue areas. As Table 2 and Figure 4 demonstrate, among South Korean respondents, treatment conditions (i.e. issue areas) that were hypothesized to be comparably less affected by past noncompliance were not always less affected. A new treaty in a different issue area in the same domain—the preservation of artifacts during war, which is in the domain of a security treaty—increased South Korean second-order beliefs, compared to a new treaty in the same issue area as the original violation. The difference was a statistically significant 5.9 percentage points. However, once moving beyond the issue of original violation, and comparing separate issues across domains, South Koreans did not believe Americans would distinguish between separate domains. Instead, they believed that Americans would maintain a constant level of skepticism about future compliance. Second-order beliefs about future South Korean compliance for a new treaty in a different issue area in a separate domain—a new treaty on trade tariffs which is in the domain of trade not security—was statistically indistinguishable to second-order beliefs about a new treaty in a different issue area in the same domain as the original violation. The difference was a statistically insignificant 1.7 percentage points. To summarize, I failed to find evidence for my fifth hypothesis regarding comparable second-order beliefs about compliance for different issue areas within the same domain. However, I did find strong evidence for my sixth hypothesis regarding comparable second-order beliefs about compliance for different issue areas in separate domains. Once a state fails to comply, its

**FIGURE 4** Estimated Spillover Effects of Noncompliance (OLS)



Each treatment condition is compared to their respective countries' POW treaty treatment condition denoted by the dotted line. The figure depicts average treatment effects from both states using OLS models that do not include pre-treatment covariates



**TABLE 2** Estimated First-Order and Second-Order Spillover Effects (OLS)

|                              | US (Model 1)            | US (Model 2)            | South Korea (Model 1)  | South Korea (Model 2)   |
|------------------------------|-------------------------|-------------------------|------------------------|-------------------------|
| Same domain                  | 0.047**<br>(0.021)      | 0.051**<br>(0.021)      | 0.059***<br>(0.020)    | 0.060***<br>(0.019)     |
| Different domain             | 0.146***<br>(0.023)     | 0.151***<br>(0.022)     | 0.076***<br>(0.020)    | 0.070***<br>(0.020)     |
| Gender                       |                         | 0.073***<br>(0.018)     |                        | 0.044**<br>(0.017)      |
| Political ideology           |                         | -0.009<br>(0.006)       |                        | 0.018**<br>(0.008)      |
| Knowledge                    |                         | 0.179***<br>(0.059)     |                        | 0.092**<br>(0.038)      |
| Cooperative internationalism |                         | 0.090*<br>(0.048)       |                        | -0.029<br>(0.068)       |
| Military internationalism    |                         | 0.035<br>(0.043)        |                        | 0.052<br>(0.047)        |
| Isolationism                 |                         | -0.108**<br>(0.043)     |                        | -0.008<br>(0.024)       |
| Constant                     | 0.419***<br>(0.015)     | 0.424***<br>(0.075)     | 0.393***<br>(0.014)    | 0.258***<br>(0.063)     |
| Observations                 | 840                     | 840                     | 750                    | 750                     |
| R <sup>2</sup>               | 0.049                   | 0.130                   | 0.021                  | 0.069                   |
| Adjusted R <sup>2</sup>      | 0.046                   | 0.112                   | 0.019                  | 0.052                   |
| Residual Std. Error          | 0.262 (df = 837)        | 0.253 (df = 822)        | 0.220 (df = 747)       | 0.216 (df = 736)        |
| F Statistic                  | 21.355*** (df = 2; 837) | 7.227*** (df = 17; 822) | 8.197*** (df = 2; 747) | 4.178*** (df = 13; 736) |

\*p < 0.1; \*\*p < 0.05; \*\*\*p < 0.01.

