LABORATORY ON INTERNATIONAL LAW & REGULATION
UC San Diego School of Global Policy & Strategy

How The Prospect of Non-Compliance Affects Elite Preferences for International Cooperation: Evidence Fro[...] "Lab in the Field" Experiment

EMILIE M. HAFNER-BURTON, BRAD LEVECK AND DAVID G. VICTOR

About the Laboratory on International Law and Regulation
(ILAR)

The Laboratory on International Law and Regulation (ILAR) is an international, interdisciplinary laboratory that explores when and why international laws actually work. Among scholars, this question has triggered a lively debate that ILAR is engaging with better theories and evidence. ILAR research examines a wide array of issues from environment and energy to human rights, trade and security issues. The ILAR team looks at these issues from the international perspective and also through comparisons across countries.

The Laboratory is part of the School of Global Policy and Strategy at the University of California, San Diego. ILAR gratefully acknowledges anchor funding from the nonpartisan Electric Power Research Institute, BP, plc, the Norwegian Research Foundation and from UC San Diego's School of International Relations and Pacific Studies.

About the Authors

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 and the Department of Political Science at the University of California, San Diego and Director of the Laboratory on International Law and Regulation, La Jolla, CA 92093. She can be reached at ehafner@ucsd.edu.

Brad L. LeVeck is Assistant Professor of Political Science at the University of California, Merced. He can be reached at bleveck@ucmerced.edu.

David G. Victor is Professor of International Relations at the School of Global Policy and Strategy and the Department of Political Science at the University of California, San Diego and Director of the Laboratory on International Law and Regulation, La Jolla, CA 92093. He can be reached at david.victor@ucsd.edu.

How The Prospect of Non-Compliance Affects Elite Preferences for International Cooperation: Evidence From a "Lab in the Field" Experiment

Abstract:

It is well known that strategic uncertainty about other states' intentions and payoffs shapes leaders' decisions to cooperate. Less understood is whether leaders will cooperate when confronted with the prospect that their *own* government will likely default from an international commitment. Existing theories suggest that such a risk could deter leaders from cooperating in the first place, and that this aversion effect may vary with the design of the agreement—leaders will be more concerned about the prospect of their country's non-compliance with agreements that can formally impose costs for defection through enforcement. Here, we describe the results of a unique experimental survey conducted on 95 high level policy elites in the United States that allows us to look causally at the link between 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 policy makers become less willing to join international agreements. However, contrary to what many theories would predict, the presence of a formal enforcement mechanism does not explain the aversion to cooperation. Elites do not like to make false promises even to agreements that are not formally enforceable. Instead, the aversion to joining agreements when compliance is difficult to assure is rooted in decision makers' own perceptions of the future—elites who have lower discount rates are particularly sensitive to the prospect of not honoring commitments.

Word Count: 7231 (main text, excluding abstract, notes, figures and tables)

It is well known that strategic uncertainty about other states' intentions and payoffs shapes decisions to cooperate internationally. Less understood is whether leaders will cooperate when confronted with the prospect that their *own* government will default from an international commitment.[1] Do governments willingly commit to international agreements if they are not sure they can comply? Or does the prospect of a possibly unattainable obligation compel leaders to avoid cooperation in the first place? This paper demonstrates that the prospect of future default strongly inhibits actual policy makers' willingness to commit to international agreements. However—and contrary to what many theories would predict—the presence of a formal institutional enforcement mechanism does not explain the aversion. Elites do not like to make false promises of cooperation even when international agreements cannot be formally enforced. Instead, we argue, the extent to which the risk of default deters cooperation depends more on how decision makers, themselves, weight the future.

There are different reasons why this particular type of risk might deter leaders from joining international agreements. A central one concerns the preference to avoid costly punishments imposed by the institution itself. A government that is uncertain about the country's future compliance may avoid making demanding commitments to institutions that credibly enforce those obligations.[2] By this logic, governments that know they have a substantial prospect of default should not commit to agreements that have costly enforcement procedures and will

---

[1] On risk taking in international relations more generally, see: McDermott 1998. On how uncertainty shapes the design of international institutions, see for example Rosendorff and Milner 2001 and Koremenos 2005.

