

# No False Promises: How the Prospect of Non-Compliance Affects Elite Preferences for International Cooperation

## RESEARCH NOTE

EMILIE M. HAFNER-BURTON

*University of California, San Diego*

BRAD L. LEVECK

*University of California, Merced*

AND

DAVID G. VICTOR

*University of California, San Diego*

Why would leaders engage in international cooperation if they believe that their *own* government might default from their commitments? Some suggest that when leaders do so, they are essentially trying to profit from false promises—from making international commitments that they likely cannot, or will not, actually fulfill. In contrast, others expect that fears of such non-compliance will deter leaders from engaging in international cooperation. Moreover, some theories suggest that the design of cooperative agreements themselves should affect how leaders respond to these possibilities. That is, leaders should be more concerned about the prospect of their country's non-compliance with agreements that impose, through formal means, sizeable costs on recalcitrant states. We describe the results of an experimental survey conducted on 95 high-level policy elites in the United States that allows us to examine the causal dynamics that underlie this debate. We focus on one key institutional design feature—formal enforcement—and preferences for international cooperation under different perceptions of risk about future compliance. We provide the first elite-level evidence that, as the prospect of defection rises, actual policymakers become less willing to join international agreements. However, contrary to what many theories of international institutions would predict, the presence of a formal enforcement mechanism fails to explain their aversion to cooperation. Elites dislike making false promises even when their commitments are not formally enforceable. By measuring these elites' patience (along with other traits), we tentatively suggest that this aversion may be linked to decision-makers' own perceptions of the future—elites who have lower discount rates are particularly sensitive to the prospect of not honoring commitments.

Almost all scholars agree that strategic uncertainty—about other states' intentions and payoffs—shapes decisions to cooperate internationally. But we know very little about whether leaders will cooperate when confronted with the prospect that their *own* government will default on an international commitment.<sup>1</sup> Some theories suggest that leaders will be untroubled by the prospect of making false promises by joining international commitments under

these circumstances. From that perspective, such promises are a form of cheap talk—meaning the costs of defection, if any, are too small to concern the leader making them. We find examples of such false promises in many issue areas, including human rights, climate change, security, and trade policy.<sup>2</sup> Other theories point in opposite directions. From those rival perspectives, leaders will seldom willfully disregard existing commitments, even if the formal costs of violation are small. Leaders will not make new international commitments unless they believe that their government can live up to them.

This divergence in theoretical predictions is rooted in claims about how leaders think about the benefits and costs of false promises. One logic emphasizes that leaders prefer to avoid costly punishments that an international institution might mete out in response to non-compliance. The long tradition of research on the rational design of international institutions—and on the rational behavior by states in response to their institutional environments—suggests that government leaders who are uncertain about their country's future compliance may avoid making demanding commitments. It also suggests that aversion to such commitments should rise in the presence of mechanisms that can credibly enforce

---

Emilie M. Hafner-Burton is the John D. and Catherine T. MacArthur Professor of International Justice and Human Rights at the School of Global Policy and Strategy at the University of California at San Diego and co-director of the Laboratory on International Law and Regulation.

Brad L. LeVeck is an assistant professor at the University of California—Merced.

David G. Victor is a professor at the School of Global Policy and Strategy at the University of California—San Diego and co-director of the Laboratory on International Law and Regulation. Prior to UC—San Diego, Victor directed the Program on Energy and Sustainable Development at Stanford.

*Authors' note:* We thank Burcu Bayram, Adam Chilton, Todd Hall, Joshua Kertzer, Andy Kydd, Rick Herrmann, Megumi Naoi, Paul Poast, Janice Stein, Dustin Tingley and Mike Tomz for comments on an earlier draft, as well as participants in the American Political Science Association annual meeting, 2013, and the Behavioral Revolution in International Relations workshop at UC San Diego, 2014. The Laboratory on International Law and Regulation (ILAR) is funded by the School of International Relations and Pacific Studies (IR/PS) at UC San Diego, BP plc, the Electric Power Research Institute and the Norwegian Research Foundation.

<sup>1</sup>On risk-taking in international relations more generally, see McDermott (1998). Other studies look at how uncertainty shapes the rational design of international institutions (see Koremenos 2005; Rosendorff and Milner 2001).

<sup>2</sup>See Hathaway (2005) on human rights, Victor (2001) on climate, and Mercer (2012), Jervis (1989), and Hafner-Burton (2005) on trade.

Hafner-Burton, Emilie M., Brad L. LeVeck, and David G. Victor. (2017) No False Promises: How the Prospect of Non-Compliance Affects Elite Preferences for International Cooperation. *International Studies Quarterly*, doi: 10.1093/isq/sqw047

© The Author (2017). Published by Oxford University Press on behalf of the International Studies Association.

All rights reserved. For permissions, please e-mail: journals.permissions@oup.com  
Downloaded from <https://academic.oup.com/isq/article-abstract/61/1/136/3837134>  
by University of California, San Diego Libraries user  
on 02 November 2017

international obligations (Downs and Rocke 1995; Downs, Rocke, and Barsoom 1996; Koremenos 2005; Rosendorff and Milner 2001; von Stein 2005).<sup>3</sup> By this logic, leaders who know their country has a substantial prospect of default will not commit to agreements that have costly enforcement procedures. Enforcement will deter cooperation altogether, or it will lead policymakers to seek more flexible (and perhaps less demanding) forms of cooperation.<sup>4</sup>

Alternatively, decision-makers may adhere to international agreements for other reasons: a sense of obligation, a concern for reputation, or fear of informal retaliation (Chayes and Chayes 1993; Finnemore and Toope 2001; Franck 1990; Guzman 2008; Thompson 2009; Tomz 2007). By these logics, formal enforcement is beside the point. Instead, compliance flows from the knowledge that interactions will be repeated and that adherence to norms, even in the absence of formal institutionalized enforcement mechanisms, can be an optimal strategy for states and their leaders (Axelrod 1984; Axelrod and Keohane 1985; Tingley 2011).<sup>5</sup> Indeed, the question of why states join and comply with international obligations lies at the heart of one of the great debates in international relations over the past two decades (Downs, Rocke, and Barsoom 1996; Simmons 2000; Martin 2013). It is one of the areas where in addition to debates within political science the field has also engaged scholars in international law (Raustiala and Slaughter 2002; Hafner-Burton, Victor, and Lupu 2012).

The absence of reliable information about how real policy leaders actually evaluate international commitments in the context of risk and uncertainty makes it difficult to adjudicate these debates. This article is the first, to our knowledge, to offer that information in an experimental setting where it is possible to eliminate or control many confounding factors and selection effects (Martin 2013). Such an experimental approach confers important advantages. In the real world, we find huge variations across issue areas and commitments. Moreover, the real world is unlikely to offer opportunities to observe the effects of an agreement's content *independent* of its design; there are strong reasons to expect that the presence or absence of enforcement is endogenous to the interests and expectations of decision-makers (Fearon 1998, 275; Downs, Rocke, and Barsoom 1996).<sup>6</sup>

We can resolve these difficulties by systematically and independently varying both, first, the perception that a country might not comply and, second, whether an agreement has a formal enforcement mechanism.<sup>7</sup> That approach

<sup>3</sup>A similar logic also applies when treaties are inflexible about the legal commitment they require states to implement (Rosendorff and Milner 2001).

<sup>4</sup>On the broader-deeper trade-off more generally, see Gilligan (2004).

<sup>5</sup>For an alternative view that argues that a longer shadow of the future may exacerbate commitment problems, see Tingley (2011).

<sup>6</sup>Even if there is a positive association between formal enforcement and the depth of cooperation—as some theories predict—there are several possible explanations for that association. When the stakes of cooperation are relatively low, states might appear more willing to both sign and violate agreements without a formal enforcement mechanism. Yet states' willingness to sign and abrogate these agreements might simply reflect the low-stakes nature of the agreement, rather than the presence or absence of enforcement. Because we cannot estimate a true causal effect without random assignment of enforcement, we are limited in our ability to draw inference from observed behavior.

<sup>7</sup>A related concern, present in many survey experiments, is that treating subjects with the presence of an enforcement mechanism could also change their beliefs about other factors that are empirically correlated with enforcement, such as the content of the treaty or the nature/depth of participants' interests. Our experiment was designed to control subjects' beliefs about these variables by directly specifying the content and scope of the agreement, as well as how countries and their citizens might benefit.

allows us to examine the causal link between the presence of institutional enforcement and decision-maker preferences for international cooperation, as they both vary independently from one another.<sup>8</sup> This enables us to determine both whether compliance uncertainty deters leaders from making false promises, and whether this potential aversion effect varies with the design of the institution in which the commitment is made. In our survey experiment, we use a scenario that invokes international commitments on trade. However, the structure of our scenario mimics a wide array of other international commitments. For example, it includes controlling emissions that cause climate change and the adoption of technology standards that affect market access—where joining and complying are strategic complements (Peinhardt and Sandler 2015). We survey a sample of 95 actual US policy elites who have on average 22 years' work experience in the parts of government—including former members of Congress as well as senior officials in Treasury and the US Trade Representative's office—business, and non-governmental interest groups that are directly implicated by international trade agreements and centrally involved in lobbying and policy decision-making.<sup>9</sup>

Sampling actual elites matters for this study because they differ in some key ways from non-elites. For example, elite populations tend to be more strategic, patient, and fair than college students and others used in the convenience samples typical of experimental research (LeVeck et al. 2014). They also tend to be more cooperative and there is some evidence that elites differ in other ways, such as being more likely to update decision-making heuristics in light of new information (Hafner-Burton, Hughes, and Victor 2013).