Note: The reference group is the noncompliance treatment group that considers the same issue as the original treaty, post-noncompliance. Not all pre-treatment covariates that were used in estimating each state's respective Model 2 have been shown above. Additional pre-treatment covariates not shown above are age, education, employment status, income, race (for US only), and political involvement. Political ideology represents individual respondents' political leaning with 0 being liberal and 1 being conservative. The dependent variable has been converted to percentage points. For first-order beliefs (US), 1 signifies a 100 percentage point increase in U.S. respondents' beliefs regarding the likelihood of South Korea's future compliance. For second-order beliefs (South Korea), 1 signifies a 100 percentage point increase in South Korean respondents' beliefs regarding U.S. beliefs about the likelihood of future South Korean compliance.

constituents believe that their state has lost much credibility on the issue of the original violation, and they also believe that their state has incurred a consistent loss in credibility in different other issue areas, no matter how far removed these issues are from the original violation.

I now turn to differences between first-order and second-order beliefs. As mentioned earlier, a simple comparison of cell means across states is not appropriate. Cross-national differences in cell means could be the result a variety of confounding factors between the two states. Instead, I compared the differences in the treatment effects within each state. As Figures 4 and 5 show, South Koreans believed that Americans would treat separate domains equally and would not draw a distinction between separate domains. In actuality, Americans did take into account different domains and treated separate issues and domains with progressively smaller reputational spillovers.

One key limitation of this study should be addressed. As mentioned above, it is possible that differences in post-noncompliance beliefs about future compliance across different issue areas were due to differing baseline expectations about compliance in certain issue areas, rather than decreasing issue-linkage. To address this concern, the experiment asked respondents to provide open-ended responses about the reasoning for their decisions. Respondents in all treatment groups were prompted to provide a reasoning for their beliefs about compliance. Using structural topic models, I then identified several common reasonings (i.e. “topics”) provided by the respondents and afterwards determined which treatment condition was more likely to trigger certain reasonings. In short, I found some evidence that past noncompliance affected first-order and second-order beliefs about future compliance due to reputational concerns, not due to varying baseline beliefs about compliance in certain issue areas.

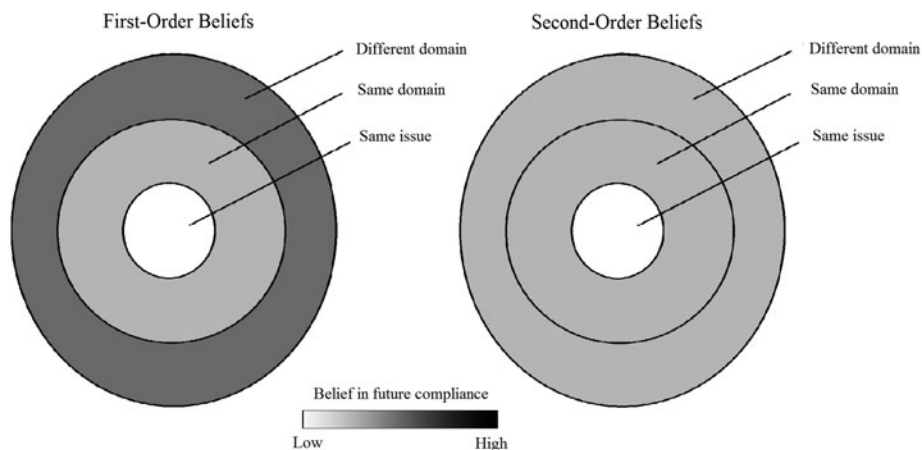
For first-order beliefs, American respondents assigned to noncompliance treatment groups, who were progressively more likely to believe in future South Korean compliance as the issue area was further removed from the original violation, were more likely to justify their beliefs by citing how they viewed different issue areas as separate issues in which previous noncompliance mattered less, not by citing different baseline expectations about compliance in a particular issue area. For instance, one US respondent who received the noncompliance treatment about a new treaty in a different issue area as the original violation reasoned that “I think that they [South Korea] will uphold their

**TABLE 3 Estimated Spillover Effects of Noncompliance (OLS)**

|                  | US    | Difference in effect sizes | South Korea | Difference in effect sizes |
|------------------|-------|----------------------------|-------------|----------------------------|
| Same issue       | 0.419 |                            | 0.393       |                            |
| Same domain      | 0.466 | 0.047**                    | 0.451       | 0.059***                   |
| Different domain | 0.566 | 0.100***                   | 0.469       | 0.017                      |

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

Note: The table notes baseline support for compliance and average treatment effects from OLS models that do not include pre-treatment covariates.