[2] Downs and Rocke 1995; Downs, Rocke, and Barsoom 1996; Koremenos 2005; Rosendorff and Milner 2001; von Stein 2005. A similar logic also applies when treaties are inflexible about the legal commitment they require states to implement. See Rosendorff and Milner 2001.

seek more flexible cooperation or no cooperation at all.[3]

Alternatively, decision makers may adhere to international agreements out of a sense of obligation, a concern for reputation, or fear of informal retaliation.[4] By these logics, compliance flows from the knowledge that interactions will be repeated and adherence to norms, even in the absence of formal institutionalized enforcement mechanisms, can be an optimal strategy.[5] Whether a cooperative agreement includes a formal enforcement mechanism is orthogonal to this logic because governments are responsive to other mechanisms that punish non-compliance.

Relevant to our central finding—that elites prefer not to make false promises even when agreements are formally unenforceable—is the fact that choices about institutional cooperation are made by individuals who vary in how they react to the possible consequences of non-compliance.[6] Governments do not think; people do.[7] The list of factors that influence individual preferences and choices is long[8], but the formal literature on cooperation has long understood that a central factor in determining whether repeated interaction will yield cooperation is the value that decision-makers place on the future. The discount rate—what we call "patience"—is an attribute that varies substantially at the level of the decision maker.[9] Patience, we suggest, can affect cooperation in at least two ways: it can alter the incentives for reciprocity in repeated games and it can affect the weight that decision makers place on reputation. Such mechanisms can be at work even if agreements do not themselves contain formal enforcement mechanisms.

---

[3] On the broader-deeper trade-off more generally, see Gilligan 2004.
[4] Finnemore and Toope 2001; Franck 1990; Guzman 2008; Thompson 2009; Tomz 2007.
[5] Axelrod 1984; Axelrod and Keohane 1985; Tingley 2011. For an alternative view which argues that a longer shadow of the future may exacerbate commitment problems, see Tingley 2011.
[6] McDermott 2004; McDermott and Herrera 2010. On individual heterogeneity, see Kertzer 2013; Martin 1994; Renshon, Lee, and Tingley 2014.
[7] McDermott 2004; McDermott and Herrera 2010.
[8] DellaVigna 2009.
[9] Coller and Williams 1999; Fowler and Kam 2006.

Few treaties actually have institutionalized enforcement mechanisms, and comparison across existing agreements is difficult because there are so many confounding factors and selection effects.[10] Outside of an experimental setting, it is also difficult to exogenously change decision-maker beliefs about whether their own country might comply with an agreement while holding all other factors constant. Indeed, the real world is unlikely to offer opportunities to observe the effects of an agreement's content independent of its design because there are strong reasons to expect that the presence or absence of formal enforcement is endogenous to the interests and expectations of decision makers.[11]

One way to resolve these difficulties is to systematically and independently vary both the perceptions that a country might not comply and whether an agreement has a formal enforcement mechanism.[12] We develop an experimental survey that allows us to look causally at the link between the presence of institutional enforcement and decision-maker preferences for international cooperation as perceptions of the likelihood of default are varied.[13] Our survey

---

[10] Martin 2013.

[11] Fearon 1998. 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.

[12] 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 which are empirically correlated with enforcement in the minds of respondents, 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.

[13] Among the growing number of studies within international relations that also use experiments are: Chilton and Tingley 2013; McDermott 2011; Mintz, Yang, and Mc Dermott 2011; Tingley 2011.

examines the willingness of decision makers to join a trade agreement that approximates the first step in an iterated assurance game that is typical of many domains of international cooperation, such as the reduction of trade barriers, controlling emissions that cause climate change, and the adoption of technology standards that affect market access. We survey a unique sample of 95 actual U.S. 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.[14]

Sampling actual elites is important for this study because they differ in some key ways from non-elites that could affect our findings. For example, elite populations tend to be more strategic, patient, and fair than college students and other convenience samples typically used for experimental research.[15] They also tend to be more cooperative and there is some evidence that elites are more likely to update decision-making heuristics in light of new information.[16]