Our experiment demonstrates that US elites are significantly less willing to make commitments when they face the prospect of future default—operationalized here as a 50 percent chance of future defection. However—contrary to the expectations of many theories of international cooperation—the presence of formal institutional enforcement intended to impose immediate costs on defectors does not dampen this aversion. Elites in our sample prefer to avoid false promises even when agreements are not formally enforceable.

Relevant to our central finding—that elites prefer not to make false promises even when agreements are formally unenforceable—is the fact that the individuals who make choices about institutional cooperation vary in how they react to the possible consequences of non-compliance (McDermott and Herrera 2010).<sup>10</sup> Governments do not think; people do (McDermott 2004; McDermott and Herrera 2010; Hafner-Burton et al. Forthcoming). A long list of factors influences individual preferences and choices (DellaVigna 2009), but the formal literature on cooperation has long understood that a central factor in determining whether repeated interaction will yield cooperation is the weight that decision-makers place on future streams of costs and benefits. Decision-makers have widely variable discount rates—which we (and others) refer to as “patience” (Coller and Williams 1999; Fowler and Kam 2006). Patience, we suggest, affects cooperation in at least two ways: first, it can alter the incentives for reciprocity in repeated games; second, it can affect the weight that decision-makers

<sup>8</sup>Among the growing number of studies within international relations that also use experiments are Chilton and Tingley (2013), McDermott (2011), Mintz, Yang, and McDermott (2011), and Tingley (2011).

<sup>9</sup>Among the growing number of studies within international relations that also use elite samples are Milner and Tingley (2013b), Mintz, Redd, and Vedlitz (2006), Tetlock (2005), and Tomz (2009). For a review of elite experiments and differences from non-elite populations, see Hafner-Burton, Hughes, and Victor (2013).

<sup>10</sup>On individual heterogeneity, see Kertzer (2013), Martin (1994), and Renshon, Lee, and Tingley (2015).

place on reputation. Such mechanisms can operate even if agreements do not themselves contain formal enforcement mechanisms.

Because we measure the relative patience of each subject, our survey provides suggestive evidence on the possible role of this trait in how policy elites weigh proximate and distant costs of non-compliance. Patient decision-makers—with low discount rates—are more cautious about cooperation in the face of their own country's possible future default than are decision-makers who have higher discount rates and place a heavier emphasis on more proximate costs and benefits. Patient decision-makers anticipate the long-term effects of their country's broken promise.

The article proceeds, first, by defining how we measure the risk of default. We then develop the different theoretical explanations for how the prospect of default might—or might not—influence decision-maker preferences for international cooperation given variation in the design of the commitment. We test those theoretical expectations with our elite survey experiment. Armed with the results from the experiment, we explore the logics that may explain why decision-makers avoid making false promises even when defection cannot be formally enforced. Our explanation explores the role of heterogeneity in decision-maker time preferences while also suggesting other directions for future research. We conclude with broader implications for theories of cooperation.

### False Promises

One of the most enduring problems in politics is that elite decision-makers often must make important decisions under conditions of risk and uncertainty (Edelstein 2002; Jarvis 2011). There are multiple types of risks that shape those decisions (Rathbun 2007). The standard way scholars think about risk and uncertainty, common to international bargaining games of incomplete information, is strategic: actors know their own intentions and payoffs but they are not fully informed about each other's and worry about the risk of defection by other parties (Koremenos, Lipson, and Snidal 2001). Actors may not possess full information about their own payoffs from a bargain because they are unsure about their own future interests or capabilities (Iida 1993). What we call “non-compliance risk” is one example.<sup>11</sup> Leaders negotiate international agreements that best reflect their national interests, but what happens after those agreements are made can be unpredictable and sometimes beyond the control of the state (Koremenos 2005). Shocks brought on by events such as natural disasters, financial crises, or unexpected political events can mobilize domestic interest groups to push a country out of compliance with their previously made international commitments (see for example Downs and Rocke 1995; Rosendorff and Milner 2001). Unexpected changes in technology, political attitudes, or other factors can also reduce the capacity of governments to fulfill their commitments (Chayes and Chayes 1993). In addition to these cases of inadvertent non-compliance, leaders may also willfully not intend to comply. And in the real world, the line between unintentional and purposeful non-compliance may be very blurry to such a point that the formal distinction of cause is not important.

More formally, we define the false promise risk as the probability that a leader's country will not comply with a

<sup>11</sup>Kertzer and McGraw (2012) have used experiments to examine other forms of uncertainty (such as strategic uncertainty) in non-elite populations.

particular agreement. We emphasize that our experiment focuses on what happens when leaders face an *increase* in their country's risk, and therefore become *less* certain about whether their country will comply with an agreement.<sup>12</sup> Our aim is to study how elite decision-makers in the setting of one of the world's most powerful nations evaluate the consequences of potential non-compliance. Do they fear the costs of non-compliance enough to avoid making commitments in the first place? Do they see those costs as arising from formal enforcement mechanisms that are embedded in the treaty itself or from some other quarter?

### Costs of Non-Compliance

Inherent in different theoretical views about what role international institutions play in international cooperation are differing views about how government decision-makers will weight the risk of non-compliance with commitments made in the process of institutionalized cooperation. How costly is a false promise when compared to the potential benefits of renegeing?

From one perspective, it may seem perfectly logical for leaders to make false commitments—that is, to promise to cooperate knowing full well that there is a decent probability that their country will not comply with the agreement. This type of behavior is familiar in the area of human rights, where governments frequently make international commitments to agreements that they do not—or cannot—honor, in part because human rights institutions are themselves weakly enforceable by law and in part because the international community is reticent to take any credible, or costly, retaliatory actions (Goldsmith and Posner 2005; Hathaway 2005). One example is the Russian Federation, whose government has made legally binding commitments to eleven (out of eighteen) global human rights treaties, many during the era of Communism, which it overtly violates (United Nations 2016). These false commitments, however, are hardly unique to human rights.

False promises are also replete in the history of other areas, including both climate and trade. The United States negotiated and signed the Kyoto Protocol in 1997 despite the lack of any politically viable strategy to honor the deep cuts in emissions that the Protocol would have required (Depledge 2005; Hovi, Sprinz, and Bang 2012). The European Union negotiated and joined the World Trade Organization (WTO), including strict new standards for food safety regulations, even though its existing ban on imported beef produced with hormones likely violated the new WTO strictures—a finding confirmed in one of the WTO's very first high-profile disputes (WTO 1998). Important exporting countries like Bangladesh have also made false promises—to protect Bangladeshi workers in exchange for receiving Western trade privileges (USTR 2015).<sup>13</sup>

From this perspective:

**H1a:** *Increasing a state's false promise risk should have little effect on the willingness of that state's decision-makers to join international agreements.*

<sup>12</sup>Future work may want to look at what happens when states are initially certain that their country will not comply, but later become less certain of this fact.

<sup>13</sup>In 2013, for example, President Obama suspended Bangladesh from America's Generalized System of Preferences trade program in response to Bangladesh's repeated failure to meet statutory eligibility requirements related to worker rights.

Yet from several other vantage points, the prospect of default may affect preferences for cooperation for a variety of different reasons. While strategic uncertainty may drive actors to create enforceable international agreements in an effort to reduce the incentives for others to defect, uncertainty about one's own compliance prospects may do the opposite: "states may not even commit themselves to an agreement if they anticipate that circumstances will alter their expected benefits" (Koremenos 2005, 549). Moreover, if "treaties are commitment devices, then they should in fact have a screening effect, because only those governments that are willing and think they will be able to comply should sign on" (Simmons and Hopkins 2005, 624).

From this perspective:

**H1b:** *Increasing a state's false promise risk should dampen the willingness of that state's decision-makers to join international agreements.*

If we find a wariness to join commitments when there is a risk of false promise, what might explain that outcome? One explanation, rooted in theories of bargaining and institutional design, is that institutions can be designed to impose official costs on violators of an agreement. Formal enforcement mechanisms, notably, are intended to play this role. If all states would comply with obligations automatically—or if agreements were self-enforcing—then there would be no need for institutional enforcement. But when states engage in cooperation where commitments are difficult to implement, compliance is often far from automatic. Formal enforcement mechanisms—like the dispute settlement mechanism of the WTO—offer the prospect of imposing relatively swift and significant costs on defectors and thus may deter membership by countries that are not sure they can comply (Downs, Rocke, and Barsoom 1996; von Stein 2005). The WTO's dispute system, for example, follows strict timetables for filing and responding to cases, hearing evidence, decisions by panels, final decisions by the Organization's Appellate Body, and authorization of retaliatory actions against parties that don't comply (WTO 1994).

Yet the impact of formal enforcement on incentives to cooperate is a source of debate because other forces may also be at work (Thompson 2009). Scholars have long studied reciprocity that occurs between actors who repeatedly interact over long periods of time, and who may credibly threaten to stop cooperating if another actor defects (Fearon 1998; Greif, Milgrom, and Weingast 1994; Guisinger and Smith 2002). In such models, retaliation is credible because it is an equilibrium strategy of a repeated game (Bendor and Swistak 1997), a finding that holds even if retaliation is not specified as part of a formal agreement.