**FIGURE 5** Estimated Spillover Effects of Noncompliance

[new] deal with the US and the treatment of North Korean POWs is unrelated,” rather than reasoning that it was easier to comply with an artifact treaty than a POW treaty.

For second-order beliefs, Korean respondents, who mostly demonstrated a constant level of skepticism across different domains once moving beyond the issue of the original violation, often reasoned that they weighed the previous violation equally regardless of the domain of the new treaty. For instance, one South Korean respondent wrote that “If we [South Korea] did not keep the original treaty, even if we made a new treaty they [US] would not think that we would not comply well. We already lost their trust.” Although there is some anecdotal evidence of differences in baseline expectations for each issue area, it appears overall that the estimated treatment effects stem from past noncompliance and diverging beliefs about issue-linkage, rather than diverging baseline expectations. The exact methodology and full results are reproduced in the Supplementary Materials.

## CONCLUSION

Much of the reputation literature focuses on the reputation for resolve and the tendency among policymakers to exacerbate tensions due to reputational concerns. In the context of resolve, reputation is seen as a problem plaguing international politics. It deepens rivalries and prolongs conflicts because leaders are preoccupied with demonstrating their willingness to fight, not only now but also in the foreseeable future. On the other hand, there is much less scholarly focus on reputation for compliance, despite the strong belief in the reputational mechanism serving as a source for good, a commitment device of sorts to international law. Therefore, this study focuses on one particular bilateral relationship—that between the US and South Korea—to examine reputation for compliance and explore whether the public’s concern with reputational consequences can foster greater adherence to international commitments, thus strengthening the international legal order.

In short, this study first finds that noncompliance—compared to compliance—may reduce the public’s first-order and second-order beliefs about future compliance. This

finding supports the claim that the findings of a large body of literature on elite perceptions of reputation costs in the context of reputation for resolve may similarly apply to public perceptions of reputation costs in the context of reputation for compliance.

Second, this study compared first-order and second-order beliefs between the US and South Korea. The US survey suggested that first-order beliefs about compliance could be characterized as having low issue-linkage, meaning there were progressively smaller spillovers across issue areas and across domains. Americans, when informed of South Korean noncompliance in a particular issue area, were less inclined to believe in future South Korean compliance, but they tempered their skepticism if the new issue area under consideration was different from the issue area of the original violation.

On the other hand, there was difference in the degree of issue linkage between second-order beliefs and first-order beliefs. Second-order beliefs were heavily influenced by noncompliance in different issue areas, even though first-order beliefs were not. Although Americans were more forgiving of past noncompliance in different issue areas, South Koreans seemed to believe that their state would incur a reputation cost in different domains. The implication of this finding is that constituents of noncompliant states may mistakenly assume that their state's noncompliance will result in a substantial reputation cost and may demand compliance with existing treaties, fearing first-order reputation costs that are actually unwarranted in reality.

This cross-national discrepancy between first-order beliefs of Americans and second-order beliefs of South Koreans provides promising avenues for future research. To be clear, South Koreans believe that noncompliance will result in reduced US beliefs about future South Korean compliance in different domains. However, Americans are in fact more optimistic about future South Korean compliance in different domains despite South Korea's past noncompliance. A future study could further explore the causes behind this cross-national discrepancy. It is possible that there is a degree of motivated reasoning on the part of Americans. Because all treatments are post-ratification, and because the US is close allies with South Korea, Americans, unlike South Koreans, may be trying to reconcile their existing beliefs about South Korea's longstanding status as a close ally by explaining away the single instance of noncompliance to issue-specific circumstances. In order to resolve the dissonance between what Americans may assume to be true of South Koreans—given the fact that the two states have already signed a treaty—and Korean noncompliance, it appears that Americans may be more willing to forgive South Koreans, and forgive them to a greater extent than South Koreans assume Americans will forgive. To explore this further, one could vary the target state to be a perceived out-group, which would serve as a foil to the perceived in-group South Korea.