Our survey yields three novel findings. First, the prospect of non-compliance by one's own country—operationalized here as a 50% chance of future defection—significantly dampens elite willingness to join international commitments in the first place. Second, this aversion is not affected by the presence of formal institutional enforcement intended to impose costs on defectors. Elites in our sample prefer to avoid false promises even when agreements are formally

---

[14] 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; Tomz 2009. For a review of elite experiments and differences from non-elite populations see Hafner-Burton, Hughes, and Victor 2013.

[15] Hafner-Burton et al. 2014; LeVeck et al. 2014.

[16] Hafner-Burton, Hughes, and Victor 2013.

unenforceable. Third, the extent to which the risk of defaulting affects preferences for cooperation depends more on decision makers' own long-run concerns for the future than on concern for formal enforcement. 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 informal long-term effects of their country's broken promise.[17] This insight is consistent with the logic that concern for reputation or informal reciprocity through repeated interaction is influencing the calculus of cooperation, although pinning down exactly which mechanism is at work is a topic for further experimental research.

The paper 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 influence decision-maker preferences for international cooperation. We explain why heterogeneity in decision-maker time preferences affects their willingness to cooperate in the face of potential future default, describe our survey methodology, and present our results. We conclude with broader implications for theories of cooperation.

Non-Compliance Risk

One of the most enduring problems in politics is that elite decision makers often must make important decisions under conditions of risk and uncertainty.[18] There are multiple types of

---

[17] Guisinger and Smith 2002.
[18] Edelstein 2002; Jarvis 2011.

9

risks that shape those decisions.[19] The standard way to 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.[20] Actors may not possess full information about their own payoffs from a bargain because they are unsure about their own future interests or capabilities.[21] What we call "non-compliance risk" is one example.[22] Leaders negotiate the best international agreements possible but what happens after those agreements are signed can be unpredictable and sometimes beyond the control of the state.[23] 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.[24] Unexpected changes in technology, political attitudes or other factors can also reduce the capacity of governments to fulfill their commitments.[25]

More formally, we define non-compliance risk as the probability that a leader's country will not be able to comply with a 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.[26] Our aim is to study how elite decision-makers evaluate the potential consequences of potential non-

---

[19] Rathbun 2007.
[20] Koremenos, Lipson, and Snidal 2001.
[21] Iida 1993.
[22] Kertzer and McGraw 2012 have used experiments to examine other forms of uncertainty (such as strategic uncertainty) in non-elite populations.
[23] Koremenos 2005.
[24] e.g., Downs and Rocke 1995; Rosendorff and Milner 2001.
[25] Chayes and Chayes 1993.
[26] 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.

compliance. Do they fear punishment that is specified by the treaty itself, long-term

consequences that may be imposed outside of the treaty's formal institutions, both, or neither?[27]

We contend that this type of risk may affect preferences for international cooperation.

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".[28] 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".[29]

Hypothesis 1: Increasing a state's non-compliance risk dampens decision-makers'

willingness to join international agreements compared to compliance certainty.

Costs of Non-Compliance

The prospect of default may affect preferences for cooperation for a variety of reasons.

Perhaps the dominant theoretical explanation is that institutions can be designed to impose costs

---

[27] In addition to the consequences of non-compliance risk, one might also wonder about other
decision-making factors, such as decision-makers risk-aversion. While risk preferences and
perceptions, such as Arrow-Pratt risk aversion (O'Neill 2001) or the framing effects found in
Prospect Theory (McDermott 1992; 1998; McDermott, Fowler, and Smirnov 2008) are of
immense importance to international relations theory, they are not the focus of our paper. In our
empirical tests, we simply control for subjects' relative level of risk aversion to make sure those
types of risk preferences are not confounding our main results. With respect to Prospect Theory,
subjects only see one framing of the potential losses or gains from joining a treaty. So framing
effects cannot explain our results, though they might affect the generalizability of our findings.
For instance, subjects might be more risk accepting if they were facing the loss of a treaty they
were already in, instead of considering whether to gain membership a new treaty.
[28] Koremenos 2005, 549.
[29] Simmons and Hopkins 2005, 624.