Reputation, as well, may influence the incentives to cooperate. Indeed, there has been a resurgence of scholarship on reputation, and one of the central themes in that work is how reputation intersects with reciprocity.<sup>14</sup> Michael Tomz (2007) has shown that governments with poor reputations may face higher borrowing costs and lenders use contracts to manage their exposure to governments with poor reputations. In the extreme, governments with poor reputations are shunned or deterred from participation in demanding international agreements (Tomz 2007). Beth Simmons (2000) has shown how IMF obligations raise expectations that create reputational costs through market

forces for governments that violate their commitments. Brad LeVeck and Neil Narang (Forthcoming) have shown that states pay a reputational cost for defaulting on their alliance obligations, as defaulting states have a more difficult time forming future alliances. More generally, Andrew Guzman (2008) has suggested that governments that interact repeatedly care about their reputation for compliance to such a degree that they will comply with international obligations even in the absence of effective formal enforcement mechanisms. Others have even argued that rational citizens might anticipate international punishment for broken promises and remove leaders who violate the terms of agreements they sign (Guisinger and Smith 2002). Indeed, when leaders go to major conferences where they adopt formal agreements, they often rhetorically focus on the credibility of their promises—something that was on full display at the Paris climate change conference in late 2015, for example, where leaders emphasized the credibility of their promises in a bid to encourage deeper cooperation (US White House 2015).

Concern about reputation also resonates with several lines of scholarship that emphasize lawmaking as a process that works through the establishment of norms, practices, and legitimacy (see for example Brunnée and Toope 2010; Finnemore and Toope 2001; Franck 1990; Koh 1997). Law works, even in the absence of formal enforcement, because good-standing membership in treaties has become expected behavior for modern nations (Chayes and Chayes 1998).

Thus we have contrasting models of whether—and why—governments avoid making international commitments in the face of non-compliance risk. One model emphasizes that international commitments are relatively cheap talk, little encumbered by enforcement, easy to break with impunity, or simply worth the low costs of defection. In that model, defection is economical and therefore the risk of non-compliance is not a deterrent to making commitments. Another model emphasizes just the opposite: that making false promises is costly. However, that model leads to different perspectives on why leaders avoid false promises. One perspective focuses on the design of institutions and suggests that formal enforcement mechanisms may deter countries from joining cooperative agreements because non-compliance would lead to costly, official punishments. Another emphasizes concerns for the future that arise from reputation, long-term reciprocity between states, a sense of obligation, or fear of informal tit-for-tat retaliation rather than formal enforcement by the institution.

Our contribution to this discussion focuses on how elite decision-makers in the United States assess the trade-offs between membership in an agreement and the potential consequences of non-compliance. Do they knowingly make false promises, or do they balk at the idea of international cooperation when they believe the United States may not comply? Are they deterred by the prospect that the United States may be punished for non-compliance formally by the institution or informally by some other form of informal retaliation? We state the hypothesis in its simplest form:

**H2:** *Formal enforcement mechanisms are a screen that dampens decision-makers' willingness to join cooperative agreements when their state faces a higher level of non-compliance risk.*

Failure to falsify the hypothesis would provide evidence that decision-makers operating under a high level of non-compliance risk are deterred from cooperation out of fear

<sup>14</sup>For a more skeptical view, see Downs and Jones (2002).

of institutional enforcement—they prefer to select out of agreements to avoid formal retaliation by the institution. If Hypothesis 2 is falsified but decision-makers are still choosing not to join agreements in the face of non-compliance risk, then other, we assume longer-run non-institutional factors—such as reputation, normative obligation, or informal retaliation—must be at work.

Both of our hypotheses are about the preferences of individual decision-makers that are plausibly related to state policy choices. We thus follow an already-large and growing literature that seeks to determine sources in the variation of individual preferences for international cooperation (Mansfield and Mutz 2009; Milner and Tingley 2013a, 2013b; Rho and Tomz 2013, 2015; Tingley and Tomz 2012). In the real world, state policy of course reflects many other factors, such as how individual preferences aggregate up into collective decisions. By focusing on elites at the “top” of that policy process, however, we may substantially reduce that aggregation problem (Hafner-Burton et al. forthcoming).

### The Experiment

To test these hypotheses about the effect of increasing non-compliance risk on decision-maker willingness to cooperate, we recruited a unique sample of 95 American policy elites, including former members of the US Congress, their senior staff, top US trade and economic policy negotiators, senior executives in firms whose operations are exposed to international trade, NGOs centrally involved in economic policy decisions, and civil servants in federal and state government. Recruiting was done through the professional networks of decision-makers who held senior positions in government or were senior policy advisers to government officials as well as through conferences that involved senior policymaking, business, and NGO officials. A demographic profile of the sample is reported in the [Supplementary Information \(SI\)](#). The invitations were worded to screen for individuals who might plausibly find themselves in the position to sign an international trade agreement, advise a member of Congress to vote for ratifying and implementing legislation, or who would be highly familiar with the process. Subjects were informed that we had recruited them for their experience in the area of trade and trade-related economic policy; they were also informed that the broad purpose of the study was to examine decision-making. However, they were not informed about any specific hypothesis.

We asked these individuals to participate in a survey experiment designed to test whether, independent of concerns about external enforcement, elite decision-makers would be less likely to sign a trade agreement after receiving information that their country might not be able to comply with the agreement’s terms. The experiment, pictured in [Figure 1](#), first presented subjects with a vignette describing an international trade agreement. In our initial description, we explicitly framed the decision to join the treaty as a situation where joining is a strategic complement. That is, states benefit more from joining and abiding by the agreement if other countries do the same. In this type of strategic situation, conditional cooperation can be an equilibrium strategy even without enforcement, though enforcement makes it more certain that no state has a dominant strategy to free ride. Consistent with this logic, other studies have reported that the presence of an enforcement mechanism does in fact make subjects

initially more likely to join the agreement (Hafner-Burton et al. 2014).

Within the vignette, we randomly<sup>15</sup> varied whether or not the agreement was described as containing a formal enforcement mechanism.<sup>16</sup> This allowed us to consider whether the threat of formal institutional enforcement would further deter leaders from signing an agreement above and beyond concerns that always lurk in the background of international relations, such as negative reciprocity or reputational loss.<sup>17</sup> Specifically subjects were either informed that:

“An independent enforcement mechanism *promptly and credibly* punishes any country that does not comply by taking away some of the benefits of the treaty from the country that breaks the rules.”

Or that:

“The treaty does not provide any formal mechanism to punish countries that fail to comply.”

We intentionally designed this treatment to reflect a maximally plausible level of institutional enforcement. To do so, we beta tested three versions of the treatment on several elite decision-makers directly involved in US trade decisions.<sup>18</sup> They judged the other two versions—one with automatic enforcement, the other with additional penalties such as financial transfers—implausible in any foreseeable real world of international trade law. We note that this experiment is probably a “best case” for observing the effects of prompt institutional enforcement since this is the domain of international relations where commitments are often deep and enforcement is most elaborate and credible. Existing trade law enforcement mechanisms are governed by strict timetables for hearing and resolving disputes, and thus for informed, elite populations it is plausible that these mechanisms are viewed as prompt and swift. By design, our survey focused on formal enforcement mechanisms and explicitly did not refer to other mechanisms—such as derogations, unilateral reinterpretations, withdrawal, or other flexibility instruments—as a response to non-compliance.

After presenting subjects with the vignette, we measured the likelihood that they would sign such an agreement.<sup>19</sup> This allowed us to control for any factors that might cause a subject to initially favor or oppose signing such an agreement. We then presented the subjects with information indicating that there was now a 50 percent chance that their country would not be able to meet their

<sup>15</sup>A balance check is reported in the [Supplementary Information](#) for key covariates.

<sup>16</sup>The full text of this vignette can be found in the [Supplementary Information](#).

<sup>17</sup>To gain experimental control over these factors, we primed subjects to think about issues of reputational loss and negative reciprocity in both the enforcement and non-enforcement condition. This allowed us to rule out the possibility that subjects vary substantially in whether they even considered such a threat, and focus more on whether different types of subjects were more and less impacted by such factors across enforcement conditions.

<sup>18</sup>Those decision-makers did not take the subsequent survey.

<sup>19</sup>Respondents picked one of five responses, each of which corresponded to a probability interval. For example, choosing the lowest category meant that the respondent thought there was a chance between 0% and 20% that they would sign the treaty.

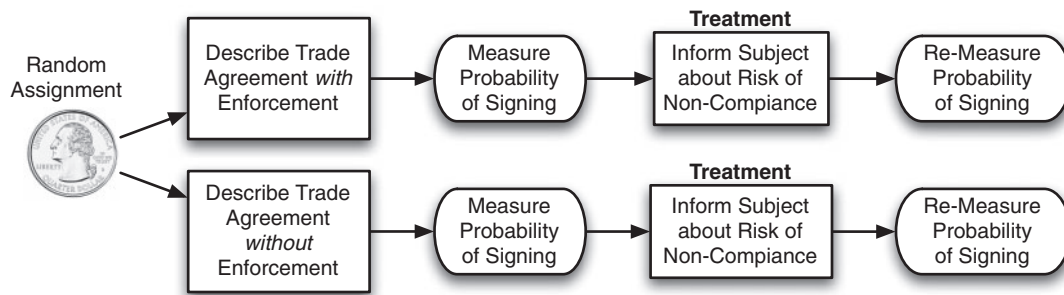


Figure 1. Survey experiment design

obligations under the agreement, and re-measured the likelihood that the subject would sign such an agreement. This is a very large risk of non-compliance, but was deemed by elites we interviewed (but did not use in our study) during design of the instrument to be within the realm of possibility. We made the risk large primarily because we had a necessarily small sample of respondents, and because the ordered response scales used in most survey experiments (including ours) are known to have unavoidable amounts of measurement noise, which can dampen estimates of the true effect (Achen 1975; Ansolabehere, Rodden, and Snyder 2008). A large “dosage” was therefore necessary to help ensure our ability to detect an effect if it did exist.