The asymmetrical nature of the US–South Korea relationship should also be explored as a potential reason for the cross-national discrepancy. The case selection of the US–South Korean relationship was an opportunity to consider not only how these two specific states perceive of reputation costs but also how states in asymmetrical bilateral relationships perceive of reputation costs. Similar to many other junior partners of the US, South Korea relies on its military alliance with the US to fend off an existential threat, while the US does not. It is possible that because of the asymmetric relationship, the US public was not as concerned with the noncompliance of its junior partner, one of many junior partners, but South Koreans were especially concerned about their primary military ally's perception of their

noncompliance. For South Koreans, the fear of US derogation from a critical military alliance may have exaggerated the perceived reputational consequences of their own noncompliance. A future study might fruitfully explore whether the observed cross-national discrepancy was the result of the asymmetrical nature of the relationship by comparing dyads with relatively equal power and unequal power.

Moreover, a future study could examine discrepancies in first-order and second-order beliefs post-*compliance*, rather than post-*noncompliance*. Does a state's past compliance lead foreign publics to have more optimistic first-order beliefs about the compliant state's future compliance in other issue areas? Meanwhile, does the domestic public believe their state's past compliance with a particular issue area will lead to more optimistic beliefs among foreign publics about the compliant state's future compliance in other issue areas? Overall, the discrepancy in first-order and second-order beliefs found in this study poses important questions to advance the literature on reputation costs.

In closing, by focusing on how the public evaluates the likelihood of future compliance based on past state behavior, the study elaborates on reputation as a possible reason for the public's post-ratification demand for compliance. At a time when activist organizations in the United States are increasingly demanding that their government better comply with existing international legal commitments, this study suggests that the US public's perceived reputation costs and issue-linkages may be a key component driving public support for compliance.

*Matthew Kim* is a student at Harvard Law School and an affiliate at the Institute for Quantitative Social Science. He earned his Ph.D. in political science at Harvard University's Government Department in 2017.

## NOTES

The author would like to thank Beth Simmons, Josh Kertzer, Muhammet Bas, Seth Soderborg, David Romney, Anton Strezhnev, participants of the Harvard International Relations Workshop, and editors of the *Journal of East Asian Studies* for offering helpful critique and advice. The study was funded by the Harvard Institute for Quantitative Social Science and the Harvard Weatherhead Center for International Affairs.

1. Dafoe, Renshon, and Huth (2014) discuss first-order and second-order beliefs about reputation in greater detail. First-order and second-order *beliefs about reputation* for compliance should not be confused with the unrelated concept of first-order and second-order compliance. First-order compliance is "compliance with the substantive provisions of a treaty arrangement," and second-order compliance is "compliance with the authoritative ruling of a third party when the substantive rules are under dispute" (Simmons 2010).

2. Anecdotally, there is at least one case in which an issue area was sufficiently salient to affect the domestic public's second-order beliefs but not sufficiently salient to affect a foreign public's first-order beliefs: An elected official, while contemplating potential noncompliance with an international treaty, faced immense domestic pressure to comply due to the public's inflated second-order beliefs about reputation costs and high issue-linkage. Moreover, the public's fear about the severe and far-reaching reputational consequences seems unfounded in hindsight. Although an in-depth case study is beyond the scope of this study, in 2002 President George W. Bush "unsigned" or nullified America's signature on the Rome Treaty that established the International Criminal Court [ICC]. At the time, Scheffer argued that, "Other governments might take American nullification of its signature on this treaty as an opportunity to unsign other treaties like the Chemical Weapons Convention that are critical in our campaign against terrorism. But an even greater danger is that rejecting the international criminal court outright would relegate America to the backwaters of international justice" (Scheffer 2002). However, after President Bush nullified the signature, US noncompliance did not result in the ICC choosing to "relegate America to the backwaters of international justice" nor did it result in significant international legal repercussions in other domains.

3. Because it was important that respondents not reject the premise of the treatment, the vignette had to depict a hypothetical that was plausible. In reality, many North Korean covert operatives have infiltrated the border into South Korean territory over the past half century. They are occasionally captured by South Korean officers with varying levels of media attention, but it is unclear what happens to these operatives after their capture. Given this reality and based on the analysis of open-ended responses, respondents do not seem to be rejecting the premise of the treatment. The reverse of the experiment in which the US was the non-compliant state rather than South Korea was not possible due to concerns about plausibility. Respondents would have likely rejected a hypothetical that posed the US torturing North Korean operatives.