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 formal 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 World Trade Organization—offer the prospect of imposing significant costs on defectors and thus may deter membership by countries that are not sure they can comply.[30]

Yet the impact of formal enforcement on incentives to cooperate is a source of debate because other forces may also be at work.[31] 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.[32] In such models, retaliation is credible because it is an equilibrium strategy of a repeated game[33], 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. Michael Tomz 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.[34] Beth Simmons has shown how IMF obligations raise expectations that create reputational costs through market

---

[30] Downs, Rocke, and Barsoom 1996; von Stein 2005.
[31] Thompson 2009.
[32] Fearon 1998; Greif, Milgrom, and Weingast 1994; Guisinger and Smith 2002.
[33] Bendor and Swistak 1997.
[34] Tomz 2007.

forces for governments that violate their commitments.[35] Andrew Guzman 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.[36] 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.[37]

Concern about reputation also resonates with several lines of scholarship that emphasize law-making as a process that works through the establishment of norms, practices, and legitimacy.[38] Law works, even in the absence of formal enforcement, because good standing membership in treaties has become expected behavior for modern nations.[39]

Thus we have contrasting models of what motivates governments to avoid international agreements in the face of non-compliance risk. One model emphasizes 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. Other models emphasize 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 tradeoffs between membership in an agreement and the potential consequences of non-compliance. Do they balk at the idea of international cooperation when they believe the

---

[35] Simmons 2000.
[36] Guzman 2008.
[37] Guisinger and Smith 2002.
[38] e.g., Brunnée and Toope 2010; Finnemore and Toope 2001; Franck 1990; Koh 1997.
[39] Chayes and Chayes 1998.

U.S. may not comply? Are they deterred by the prospect that the U.S. may be punished for non-compliance formally by the institution or informally by a loss of reputation or some other form of informal retaliation? We state the hypothesis in its simplest form:

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

Time Preferences

Cooperation often hinges on the shadow of the future. That is, states cooperate today so that they will gain a future stream of benefits. For cooperation to occur, leaders must value those future benefits more than the immediate gains from exploiting others or simply abstaining from cooperation. This valuation, we argue, depends strongly on how much leaders discount future gains relative to more immediate rewards.[40] Because these time preferences moderate the effect of many types of future informal penalties such as damage to reputation (but not of immediate formal punishment), they give us additional leverage in determining the extent to which informal

---

[40] e.g., Axelrod 1984; Bendor and Swistak 1997; Fearon 1998; Gilligan and Johns 2012; Greif, Milgrom, and Weingast 1994; Nowak 2006; Rand and Nowak 2013; Skyrms 2003.

mechanisms might explain why decision makers avoid agreements as the risks of non-compliance rise.

Much of the literature on cooperation has examined how time preferences affect cooperation via reciprocity in the infinitely repeated prisoners' dilemma.[41] A central insight is that cooperation is more probable when players place a higher value on future payoffs. A larger number of cooperative equilibria exist when players discount future payoffs less.[42] Recent experiments show that even when cooperation is an equilibrium strategy in the infinitely repeated prisoner's dilemma, cooperation is not guaranteed.[43] Instead, players must place a relatively high value on future payoffs for cooperation to emerge in controlled experiments.[44]

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 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.[45] In these games, uncertainty about one's ability to cooperate could lead to

---

[41] Though see Thompson 2009 and Tomz 2007 for arguments about why sanctioning and negative reciprocity might not affect cooperation in international relations.
[42] e.g., Ely 2002; Mailath and Samuelson 2006.
[43] Dal Bó 2005.
[44] Dal Bó and Fréchette 2011.
[45] Hauert et al. 2002.

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.

To give an example, if players play a grim-trigger strategy in the prisoners' dilemma[46] 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, δ 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.[47]

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

---

[46] Axelrod 1984; Bendor and Swistak 1997.