Ideally, we would have also measured subjects’ baseline uncertainty about whether their country would comply with the agreement. Without this measurement, our treatment effect for non-compliance risk must be interpreted as the effect of moving from an unknown (but presumably small) level of perceived risk to a very large, 50 percent, chance of non-compliance. As a consequence, our findings tend to understate the true effect size of non-compliance risk. However, there are three reasons to believe that the magnitude of this bias is quite small. First, nothing in the initial prompt’s language suggested to subjects that their own country’s compliance was in doubt. Second, if subjects thought there was an initially high risk of non-compliance by their country, we should see little effect. As we describe below, the effect of non-compliance risk is substantial. Third, none of the free-response comments given by subjects (listed in the SI) suggest that subjects were highly uncertain about compliance prior to the treatment.

After telling subjects that their country might not comply, we primed them for the possibility of reputational loss by noting that a failure to comply could reduce their country’s credibility. Issues of non-formal enforcement, such as the risk of reputational loss, always lurk in the background of international relations. Priming for this possibility gave us better experimental control by ensuring that variance in subjects’ responses was not driven by variation in whether subjects considered this ever-present possibility. However, subjects were still free to vary in whether they thought factors such as reputational loss were important, and informal comments (listed in the SI) show considerable heterogeneity in views.

In the enforcement condition, we also primed subjects to think about the consequences of the enforcement mechanism being invoked (to further make it salient). A potential concern is that priming subjects to think about their country’s credibility might account for why formal

enforcement has no effect. However, the credibility prime was presented in both enforcement and non-enforcement conditions. Therefore, if the formal mechanism had any additional effect over concerns of credibility, we would see an effect, which we do not. Additionally, our description of the enforcement mechanism was the only text that was bolded—if anything, subjects’ attention should have been more biased toward concerns about enforcement.

Another concern is that, by using a within-subjects research design, we trigger desirability bias in our subjects. Elites might want to portray themselves as fair-minded, honest brokers unwilling to deal in false promises. We tried to mitigate the risk of such biases by emphasizing at the beginning of our study that subjects would remain anonymous, and that we would not release any identifying information about them. We also, truthfully, told subjects that identifying information would be stripped from the dataset before analyses began. It is also hard to imagine a case in which individual elites are worried about being perceived as dishonest by the experimenters, but do not have similar concern for their future reputation when negotiating international agreements.

This research design allows us to determine whether policymakers who face a dramatic increase in their country’s prospect of non-compliance will avoid joining agreements that require deep cooperation if the treaty is formally and swiftly enforceable, and thus adjudicate between opposing views about how legal institutions function in the presence of non-compliance risk. The research design also allows us to compare the effects that formal enforcement might have on decision-makers’ preferences for cooperation with longer-term consequences that are not formally specified in the agreement itself, such as concern for long-term reputation, reciprocity, or adherence to cooperative norms.

#### *Non-Compliance Risk Deters Cooperation*

Our subjects were less likely to prefer to join the agreement once given information that there was a 50 percent chance that their country would default on the treaty’s obligations. Table 1 shows the distribution of subjects’ responses before and after being treated with non-compliance risk. The first row shows that when subjects were initially presented with the agreement, very few chose the lowest two categories (corresponding to a 0–20 percent and 21–40 percent chance of joining); the vast majority of respondents stated that there was a greater than 60 percent chance they would join. This changes dramatically after subjects are treated with the prospect of

non-compliance, with the vast majority stating that there is a less than 60 percent chance that they would still join.

We compared the mean difference between individual subjects' responses to the original question and their responses after being treated with Non-Compliance Risk. On average, subjects' stated propensity (on our 1–5 scale) decreases by 0.74, which roughly corresponds to a 15 percent decrease in the chance that an individual subject will state a preference to join the agreement (paired *t*-test, *p*-value =  $3.8 \times 10^{-8}$ ).<sup>20</sup> Contrary to an important strand of theoretical reasoning, elites—at least in the United States—are reticent to make what are likely to be false promises.

#### *Institutional Enforcement as a Screen*

Next, we test whether a credible formal enforcement mechanism is causing this effect by dampening decision-makers' willingness to join international agreements when there is high non-compliance risk, as an alternative theoretical viewpoint would suggest. Figure 2 shows the average paired difference for subjects faced with an agreement that has an enforcement mechanism, one that has no enforcement mechanism, and the difference pooling these two conditions (vertical line represents 95 percent confidence intervals). In each case, the difference is in the expected direction (negative) and statistically significant (paired *t*-test *p*-values =  $1.9 \times 10^{-5}$ ,  $6 \times 10^{-4}$ ,  $3.8 \times 10^{-8}$ ). However, there is no significant difference between the enforcement conditions. In other words, non-compliance risk decreases the probability that decision-makers will join the agreement regardless of enforcement—elites are seeking to avoid something other than the costs of formal punishment by the institution. They do not like to make false promises even to agreements that cannot be formally enforced.

In addition to statistical insignificance, Figure 2 shows that the difference between the enforcement conditions is substantively small. On average, non-compliance risk decreases leaders' propensity to join an agreement by only an additional 3 percentage points when institutional enforcement is present. The risk of non-compliance deters leaders from joining cooperative agreements regardless of whether formal enforcement is present, and we can therefore confidently reject Hypothesis 2. Formal enforcement is not deterring subjects from joining under a high level of non-compliance risk.

#### **Discussion: Understanding Which Mechanisms Are at Work**

From the standpoint of the existing literature, it is interesting and surprising that formal enforcement has little additional effect on whether elites are deterred from making false promises. This finding suggests that many elites feared consequences, such as reputational loss or negative reciprocity, not specified in the treaty itself; future research should focus on determining which of these potential factors are actually driving elites' aversion to false promises.

To aid future research on these questions, we argue that our results are particularly consistent with two mechanisms: reputational loss and negative reciprocity. It would therefore be useful for future research to start by

<sup>20</sup>This should be taken to represent the average decrease in subjects' propensity to join the treaty, and not a statement about the percent of people who became less likely to sign the treaty.

directly testing whether either of these mechanisms can explain a substantial amount of the variation in elites' aversion to making false promises. Our argument for these mechanisms centers around the fact that they take a long time to fully play out.<sup>21</sup> As a result, their deterrent effect depends strongly on how much decision-makers discount the future (see for example Axelrod 1984; Bendor and Swistak 1997; Fearon 1998; Gilligan and Johns 2012; Grief, Milgrom, and Weingast 1994; Nowak 2006; Rand and Nowak 2013; Skyrms 2003). Therefore, if it is these consequences that deter false promises, then decision-makers who discount the future less should be more deterred. Consistent with this conjecture, we find non-experimental evidence that subjects' time preferences moderate the effect of non-compliance risk.

#### *The Role of Time Preferences in Reciprocity and Reputation*

The effect of time preferences on reciprocity has been most studied in the context of the infinitely repeated prisoners' dilemma.<sup>22</sup> A central insight is that (conditional) cooperation is more probable when players place a higher value on future payoffs. A larger number of conditionally cooperative equilibria exist when players discount future payoffs less (see Ely and Välimäki 2002; Mailath and Samuelson 2006). Recent experiments show that even when it is an equilibrium strategy to conditionally cooperate in the infinitely repeated prisoners' dilemma, such cooperation is not guaranteed (Dal Bó 2005). Instead, players must place a relatively high value on future payoffs for cooperation to emerge in controlled experiments (Dal Bó and Fréchet 2011).

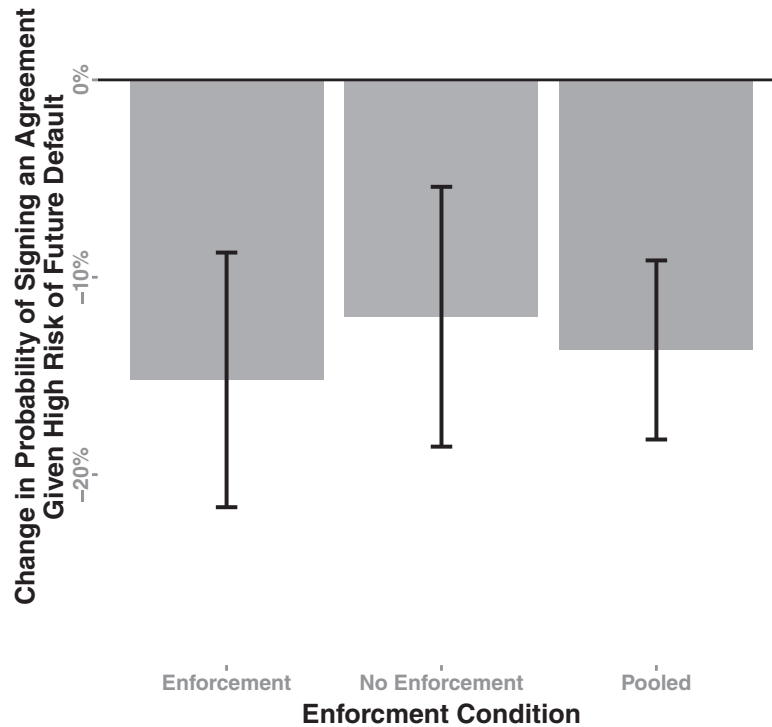
This literature is useful for understanding how one form of informal cooperation—reciprocity—depends on time-discounting and how players value the future. However, it does not directly speak to our specific question of whether the risk of future non-compliance deters actors from participating in a cooperative agreement. This is because the standard repeated prisoners' dilemma assumes that players always interact with one another, and hence always “participate” in some interaction. They therefore cannot “screen” themselves by not participating. However, there are augmented prisoners' dilemma models where, beyond cooperating or defecting, players have a third option to act as a non-participant where they earn a fixed payoff that is lower than the payoff to mutual cooperation but higher than the payoff to mutual defection (see for example Hauert et al. 2002; Abdallah et al. 2014). In these games, uncertainty about one's ability to cooperate could lead to non-participation due to fear that one's own defection will be met by defection from other players in future rounds. This would happen if players prefer the stream of benefits that come from non-participation to those that come from mutually cooperating for a few rounds and then mutually defecting in future rounds (after they themselves have defected). Such a preference also depends on players' time-discounting. If players sufficiently discount the payoffs of future rounds, they will not care that the payoff to mutual defection in future rounds is lower than the payoff to non-participation.