4. The exact question wording for political ideology was “Which best describes your political leanings?” For knowledge level, I asked the following six questions. (1) “Do you happen to know which of the following countries are members of the United Nations?” (2) “Where is the International Court of Justice located?” (3) “Do you happen to know which US president was the chief advocate for the League of Nations?” (4) “On which of the following does the US federal government currently spend the least?” (5) “Do you happen to know which party has the most members in the House of Representatives in Washington DC?” and (6) “Who has the final responsibility to decide if a law is constitutional or not?” The number of correct responses for each respondent was tallied and weighed equally to create an index of individual knowledge level. For political activism, I asked whether the respondent had ever (1) “worked for a political campaign,” (2) “donated money to a campaign,” (3) “put up a political sign such as a lawn sign or a bumper sticker,” and (4) “attended local political meetings.” The number of positive responses for each respondent was tallied and weighed equally to create an index of individual political activism.

5. For cooperative internationalism, respondents were asked to state the extent to which they agreed with the following three statements: (1) “The US needs to cooperate more with the United Nations,” (2) “It is essential for the US to work with other nations to solve problems such as overpopulation, hunger, and pollution,” and (3) “Helping to improve the standard of living in less developed countries is of utmost importance.” A response was asked for each statement, and the three responses were averaged to create an index for cooperative internationalism. For militant internationalism, respondents were asked to state the extent to which they agreed with the following three statements: (1) “The US should take all steps including the use of force to prevent aggression by any expansionist power,” (2) “Going to war is unfortunate but sometimes the only solution to international problems,” and (3) “American military strength is NOT the best way to ensure world peace.” Again, a response was asked for each statement, and the three responses were averaged to create an index for militant internationalism. For isolationism, respondents were asked to state the extent to which they agreed with the following three statements: (1) “We should not think so much in international terms but concentrate more on our own national problems,” (2) “The US needs to play an active role in solving conflicts around the world,” and (3) “America’s conception of its leadership role in the world must be scaled down.” Again, a response was asked for each statement, and the three responses were averaged to create an index for isolationism.

6. Studies on MTurk can be replicated using nationally representative samples despite the bias in MTurk samples (Berinsky, Huber, and Lenz 2012). Based on the sample characteristics included in the Supplementary Materials, the bias in this study’s US sample is not severe enough to warrant serious concerns about bias. Moreover, weighted least squares regressions using post-stratification weights replicate all of the analyses and have been provided in the Supplementary Materials. Demographics from the 2016 American National Election Studies (ANES) were used to construct the weights. It should be noted that using post-stratification weights can lead to covariate imbalance across the treatment conditions and also complicate the estimation of the sampling variance of the treatment effects. The weighted least squares regression models in the Supplementary Materials should thus be considered only as robustness checks (Franco, Malhotra, Simonovits, and Zigerell 2017; Gelman 2007).

7. The response variable, beliefs about willingness to comply, was converted from the original Likert scale to percentage points, where “strongly agree” on the Likert scale was converted to 1.000, “agree” to 0.833, “somewhat agree” to 0.667, “neither agree nor disagree” to 0.500, “somewhat disagree” to 0.333, “disagree” to 0.167, and “strongly disagree” to 0.000.

## REFERENCES

- Barton, Dona-Gene, Courtney Hillebrecht, and Sergio C. Wals. 2017. “A Neglected Nexus: Human Rights and Public Perceptions.” *Journal of Human Rights* 16 (3): 293–313.