[47] This follows from the fact that the left-hand side of the inequality is positive and approaches infinity as δ approaches 1. The right-hand side is also positive, but finite. Therefore, the player will always prefer L to defecting in the nth round if δ is sufficiently close to 1.

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.[48]

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 does a decision maker value rewards in the distant future compared to rewards in the here and now? While previous literature has mapped the time-discounting parameter in existing models onto institutional features of a state, such as the expected longevity of a regime[49], patience is a trait that exists at the level of an individual decision maker. Empirically, this trait can vary substantially between individuals[50] but exhibits a great deal of stability over an

---

[48] Downs and Jones 2002; Downs, Rocke, and Barsoom 1996.
[49] e.g., see Guzman 2002.
[50] Coller and Williams 1999; Fowler and Kam 2006; Harrison, Lau, and Williams 2002; Mischel 1974.

individual's lifetime and across different environments.[51] 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 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.[52] 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 our third hypothesis:

> *Hypothesis 3: 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.*

All three of our hypotheses are about the preferences of individual decision makers. A

full blown theory of policy making, which is beyond the scope of this article, would need to

connect individual preferences to the ultimate decisions made by states—a process that is

---

[51] Casey et al. 2011; Meier and Sprenger 2010.
[52] Brewster 2009; Downs and Jones 2002; Guzman 2002; 2008; Kreps and Wilson 1982.

mediated by layers of intervening institutions that aggregate preferences as well as a variety of

institutional rules that affect freedom of choice. Nonetheless, there are at least two reasons why

a focus on individual policy elites is a useful way to begin understanding how collective state

decisions about non-compliance risk might be made. First, it is reasonable to suspect that the

individual preferences of policy elites shape state policies. Many policy decisions are made at the

"top" of government by a relatively small group of actors who often have substantial authority

and autonomy—they are not merely perfect agents for underlying interests and structures.[53]

Second, it is reasonable to assume that in many situations the relationship between individual

preferences and state policy making is weakly monotonic. That is, if a relevant set of foreign

policy decision makers want to take a particular action, that action becomes more likely (or at

least not less likely). In the cases where these two assumptions hold, one can (in a limited way)

extend our findings on individual choices to the choices adopted by states.[54] We thus follow an

already-large and growing literature that seeks to determine sources in the variation of individual

preferences for international cooperation.[55]

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,

---

[53] e.g., see Evans, Jacobson, and Putnam 1993; Hafner-Burton et al. 2014; Putnam 1988; Shamir and Shikaki 2005. This point is also emphasized by "insider" accounts of international negotiations. For example, on the 1992 Rio Conference see Brenton 1994; on the US-Canada free trade talks that were precursors to NAFTA see Hart 1994.

[54] Future research could look more closely at how individual elite-level decisions aggregate into collective policy and at how non-elite pressures, such as voter attitudes, affect that aggregation process.

[55] Mansfield and Mutz 2009; Milner and Tingley 2013a; 2013b; Rho and Tomz 2013; 2015; Tingley and Tomz 2012.

including former members of U.S. Congress, their senior staff, top U.S. 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 policy-making, business and NGO officials. A demographic profile of the sample is reported in the Supporting 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.

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, 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.[56] Consistent with this logic, we have reported in other studies that

---

[56] This type of situation is sometimes modeled formally as a global game where cooperative actions are strategic complements but where actors lack common knowledge of the payoff to cooperation and may even have different payoffs. Typically actors in these games play a

the presence of an enforcement mechanism does in fact make subjects initially more likely to join the agreement.[57]

Within the vignette, we randomly[58] varied whether or not the agreement was described as containing a formal enforcement mechanism.[59] 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.[60] 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 U.S trade decisions.[61] They judged the other two versions—one with automatic enforcement, the other with additional penalties such as financial transfers—

threshold strategy, where they cooperate if their signal about the payoff is high enough and if their uncertainty about others' incentive to defect is low enough.

[57] Hafner-Burton et al. 2014.

[58] A balance check is reported in the Supporting Information for key covariates.

[59] The full text of this vignette can be found in the Supporting Information.