<sup>21</sup>This is unlike our formal enforcement mechanism, which was specified as being immediate in the text of the treatment.

<sup>22</sup>Though see Thompson (2009) and Tomz (2007) for arguments about why sanctioning and negative reciprocity might not affect cooperation in international relations.

**Table 1.** Propensity (as percentage) to join (on a scale of 1–5) before and after non-compliance risk treatment

	(1) 0–20% chance	(2) 21–40% chance	(3) 41–60% chance	(4) 61–80% chance	(5) 81–100% chance
Before Non-Compliance Risk	5	3	26	45	16
After Non-Compliance Risk	12	22	25	27	9



**Figure 2.** Effect of risk of future default on probability of signing a trade agreement

To give an example, if players play a grim-trigger strategy in the prisoners’ dilemma (Axelrod 1984; Bendor and Swistak 1997) and do not expect others to defect before they do, then a player who thinks she might defect will only prefer non-participation to participation if

$\sum_{t=n+1}^{\infty} \delta^t L - \sum_{t=n+1}^{\infty} \delta^t D > \delta^n T + \sum_{t=0}^{n-1} \delta^t C$ , where  $L$  is the per period payoff to non-participation,  $D$  is the per period payoff to mutual defection,  $T$  is the temptation payoff to unilaterally defecting for one round,  $C$  is the per period payoff to mutually cooperating,  $\delta$  is players’ discount rate, and  $n$  is the round in which a player expects to defect (as in all prisoners’ dilemmas, it is also assumed that  $T > C > L > D$ ). In other words, for screening to occur, players must not place too much weight on the payoff to mutual cooperation in early rounds relative to the downside of mutually defecting in all future rounds. Importantly in this example, if players place a sufficiently high value on the future, they will always choose non-participation.<sup>23</sup>

Time-discounting is also central to theories of reputation, where players are concerned about the future benefits that are lost from having a bad reputation. Unlike

reciprocity, reputation does not rely on the same players interacting repeatedly. Instead, a player complies with an agreement today because other players will observe her history of compliance, which can reveal valuable information about a player’s underlying propensity to be a reliable partner in cooperation. If a player’s history suggests a general tendency to renege on agreements, then other players will be less likely to form cooperative agreements with that player in the future. This can give players an incentive to signal a cooperative tendency by complying with the agreements they sign, but only if they sufficiently value the stream of future cooperation that will be generated by a good reputation. If this stream of cooperation is not sufficiently valued (relative to the immediate payoff to defection), then players may have an incentive to defect. Applying this logic to international cooperation, the effect of reputation on compliance relies on the extent to which decision-makers discount the future. What remains unknown is whether this incentive is sufficiently strong to prompt leaders to actually comply with agreements (Downs and Jones 2002; Downs et al. 1996).

Patience is a behavioral construct that directly corresponds to the time-discounting factor in formal models of cooperation and decision making. By “patience,” we simply mean how much a decision-maker values rewards in the distant future compared to rewards in the here and now. While previous literature has mapped the time-

<sup>23</sup>This follows from the fact that the left-hand side of the inequality is positive and approaches infinity as  $\delta$  approaches 1. The right-hand side is also positive, but finite. Therefore, the player will always prefer  $L$  to defecting in the  $n$ th round if  $\delta$  is sufficiently close to 1.



discounting parameter in existing models onto institutional features of a state, such as the expected longevity of a regime (see Guzman 2002), patience is a trait that exists at the level of an individual decision-maker. Empirically, this trait can vary substantially between individuals (Coller and Williams 1999; Fowler and Kam 2006; Harrison, Lau, and Williams 2002; Mischel 1974) but exhibits a great deal of stability over an individual's lifetime and across different environments (Casey et al. 2011; Meier and Sprenger 2015). It is therefore likely that the level of patience that decision-makers exhibit tells us something important about their preferences for international cooperation—including how those preferences might vary with the risk of non-compliance.

Because failure to comply could reduce a country's ability to gain the benefits of future cooperation, decision-makers who place a higher value on future payoffs should be less willing to risk non-compliance. They will be deterred from joining an agreement when their ability to comply comes into question (Brewster 2009; Downs and Jones 2002; Guzman 2002, 2008; Kreps and Wilson 1982). This relationship between patience and non-compliance risk may hold because decision-makers are strategic and calculating, or because patient decision-makers are more likely to evolve and adopt norms of compliance. The relationship may exist because the players interact repeatedly and thus are more attentive to future reciprocity, or because they are concerned about the future benefits that flow from a good reputation. By contrast, decision-makers who care less about future rewards may focus more on whether they like the terms of an agreement today and less on whether their country will be able to comply with those terms in the future. This leads us to propose a third hypothesis:

**H3:** *On average, more patient decision-makers will be less willing than impatient decision-makers to join an agreement if they possess information that their country will likely default on the agreement's terms.*

#### *Measuring Time Preferences*

To test our hypothesis on patience, we used standard tasks from behavioral decision theory (Camerer 2003). To measure how much subjects value the future—that is, their level of *patience*—we adapted a “choice game” introduced by Coller and Williams (1999). Here we refer to this game as a time-discounting task in order to more intuitively evoke the game's purpose. Past studies have linked behavior in this task to real-world behavior, such as savings rates (Harrison et al. 2002) and voting behavior (Fowler and Kam 2006). It has also been noted that time preferences should theoretically affect individual preferences for any policy option with long-run implications, such as climate restrictions (Lind 1995; Nordhaus 1997), but whether this is the case remains an empirical question (which we investigate in the domain of international agreement compliance). Consistent with this hypothesis, our recent work has found a relationship between time-discounting and preferences for policy-related outcomes, such as whether individuals prefer to conduct international negotiations with larger groups (Hafner-Burton et al. 2014), and whether international policy elites prefer to reject inequitable bargains (LeVeck et al. 2014).

In our study (as in others using this choice game), subjects were asked to make twenty different choices between

a \$100 prize that would be paid to them within thirty days after taking the study and a variable, larger prize that would be paid within sixty days. For each subject, a measure of patience is the number of sixty-day choices. Time-indifferent, highly patient players will always choose the sixty-day prize even if it is just a tiny bit larger than the \$100 offered at thirty days; players accustomed to loan sharks' payday borrowing will usually choose the more immediate prize. (Additional discussion of how these choices relate to discount rates is in the SI.) Figure 3 shows the distribution of sixty-day choices made by subjects in our study, which is similar to choices found in many other studies (see Coller and Williams 1999; Fowler and Kam 2006).

Compared to previous studies, which have looked at convenience samples of undergraduates (Fowler and Kam 2006) or the general public (Harrison et al. 2002), our population of elites is more patient on average. However, the distribution of patience among elites shares many features of the distributions found in convenience samples. Like in other studies, the distribution is multimodal with spikes around salient points such as the extremes. Most importantly, like in other studies, there is substantial heterogeneity in our elite population. So to the extent that patience affects elite decision-making under an increased risk of non-compliance, a decision may depend substantially on which decision-makers are in the room.

#### *Patience Deters Cooperation as Non-Compliance Risks Rise*

We have established that non-compliance risk dampens decision-makers' willingness to cooperate, and that the cause of this screening effect is not—as some theories would expect—the threat of formal enforcement. Here, we report—using non-experimental data—that concerns over long-term factors also correspond to decision-maker preferences for cooperating. To do this, we examine whether patience made subjects even less willing to sign a trade agreement after being given information that there was a 50 percent chance that their country would default on the agreement's obligations (Hypothesis 3).

To show that the effect of patience further affects decision-makers' propensity to join under non-compliance risk, we analyzed the following model, which we estimated using OLS<sup>24</sup> regression with robust standard errors.<sup>25</sup>

$$y = \beta_0 + \beta_1 \text{Enforcement} + \beta_2 \text{Patience} + \beta_3 \text{RiskAversion} + \beta_j \text{InitialPropensity}_j \quad (1)$$

Here,  $y_i$  is the subject  $i$ 's measured propensity to join the agreement after they have been treated with non-compliance risk (on a scale of 1–5).

*InitialPropensity<sub>j</sub>* is a set of four dummy variables, which control for a subject's initial propensity to join the agreement (prior to being treated with non-compliance risk), with the excluded category corresponding to subjects who

<sup>24</sup>OLS is appropriate since each category on our dependent variable represents an equally spaced probability interval. However, we further tested this assumption by estimating an ordered probit model. A log-likelihood ratio test shows that we cannot reject the equal-spacing assumption.

<sup>25</sup>We also obtain a similar result if we look at the Spearman rank correlation between the difference in subjects' responses,  $\Delta r = r_n - r_i$  (their stated propensity to join after being treated with non-compliance risk minus their initial propensity to join), and our measure of patience,  $\rho = -.24$  ( $Z = -2.30$ , two-sided  $p$ -value = 0.02).