- Berinsky, Adam, Gregory Huber, and Gabriel Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20: 351–368.
- Brutger, Ryan, and Joshua Kertzer. 2018. "A Dispositional Theory of Reputation Costs." *International Organization* 72 (3): 693–724.
- Cardenas, Sonia. 2007. *Conflict and Compliance: State Responses to Human Rights Pressure*. Philadelphia: University of Pennsylvania Press.
- Chilton, Adam. 2014a. "The Influence of International Human Rights Agreements on Public Opinion: An Experimental Study." *Chicago Journal of International Law* 15: 110–137.
- . 2014b. "The Laws of War and Public Opinion: An Experimental Study." Coase-Sandor Institute for Law and Economics, University of Chicago, July 21. [https://chicagounbound.uchicago.edu/law\\_and\\_economics/684/](https://chicagounbound.uchicago.edu/law_and_economics/684/).
- Crow, David. 2017. "Rights Trap or Amplifier? Crime and Attitudes toward Local Human Rights Organizations in Mexico." *Journal of Human Rights* 16 (3): 332–350.
- Dafoe, Allan, and Devin Caughey. 2016. "Honor and War: Southern U.S. Presidents and the Effects of Concern for Reputation." *World Politics* 68 (2): 341–381.
- Dafoe, Allan, Jonathan Renshon, and Paul Huth. 2014. "Reputation and Status and Motives for War." *Annual Review of Political Science* 17: 371–393.
- Dai, Xinyuan. 2005. "Why Comply? The Domestic Constituency Mechanism." *International Organization* 59 (2): 363–398.
- Downs, George, David Rocke, and Peter Barsoom. 1996. "Is the Good News about Compliance Good News about Cooperation?" *International Organization* 50 (3): 379–406.
- Franco, Annie, Neil Malhotra, Gabor Simonovits, and L. J. Zigerell. 2017. "Developing Standards for Post-Hoc Weighting in Population-Based Survey Experiments." *Journal of Experimental Political Science* 4 (2): 161–172.
- Gelman, Andrew. 2007. "Struggles with Survey Weighting and Regression Modeling." *Statistical Science* 22 (2): 153–164.
- Gerber, Theodore. 2017. "Public Opinion on Human Rights in Putin-Era Russia: Continuities, Changes, and Sources of Variation." *Journal of Human Rights* 16 (3): 314–331.
- Guzman, Andrew. 2008. *How International Law Works: A Rational Choice Theory*. New York: Oxford University Press.
- Hafner-Burton, Emilie, and Kiyoteru Tsutsui. 2005. "Human Rights in a Globalizing World: The Paradox of Empty Promises." *American Journal of Sociology* 110 (5): 1373–1411.
- Hopf, Theodore. 1994. *Peripheral Visions: Deterrence Theory and American Foreign Policy in the Third World, 1965–1990*. Ann Arbor: University of Michigan Press.
- Huth, Paul. 1997. "Reputation and Deterrence: A Theoretical and Empirical Assessment." *Security Studies* 7 (1): 72–99.
- Kagan, Donald. 1996. *On the Origins of War and the Preservation of Peace*. New York: Doubleday.
- Kertzer, Joshua. 2016. *Resolve in International Politics*. Princeton: Princeton University Press.
- Kertzer, Joshua, Kathleen Powers, Brian Rathbun, and Ravi Iyer. 2014. "Moral Support: How Moral Values Shape Foreign Policy Attitudes." *Journal of Politics* 76 (3): 825–840.
- Koo, Jeong-Woo. 2017. "The Construction of Human Rights Actorhood: Findings from the Korean General Social Survey." *Journal of Human Rights* 16 (3): 261–275.
- Kreps, Sarah, and Geoffrey Wallace. 2016. "International Law, Elites, and Public Support for Drone Strikes." *Journal of Peace Research* 53 (6): 830–844.
- Lebow, Richard. 2008. *Why Nations Fight: Past and Future Motives for War*. Cambridge, UK: Cambridge University Press.
- Lipson, Charles. 2003. "Why Are Some International Agreements Informal?" *International Organization* 45 (4): 495–538.
- Mercer, Jonathan. 1996. *Reputation and International Politics*. Ithaca, NY: Cornell University Press.
- Moravcsik, Andrew. 2005. *The Paradox of U.S. Human Rights Policy*. Princeton: Princeton University Press.
- Morrow, James. 2014. *Order within Anarchy: The Laws of War as an International Institution*. Cambridge, UK: Cambridge University Press.
- Murdie, Amanda, and Carolin Purser. 2017. "How Protest Affects Opinions of Peaceful Demonstration and Expression of Rights." *Journal of Human Rights* 16 (3): 351–369.