[60] 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.

[61] Those decision makers did not take the subsequent survey.

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.

After presenting subjects with the vignette, we measured the likelihood that they would sign such an agreement.[62] 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% chance that their country would not be able to meet their 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 during design of the instrument (but did not use in our study) 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 un-avoidable amounts of measurement noise, which can dampen estimates of the true effect.[63] 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

---

[62] Respondents picked 1 of 5 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.
[63] Achen 1975; Ansolabehere, Rodden, and Snyder 2008.

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%, 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 supporting information) suggest that subjects were highly uncertain about compliance prior to the treatment.

After telling subjects that their country might not comply, we primed 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 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 Supporting Information) 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 towards concerns about enforcement.

Another risk 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 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 policy makers 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 such as concern for long-term reputation, reciprocity or adherence to cooperative norms.

Figure 1: Survey Experiment Design
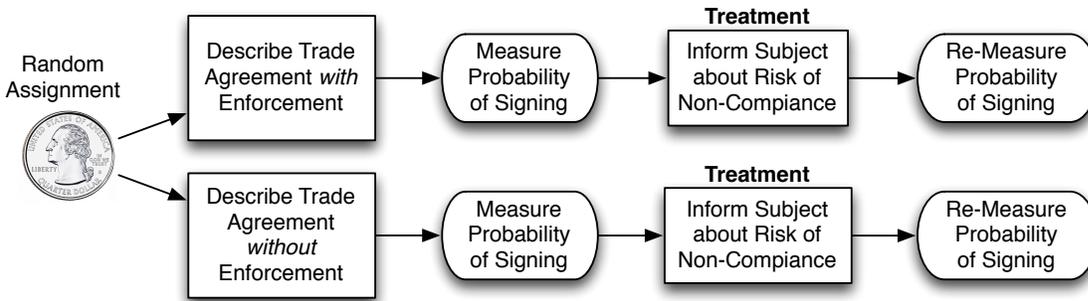
Measuring Time Preferences

To test our hypotheses on patience, we used standard tasks from behavioral decision-theory.[64] To measure how much subjects value the future – that is, their level of *patience* – we adapted a "choice game" introduced by Coller and Williams.[65] 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[66] and voting behavior[67]. In our study (as in others using this choice game), subjects were asked to make 20 different choices between a $100 prize that would be paid to them within 30 days after taking the study and a variable, larger prize that would be paid within 60 days. For each subject, a measure of patience is the number of 60-day choices. Time-indifferent, highly patient players will always choose the 60-day prize even if it is just a tiny bit larger than the $100 offered at 30 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 2 shows

[64] Camerer 2003.
[65] Coller and Williams 1999.
[66] Harrison, Lau, and Williams 2002.
[67] Fowler and Kam 2006.

the distribution of 60-day choices made by subjects in our study, which is similar to choices found in many other studies.[68]

Figure 2: Distribution of Elite Patience
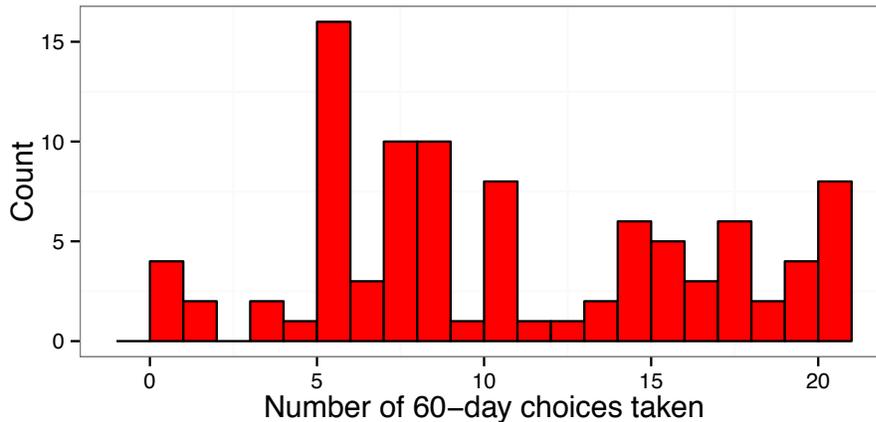
Compared to previous studies, which have looked at convenience samples of undergraduates[69] or the general public[70], our population of elites is more patient on average.[71] However, the distribution of patience among elites shares many features of the distributions found in convenience samples. Like 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.