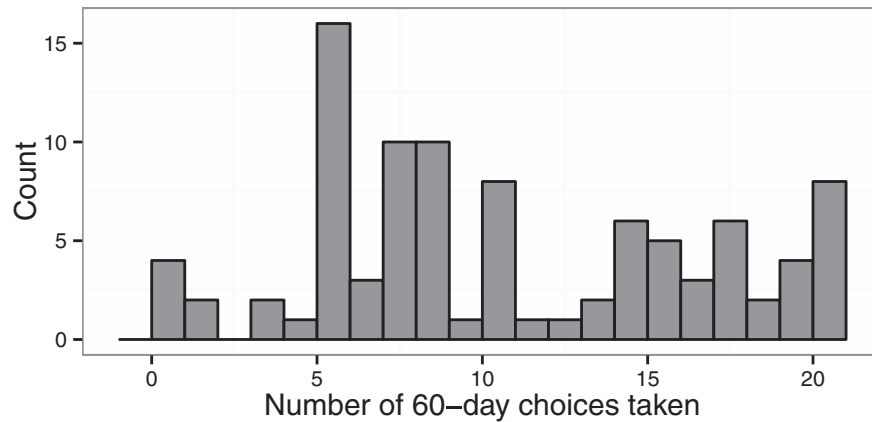


Figure 3. Distribution of elite patience

said their initial propensity to join was 1 (0–20 percent). For example, if a subject stated that their initial propensity to join the agreement was “3” (indicating a 40–60 percent chance they would join the agreement), then  $InitialPropensity_3$  was coded as 1 and all other values of  $InitialPropensity_{j \neq 3}$  were coded as 0. This set of dummy variables effectively controls for any factor that would make a decision-maker more or less likely to join the particular agreement prior to being treated with increased non-compliance risk. Including these dummies therefore allows us to isolate how *Enforcement*, *Patience*, and *RiskAversion* further change decision-makers’ response to increasing the risk of non-compliance, independent of other treaty-related considerations.<sup>26</sup>

*Enforcement* is a dummy variable coded 1 if the subject was in the enforcement condition, and 0 if they were in the non-enforcement condition. *Patience* is our measure for time preferences. To make the magnitude of this coefficient more interpretable, we subtracted the mean of *Patience* and divided by 2 standard deviations, which puts it on roughly the same scale as a binary indicator (such as *Enforcement*).

We also include a measure of risk aversion because risk aversion is known to be empirically related to patience but may also exert an independent effect on whether decision-makers are affected by non-compliance risk. Crucially for our study, the effect of risk aversion on willingness to cooperate—unlike the effect of patience—operates over all time horizons. Risk aversion should moderate the effect of increased non-compliance risk regardless of whether an agreement is enforced by swift institutional punishment or whether it is enforced by the shadow of the future. Therefore, to the extent that risk aversion and patience are empirically related, we need to control for risk aversion. This allows us to draw stronger conclusions about whether decision-makers are concerned with the long-term consequences of non-compliance; if patience still moderates decision-makers’ responses to non-compliance risk after controlling for risk aversion, then it is highly likely that concerns about compliance are related to mechanisms that rely on the shadow of the future. To measure subjects’ risk aversion, we used a

multiple price-list task adapted from Holt and Laury (2002), which we describe in detail in the SI to economize on space.<sup>27</sup>

Column 1 of Table 2 below shows a model that only includes the effect of *Enforcement* by itself. The sign on the estimated coefficient is negative; however, it is not statistically distinguishable from 0. This is consistent with what we found using paired *t*-tests in the previous section.

Columns 2 and 3 of Table 2 show a model that includes an estimate of how *Patience* and *RiskAversion* affect subjects’ propensity to join under increased non-compliance risk (after controlling for their initial propensity to join the agreement). The sign on both coefficients is negative and statistically significant. The coefficient of  $-0.56$  on *Patience* (in column 2) roughly corresponds to an additional 11 percent drop in subjects’ propensity to join. Given how we normalized our measure, this means that compared to the least patient person in our sample, additional risk of non-compliance decreases the most patient subjects’ propensity to join by an additional 11 percent. Therefore the difference in the effect of risk between the most and least patient is about 70 percent as large as the average effect of risk itself. Likewise, the coefficient of  $-0.58$  on *RiskAversion* roughly means that, compared to the least risk-averse subject in our sample, the most risk-averse person is approximately 12 percent less likely to join the agreement. Column 3 shows that nothing changes when we combine all variables in one regression.

The fact that patience substantially changes subjects’ responses to non-compliance risk is consistent with (though not causal proof of) the idea that subjects are concerned about the long-term consequences of defection. If concerns were only related to the more immediate consequences on non-compliance, then only risk aversion would matter. However, patience moderates the effect of non-compliance risk independent of risk aversion. Strikingly, compared to formal enforcement, patience has

<sup>26</sup>Another way to model this would be to use a repeated measures model, where each subject has a fixed effect, and variables such as enforcement and patience are interacted with a dummy for non-compliance risk. Such a model gives us essentially the same results, but early reviewers of the paper, who were not used to these models, found the presentation confusing. This version of the results is available from the authors upon request.

<sup>27</sup>O’Neill (2001) makes an interesting critique when it comes to using concepts such as Arrow–Pratt risk aversion, which unlike many gambles in international relations are based on preferences over monetary lotteries with known probabilities and known payoffs. However, psychologists have validated our risk-aversion task as a measure that predicts decision-makers’ willingness to take risky decisions across a number of domains where individuals do not perfectly know the stakes or probabilities, such as smoking and seat-belt use (Anderson and Mellor 2008). The measure may therefore still account for variation in subjects’ willingness to take risks in general, even if choices over monetary lotteries do not perfectly map onto the risks faced by decision-makers in international relations.

**Table 2.** Regression on likelihood that a decision-maker will prefer to sign an agreement after they face an increased risk that their country will not comply

	<i>Enforcement</i>	<i>Traits</i>	<i>Combined</i>
Enforcement	-0.19 (0.222)		-0.25 (0.210)
Patience (normed)		-0.56** (0.203)	-0.56** (0.202)
Risk Aversion (normed)		-0.58** (0.213)	-0.61** (0.214)
Initial Propensity to Join:			
Intercept	2.35*** (0.492)	2.65*** (0.454)	2.87*** (0.489)
(Chose 1: 0-20%)			
(2) 21-40%	-0.96 (0.757)	-1.45* (0.725)	-1.59* (0.732)
(3) 41-60%	0.04 (0.507)	-0.36 (0.491)	-0.47 (0.498)
(4) 61-80%	1.02* (0.491)	0.59 (0.483)	0.48 (0.491)
(5) 81-100%	1.62** (0.526)	1.23** (0.511)	1.21** (0.510)
<i>N</i>	95	95	95
<i>R</i> <sup>2</sup>	0.38	0.38	0.39
Adj <i>R</i> <sup>2</sup>	0.34	0.34	0.34

Standard errors in parentheses. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

a larger estimated impact on subjects' responses to non-compliance risk.

### Conclusions and Implications

The debate over whether institutional enforcement is desirable for international cooperation has been running for decades. It is of critical importance to many larger theoretical controversies in the fields of international relations and of public international law. But the debate has lacked much in the way of clean, empirical tests of the competing perspectives.

Our methods and findings accomplish just that. First, we provide the first direct elite-level evidence that increasing the risk of non-compliance with international obligations decreases real policymakers' willingness to cooperate by joining agreements. Actual decision-makers believe in eschewing commitments unless their government can likely honor them. Elite decision-makers, at least in the context of US foreign policy, are reluctant to make false promises.

Second, we provide important evidence that this aversion effect is not driven by variation in the design of the commitment—specifically, by the presence of a formal enforcement mechanism. Elites may very well seek to avoid retaliation through formal enforcement. However, institutional retaliation is not the only factor they consider; other factors must explain why decision-makers seek to avoid commitments with uncertain prospects for compliance even when agreements are not formally enforceable. These may include the fear of informal retaliation, reputational loss, or the desire to abide by other international norms.

The combination of the first two findings suggests that— from the perspective of the policy elites who actually make state policy decisions about designing, joining, and complying with international agreements—enforcement mechanisms are relatively unimportant. However, the design of commitments to allow compliance proves crucial. States can achieve compliance through shallow cooperation, of

course (Downs, Rocke, and Barsoom 1996). But deep cooperation requires commitments that elicit changes in state behavior without raising fears of non-compliance. We need more work applied to real policy settings to reveal how flexibility—such as in content and in opt-out provisions—can yield commitments that are simultaneously demanding yet also achievable. One interpretation of the large shift in legal design on climate change—away from Kyoto-style inflexible binding targets toward a more tailored bottom-up system of state-designed commitments—is that diplomats are learning this lesson about flexibility (Victor 2015; Keohane and Victor 2016).

Third, we demonstrated that not all elite decision-makers react alike to non-compliance risk. Patient people—those with long time horizons—are more sensitive to the risk that their government will not comply. In other words, the extent to which an agreement screens participation—by deterring countries from joining—may depend not so much on the enforcement structure of the agreement but rather on how decision-makers themselves weigh the future.

Finally, this article suggests that new work would be useful on at least three fronts:

First, do these findings hold when looking at other issue areas beyond trade? We argued that the structure of the trade agreement specified in this study mimics other forms of commitments. However, cooperation in other domains can implicate different bargaining games and thus different concerns about the impacts of non-compliance on cooperation. Varying the domain and particular cooperation problem would also align with the need to examine different institutional arrangements. For example, in addition to formal enforcement mechanisms, many treaties have provisions for withdrawal and some also allow for derogations or other kinds of flexibility in the face of non-compliance. It would be interesting to see if the same evidence about the reticence to make false promises holds if policy elites think that these other institutional responses to non-compliance are available.