- Mutz, Diana, and Robin Pemantle. 2015. "Standards for Experimental Research: Encouraging a Better Understanding of Experimental Methods." *Journal of Experimental Political Science* 2 (2): 192–215.
- Neumayer, Eric. 2005. "Do International Human Rights Treaties Improve Respect for Human Rights?" *Journal of Conflict Resolution* 49 (6): 925–953.
- O'Neill, Barry. 1996. *Honor, Symbols, and War*. Ann Arbor: University of Michigan Press.
- . 2006. "Nuclear Weapons and National Prestige." Cowles Foundation Discussion Paper, Yale University, February. <https://cowles.yale.edu/sites/default/files/files/pub/d15/d1560.pdf>.
- Pandya, Archana, and James Ron. 2017. "Local Resources for Local Rights? The Mumbai Fundraiser's Dilemma." *Journal of Human Rights* 16 (3): 370–387.
- Powell, Emilia, and Jeffrey Staton. 2009. "Domestic Judicial Institutions and Human Rights Treaty Violation." *International Studies Quarterly* 53 (1): 149–174.
- Press, Daryl. 2005. *Calculating Credibility: How Leaders Assess Military Threats*. Ithaca, NY: Cornell University Press.
- Pruitt, Autry. 2016. *Planes, Steak & Water: Defending Donald J. Trump*. Austin, TX: Gatekeeper Press.
- Putnam, Tonya, and Jacob Shapiro. 2017. "International Law and Voter Preferences: the Case of Foreign Human Rights Violations." *Human Rights Review* 18 (3): 243–262.
- Ron, James. 2017. Introduction to Special Issue on "Public Opinion Polling & Human Rights." *Journal of Human Rights* 16 (3): 257–260.
- Scheffer, David. 2002. "A Treaty Bush Shouldn't 'Unsign,'" *New York Times*, April 6, 2002.
- Schelling, Thomas. 1966. *Arms and Influence*. New Haven: Yale University Press.
- Simmons, Beth. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge, UK: Cambridge University Press.
- . 2010. "Treaty Compliance and Violation." *Annual Review of Political Science* 13 (1): 273–296.
- Simmons, Beth, and Daniel Hopkins. 2005. "The Constraining Power of International Treaties: Theory and Methods." *American Political Science Review* 99 (4): 623–631.
- Smith-Cannoy, Heather. 2012. *Insincere Commitments: Human Rights Treaties, Abusive States, and Citizen Activism*. Washington, DC: Georgetown University Press.
- Snyder, Glenn, and Paul Diesing. 1977. *Conflict Among Nations: Bargaining, Decision Making, and System Structure in International Crises*. Princeton: Princeton University Press.
- Snyder, Jack, and Borghard, Erica. 2011. "The Cost of Empty Threats: A Penny, Not a Pound." *American Political Science Review* 105 (3): 437–456.
- Tang, Shipping. 2005. "Reputation, Cult of Reputation, and International Conflict." *Security Studies* 14 (1): 34–62.
- Tomz, Michael. 2008. "Reputation and the Effect of International Law on Preferences and Beliefs." Department of Political Science, Stanford University, February. <https://web.stanford.edu/~tomz/working/Tomz-IntlLaw-2008-02-11a.pdf>.
- Valentino, Benjamin, and Ethan Weinberg. 2017. "More than Words? 'Genocide', Holocaust Analogies and Public Opinion in the United States." *Journal of Human Rights* 16 (3): 276–292.
- Von Stein, Jana. 2005. "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance." *American Political Science Review* 99 (4): 611–622.
- Wallace, Geoffrey. 2013. "International Law and Public Attitudes Toward Torture: An Experimental Study." *International Organization* 67 (1): 105–140.
- . 2014. "Martial Law? Military Experience, International Law, and Support for Torture." *International Studies Quarterly* 58 (3): 501–514.
- Walt, Stephen. 2015. "The Credibility Addiction." *Foreign Policy*, January 6. <https://foreignpolicy.com/2015/01/06/the-credibility-addiction-us-iraq-afghanistan-unwinnable-war/>.
- Weisiger, Alex, and Keren Yarhi-Milo. 2015. "Revisiting Reputation: How Past Actions Matter in International Relations." *International Organization* 69 (2): 473–495.