---

[68] e.g., Coller and Williams 1999; Fowler and Kam 2006.
[69] Fowler and Kam 2006.
[70] Harrison, Lau, and Williams 2002.
[71] Hafner-Burton et al. 2014.

Non-Compliance Risk Deters Cooperation

Our subjects were less likely to prefer to join the agreement once given information that there was a 50% 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 2 categories (corresponding to a 0-20% and 21-40% chance of joining); the vast majority of respondents stated that there was a greater than 60% 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% chance they will still join.

We compared the mean difference between individual subjects' response to the original question and their response after being treated with non-compliance risk.[72] On average, subjects' stated propensity (on our 1-5 scale) decreases by 0.74, which roughly corresponds to a 15% decrease in the chance that an individual subject will state a preference to join the agreement (paired *t*-test p value $=3.8 \times 10^{-8}$).[73]

Table 1: Propensity to join (on a scale of 1-5) before and after non-compliance risk treatment

	(1) 0-20%	(2) 21-40%	(3) 41-60%	(4) 61-80%	(5) 81-100%

---

[72] The mean difference is appropriate, as each of the response categories, 1-5, are defined as equally spaced probability intervals. However, one can relax this assumption by using a non-parametric rank-based test. Using a Wilcoxon Ranksum test we also find a significant difference in the enforcement and non-enforcement condition (p-values $= 1.72 \times 10^{-6}$ and 0.002 respectively). The Hodges and Lehman estimator shows a median difference of -1.0 (95% ci -1.50, -1.0).

[73] 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.

	chance	chance	chance	chance	chance
Before Non-Compliance Risk	5%	3%	27%	47%	17%
After Non-Compliance Risk	13%	23%	26%	28%	9%

Institutional Enforcement is Not 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. Figure 3 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 represent 95% confidence intervals). In each case, the difference is negative and statistically significant (paired t-test p-values = 1.9×10^{-5}, 0.0006, 3.8×10^{-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 3 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% 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.

Figure 3: Effect of Risk of Future Default on Probability of Signing A Trade Agreement
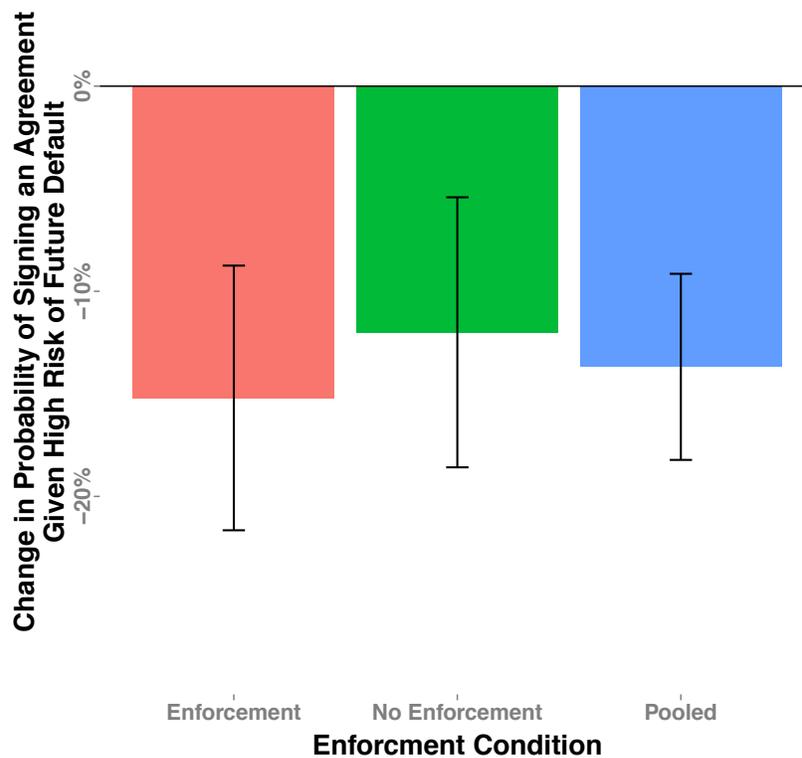
Patience Deters Cooperation As Non-Compliance Risks Rise