Second, we need more work to nail down the exact causal mechanisms at work. Our research establishes that elites prefer to avoid false commitments, regardless of whether those commitments are formally enforceable. Yet, one might reasonably wonder (as we did) why formal enforcement has little additional effect in our experiment. The fact that time preferences moderate the effect of non-compliance risk points to a set of potential mechanisms already suggested by the non-experimental literature on cooperation and compliance. Therefore, future studies could usefully test which of these potential mechanisms best explain our finding.

Third, do our results generalize to other countries? There are many reasons that American elites may differ from their foreign counterparts. For example, coming from a powerful country may make them less concerned about reputation—which may suggest that our results would be even stronger with elites in other countries. Americans may also be disinclined to think that formal enforcement mechanisms work because it is easier for a powerful country to ignore inconvenient rulings from enforcement institutions—a pattern evident not only in trade, but also in many other issue areas. And leaders in autocracies may have different attitudes about compliance risk than those from democracies. Perhaps democratic leaders, for example, are much more aware of the many ways that national political processes can yield involuntary defection—and also political pressures for compliance. This would make them more sensitive to how such outcomes harm the prospects for international cooperation. But perhaps Americans are not so different, or their differences derive from cultural factors rather than the power position of the United States. Such questions, we hope, will motivate future experiments on elite-level behavior.

### Supplementary Materials

Supplemental information is available at the *International Studies Quarterly* data archive.

### References

- ABDALLAH, SHERIEF, RASHA SAYED, IYAD RAHWAN, BRAD L. LEVECK, MANUEL CEBRIAN, ALEX RUTHERFORD, AND JAMES H. FOWLER. 2014. "Corruption Drives the Emergence of Civil Society." *Journal of the Royal Society Interface* 11(93): 1742–5662.
- ACHEN, CHRISTOPHER H. 1975. "Mass Political Attitudes and the Survey Response." *American Political Science Review* 69(4): 1218–31.
- ANDERSON, LISA R., AND JENNIFER M. MELLOR. 2008. "Predicting Health Behaviors with an Experimental Measure of Risk Preference." *Journal of Health Economics* 27(5): 1260–74.
- ANSOLABEHRE, STEPHEN, JONATHAN RODDEN, AND JAMES M. SNYDER JR. 2008. "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting." *American Political Science Review* 102(2): 215–32.
- AXELROD, ROBERT. 1984. *The Evolution of Cooperation*. New York: Basic Books.
- AXELROD, ROBERT, AND ROBERT O. KEOHANE. 1985. "Achieving Cooperation under Anarchy: Strategies and Institutions." *World Politics* 38(1): 226–54.
- BENDOR, JONATHAN, AND PIOTR SWISTAK. 1997. "The Evolutionary Stability of Cooperation." *American Political Science Review* 91(2): 290–307.
- BREWSTER, RACHEL. 2009. "Unpacking the State's Reputation." *Harvard International Law Journal* 50: 231–70.
- BRUNNÉE, JUTTA, AND STEPHEN J. TOOPE. 2010. *Legitimacy and Legality in International Law: An Interactional Account*. Cambridge: Cambridge University Press.
- CAMERER, COLIN F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.
- CASEY, B. J., LEAH H. SOMERVILLE, IAN H. GOTLIB, OZLEM AYDUK, NICHOLAS T. FRANKLIN, MARY K. ASKREN, JOHN JONIDES, et al. 2011. "Behavioral and

- Neural Correlates of Delay of Gratification 40 Years Later." *Proceedings of the National Academy of Sciences of the United States of America* 108(36): 14998–5003.
- CHAYES, ABRAM, AND ANTONIA HANDLER CHAYES. 1993. "On Compliance." *International Organization* 47(2): 175–205.
- . 1998. *The New Sovereignty: Compliance with International Regulatory Agreements*. Cambridge, MA: Harvard University Press.
- CHILTON, ADAM, AND DUSTIN TINGLEY. 2013. "Why the Study of International Law Needs Experiments." *Columbia Journal of Transnational Law* 52(1): 173–238.
- COLLER, MARIBETH, AND MELONIE B. WILLIAMS. 1999. "Eliciting Individual Discount Rates." *Experimental Economics* 2(2): 107–27.
- DAL BÓ, PEDRO. 2005. "Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games." *American Economic Review* 95(5): 1591–604.
- DAL BÓ, PEDRO, AND GUILLAUME R. FRÉCHETTE. 2011. "The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence." *American Economic Review* 101(1): 411–29.
- DELLAVIGNA, STEFANO. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47(2): 315–72.
- DEPLEDGE, JOANNA. 2005. "Against the Grain: The United States and the Global Climate Change Regime." *Global Change, Peace & Security* 17(1): 11–27.
- DOWNES, GEORGE W., AND MICHAEL A. JONES. 2002. "Reputation, Compliance, and International Law." *Journal of Legal Studies* 31(1): S95–114.
- DOWNES, GEORGE W., AND DAVID M. ROCKE. 1995. *Optimal Imperfection? Domestic Uncertainty and Institutions in International Relations*. Princeton, NJ: Princeton University Press.
- DOWNES, GEORGE W., DAVID M. ROCKE, AND PETER N. BARSOOM. 1996. "Is the Good News About Compliance Good News About Cooperation?" *International Organization* 50(3): 379–406.
- EDELSTEIN, DAVID M. 2002. "Managing Uncertainty: Beliefs about Intentions and the Rise of Great Powers." *Security Studies* 12(1): 1–40.
- ELY, JEFFREY C., AND JUUSO VALIMÄKI. 2002. "A Robust Folk Theorem for the Prisoner's Dilemma." *Journal of Economic Theory* 102(1): 84–105.
- FEARON, JAMES D. 1998. "Bargaining, Enforcement, and International Cooperation." *International Organization* 52(2): 269–305.
- FINNEMORE, MARTHA, AND STEPHEN J. TOOPE. 2001. "Alternatives to 'Legalization': Richer Views of Law and Politics." *International Organization* 55(3): 743–58.
- FOWLER, JAMES H., AND CINDY D. KAM. 2006. "Patience as a Political Virtue: Delayed Gratification and Turnout." *Political Behavior* 28(2): 113–28.
- FRANCK, THOMAS M. 1990. *The Power of Legitimacy Among Nations*. New York: Oxford University Press.
- GILLIGAN, MICHAEL J. 2004. "Is There a Broader-Deeper Trade-Off in International Multilateral Agreements?" *International Organization* 58(3): 459–84.
- GILLIGAN, MICHAEL J., AND LESLIE JOHNS. 2012. "Formal Models of International Institutions." *Annual Review of Political Science* 15: 221–43.
- GOLDSMITH, JACK L., AND ERIC A. POSNER. 2005. *The Limits of International Law*. New York: Oxford University Press.
- GREIF, AVNER, PAUL MILGROM, AND BARRY R. WEINGAST. 1994. "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild." *Journal of Political Economy* 102(4): 745–76.
- GUISINGER, ALEXANDRA, AND ALASTAIR SMITH. 2002. "Honest Threats: The Interaction of Reputation and Political Institutions in International Crises." *Journal of Conflict Resolution* 46(2): 175–200.
- GUZMAN, ANDREW T. 2002. "A Compliance-Based Theory of International Law." *California Law Review* 90: 1823–88.
- . 2008. *How International Law Works: A Rational Choice Theory*. Oxford and New York: Oxford University Press.
- HAFNER-BURTON, EMILIE M. 2005. "Trading Human Rights: How Preferential Trade Agreements Influence Government Repression." *International Organization* 59(3): 593–629.
- HAFNER-BURTON, EMILIE M., STEPHAN HAGGARD, DAVID A. LAKE, AND DAVID G. VICTOR. Forthcoming. "The Behavioral Revolution and the Study of International Relations." *International Organization*.
- HAFNER-BURTON, EMILIE M., D. ALEX HUGHES, AND DAVID G. VICTOR. 2013. "The Cognitive Revolution and the Political Psychology of Elite Decision Making." *Perspectives on Politics* 11(2): 368–86.
- HAFNER-BURTON, EMILIE M., BRAD L. LEVECK, DAVID G. VICTOR, AND JAMES H. FOWLER. 2014. "Decision Maker Preferences for International Legal Cooperation." *International Organization* 68(4): 845–76.

- HAFNER-BURTON, EMILIE M., DAVID G. VICTOR, AND YONATAN LUPU. 2012. "Political Science Research on International Law: The State of the Field." *American Journal of International Law* 106(1): 47–97.
- HARRISON, GLENN W., MORTEN I. LAU, AND MELONIE B. WILLIAMS. 2002. "Estimating Individual Discount Rates in Denmark: A Field Experiment." *American Economic Review* 92(5): 1606–17.
- HATHAWAY, OONA. 2005. "Why Do Countries Commit to Human Rights Treaties?" *Journal of Conflict Resolution* 51(4): 588–621.
- HAUERT, CHRISTOPH, SILVIA DE MONTE, JOSEF HOFBAUER, AND KARL SIGMUND. 2002. "Volunteering as Red Queen Mechanism for Cooperation in Public Goods Games." *Science* 296(5570): 1129–32.
- HOLT, CHARLES A., AND SUSAN K. LAURY. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92(5): 1644–55.
- HOVL, JON, DETLEF F. SPRINZ, AND GURI BANG. 2012. "Why the United States Did Not Become a Party to the Kyoto Protocol: German, Norwegian and US Perspectives." *European Journal of International Relations* 18(1): 129–50.
- IIDA, KEISUKE. 1993. "Analytic Uncertainty and International Cooperation: Theory and Application to International Economic Policy Coordination." *International Studies Quarterly* 37(4): 431–57.
- JARVIS, DARRYL S. L. 2011. "Theorising Risk and Uncertainty in International Relations: The Contributions of Frank Knight." *International Relations* 25(3): 296–312.
- JERVIS, ROBERT. 1989. *The Logic of Images in International Relations*. New York: Columbia University Press.
- KEOHANE, ROBERT O., AND DAVID G. VICTOR. 2016. "Cooperation and Discard in Global Climate Policy." *Nature Climate Change* 6(6): 570–75.
- KERTZER, JOSHUA D. 2013. "Making Sense of Isolationism: Foreign Policy Mood as a Multilevel Phenomenon." *Journal of Politics* 75(1): 225–40.
- KERTZER, JOSHUA D., AND KATHLEEN M. MCGRAW. 2012. "Folk Realism: Testing the Microfoundations of Realism in Ordinary Citizens." *International Studies Quarterly* 56(2): 245–58.
- KOH, HAROLD. 1997. "Why Do Nations Obey International Law?" Faculty Scholarship Series Paper 2101, New Haven, CT USA: Yale Law School.
- KOREMENOS, BARBARA. 2005. "Contracting around International Uncertainty." *American Political Science Review* 99(4): 549–65.
- KOREMENOS, BARBARA, CHARLES LIPSON, AND DUNCAN SNIDAL. 2001. "The Rational Design of International Institutions." *International Organization* 55(4): 761–99.
- KREPS, DAVID M., AND ROBERT WILSON. 1982. "Sequential Equilibria." *Econometrica* 50(4): 863–94.
- LEVECK, BRAD L., D. ALEX HUGHES, JAMES H. FOWLER, EMILIE HAFNER-BURTON, AND DAVID G. VICTOR. 2014. "The Role of Self-Interest in Elite Bargaining." *Proceedings of the National Academy of Sciences of the United States of America* 111(52): 18536–41.
- LEVECK, BRAD L., AND NEIL NARANG. Forthcoming. "How International Reputation Matters: Revisiting Alliance Violations in Context." *International Interactions*.
- LIND, ROBERT. 1995. "Intergenerational Equity, Discounting, and the Role of Cost-Benefit Analysis in Evaluating Global Climate Policy." *Energy Policy* 23(4–5): 379–89.
- MAILATH, GEORGE J., AND LARRY SAMUELSON. 2006. *Repeated Games and Reputations: Long-Run Relationships*. New York: Oxford University Press.
- MANSFIELD, EDWARD D., AND DIANA C. MUTZ. 2009. "Support for Free Trade: Self-Interest, Sociotropic Politics, and Out-Group Anxiety." *International Organization* 63(3): 425–57.
- MARTIN, LISA L. 1994. "Heterogeneity, Linkage and Commons Problems." *Journal of Theoretical Politics* 6(4): 473–93.
- . 2013. "Against Compliance." In *Interdisciplinary Perspectives on International Law and International Relations: The State of the Art*, edited by Jeffrey L. Dunoff and Mark A. Pollack, 591–612. New York: Cambridge University Press.
- MCDERMOTT, ROSE. 1998. *Risk-Taking in International Politics: Prospect Theory in American Foreign Policy*. Ann Arbor: University of Michigan Press.
- . 2004. *Political Psychology in International Relations*. Ann Arbor: University of Michigan Press.
- . 2011. "New Directions for Experimental Work in International Relations." *International Studies Quarterly* 55(2): 503–20.
- MCDERMOTT, ROSE, AND YOSHIKO HERRERA. 2010. "Psychological versus Rational Models of Human Decision-Making Symposium." *APSA Comparative Politics Newsletter* 21(1): 21–3.
- MEIER, STEPHAN, AND CHARLES D. SPRENGER. 2015. "Temporal Stability of Time Preferences." *Review of Economics and Statistics* 97(2): 273–86.
- MERCER, JOHNATHAN. 2012. "Audience Costs Are Toys." *Security Studies* 21(3): 398–404.
- MILNER, HELEN V., AND DUSTIN TINGLEY. 2013a. "Public Opinion and Foreign Aid: A Review Essay." *International Interactions* 39(3): 389–401.
- . 2013b. "The Choice for Multilateralism: Foreign Aid and American Foreign Policy." *Review of International Organizations* 8(3): 313–41.
- MINTZ, ALEX, STEVEN B. REDD, AND ARNOLD VEDLITZ. 2006. "Can We Generalize from Student Experiments to the Real World in Political Science, Military Affairs, and International Relations?" *Journal of Conflict Resolution* 50(5): 757–76.
- MINTZ, ALEX, YI YANG, AND ROSE MCDERMOTT. 2011. "Experimental Approaches to International Relations." *International Studies Quarterly* 55(2): 493–501.
- MISCHEL, WALTER. 1974. "Processes in Delay of Gratification." In *Advances in Experimental Social Psychology*, vol. 7, edited by Leonard Berkowitz, 249–92. New York: Academic Press.
- NORDHAUS, WILLIAM. 1997. "Discounting in Economics and Climate Change: An Editorial Comment." *Climatic Change* 37(2): 315–28.
- NOWAK, MARTIN A. 2006. "Five Rules for the Evolution of Cooperation." *Science* 314(5805): 1560–63.
- PEINHARDT, CLINT, AND TODD SANDLER. 2015. *Transnational Cooperation: An Issue-Based Approach*. New York: Oxford University Press.
- RAND, DAVID G., AND MARTIN A. NOWAK. 2013. "Human Cooperation." *Trends in Cognitive Sciences* 17(8): 413–25.
- RATHBUN, BRIAN C. 2007. "Uncertain about Uncertainty: Understanding the Multiple Meanings of a Crucial Concept in International Relations Theory." *International Studies Quarterly* 51(3): 533–57.
- RAUSTIALA, KAL R., AND ANNE-MARIE SLAUGHTER. 2002. "International Law, International Relations and Compliance." In *Handbook of International Relations*, edited by Walter Carlsnaes, Thomas Risse, and Beth A. Simmons, 538–58. Thousand Oaks, CA: Sage Publications.
- RENSHON, JONATHAN, JULIA LEE, AND DUSTIN TINGLEY. 2015. "Physiological Arousal and Political Beliefs." *Political Psychology* 36(5): 569–85.
- RHO, SUNGMIN, AND MICHAEL TOMZ. 2013. "Industry, Self-Interest, and Individual Preferences over Trade Policy." Paper presented at the International Political Economy Colloquium, Goa, India, January 4–6.
- RHO, SUNGMIN, AND MICHAEL TOMZ. 2015. "Why Don't Trade Preferences Reflect Economic Self-Interest?" Working Paper, Stanford, CA USA: Stanford University.
- ROSENDORFF, B. PETER, AND HELEN V. MILNER. 2001. "The Optimal Design of International Trade Institutions: Uncertainty and Escape." *International Organization* 55(4): 829–57.
- SIMMONS, BETH A. 2000. "International Law and State Behavior: Commitment and Compliance in International Monetary Affairs." *American Political Science Review* 94(4): 819–35.
- SIMMONS, BETH A., AND DANIEL J. HOPKINS. 2005. "The Constraining Power of International Treaties: Theories and Methods." *American Political Science Review* 99(4): 623–31.
- SKYRMS, BRIAN. 2003. *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press.
- TETLOCK, PHILIP E. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press.
- THOMPSON, ALEXANDER. 2009. "The Rational Enforcement of International Law: Solving the Sanctioners' Dilemma." *International Theory* 1(2): 307–21.
- TINGLEY, DUSTIN. 2011. "The Dark Side of the Future: An Experimental Test of Commitment Problems in Bargaining." *International Studies Quarterly* 55(2): 521–44.
- TINGLEY, DUSTIN, AND MICHAEL TOMZ. 2012. "How Does the UN Security Council Influence Public Opinion." Working Paper, Stanford, CA USA: Stanford University and Harvard University.
- TOMZ, MICHAEL. 2007. *Reputation and International Cooperation: Sovereign Debt Across Three Centuries*. Princeton, NJ: Princeton University Press.
- . 2009. "The Foundations of Domestic Audience Costs: Attitudes, Expectations and Institutions." In *Kitai, Seido, Gurobaru-Shakai (Expectations, Institutions, and Global Society)*, edited by Masaru Kohno and Aiji Tanaka, 85–97. Tokyo: Keiso-Shobo.
- UNITED NATIONS. 2016. "Ratification of 18 International Human Rights Treaties." *UNOHCHR*. Accessed February 1, 2016, <http://indicators.ohchr.org/>.
- US WHITE HOUSE. 2015. "U.S.–China Joint Presidential Statement on Climate Change." Office of the Press Secretary, September 25.

- Accessed February 1, 2016, [www.whitehouse.gov/the-press-office/2015/09/25/us-china-joint-presidential-statement-climate-change](http://www.whitehouse.gov/the-press-office/2015/09/25/us-china-joint-presidential-statement-climate-change).
- USTR. 2015. "GSP Review of Bangladesh Recognizes Progress, Urges That More Be Done on Worker Safety and Rights." Press Release, January.
- VICTOR, DAVID G. 2001. *The Collapse of the Kyoto Protocol and the Struggle to Slow Global Warming*. Princeton, NJ: Princeton University Press.
- . 2015. "Why Paris Worked: A Different Approach to Climate Diplomacy." *Yale Environment* 360.
- VON STEIN, JANA. 2005. "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance." *American Political Science Review* 99(4): 611–22.
- WTO. 1994. *Annex 2: Understanding on Rules and Procedures Governing the Settlement of Disputes*. In *The Uruguay Round Agreements*, 353–77. Geneva, Switzerland: WTO.
- . 1998. *EC Measures Concerning Meat and Meat Products (Hormones)*. Report of the Appellate Body, WT/DS26/AB/R, WT/DS48/AB/R, January 16.