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 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% 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[74] regression with robust standard errors:[75]

$$y = \beta_1 Enforcement + \beta_2 Patience + \beta_3 RiskAversion + \beta_j InitialPropensity_j \qquad (1)$$

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$_j$ is a set of 5 dummy variables, which control for a subjects' initial propensity to join the agreement (prior to being treated with non-compliance risk). For example, if a subject stated that their initial propensity to join the agreement was "3" (indicating a 40-60% chance they would join the agreement), then *InitialPropensity$_3$* was coded as a 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 risk-aversion* further change decision makers' response to increasing the risk of non-compliance, independent of other treaty-related considerations.[76]

---

[74] OLS is appropriate since each category on our dependent variable represents an equally space 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.

[75] We also obtain a similar result if we look at the Spearman rank correlation between the difference in subjects' responses, $\Delta r = r_u - r_i$ (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).

[76] 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

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[77], which we describe in detail in the SI to economize on space.[78]

---

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.

[77] Holt and Laury 2002.

[78] 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,

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.

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.

	Enforcement	Traits	Combined
Enforcement	-0.19		-0.25
	(0.222)		(0.210)
Patience (normed)		-0.57**	-0.56**
		(0.191)	(0.203)
Risk Aversion (normed)		-0.58**	-0.61**
		(0.315)	(0.214)
Initial Propensity to Join:			
(1) 0-20%	2.35***	2.65***	2.87***
	(0.492)	(0.454)	(0.493)
(2) 21-40%	1.40*	1.19*	1.28*
	(0.598)	(0.559)	(0.529)
(3) 41-60%	2.40***	2.28***	2.40***
	(0.226)	(0.191)	(0.214)
(4) 61-80%	3.38***	3.24***	3.35***
	(0.182)	(0.146)	(0.171)

psychologists have validated our risk-aversion task as a measure that predicts decision makers' willingness to take risky decision 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 subject's 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.

(5) 81-100%	3.97***		3.87***	4.08***
	(0.314)		(0.242)	(0.296)
N	95		95	95
R2	0.90		0.92	0.92
Adj R2	0.90		0.91	0.91`

Standard errors in parentheses. * p<0.05 ** p<0.01 *** p<0.001

Columns 2 and 3 of Table 2 shows 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* roughly corresponds to an additional 11% 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%. Therefore the difference in the effect of risk between the most and least patient is about 70% as large as the average effect of risk itself. Likewise, the coefficient of -0.61 on *RiskAversion* roughly means that, compared to the least risk-averse subject in our sample, the most risk-averse person is 12% 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 definitive proof of) the idea that subjects are especially concerned about the long-term consequences, such as a potential loss of reputation, future retaliation or concern for normative obligation. 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 a larger estimated impact on subjects' responses to non-

compliance risk. It is beyond the scope of this experiment to distinguish whether it is concern for

reputation, long-term tit-for-tat, or other long-term concerns that explain the aversion—likely, all

could matter. However, our finding that the prospect of enforcement is not additionally scaring

decision makers away from cooperating is quite important in its own right and speaks to the

long-standing debate about selection bias in international organizations. Our behavioral finding

underscores this experimental finding, and further suggests that decision makers' concerns are

likely to be rooted in one of the many long-term—informal—concerns discussed by international

relations theorists.

It is, however, important to also point out that our approach, like any method, has its

relative limitations. The use of an experiment on real world elites enhances the internal validity

of our study by ensuring that our treatments are exogenous and that our subjects have the

experiential background to reason about risk and enforcement. However, in order to keep the

experiment tractable we could examine only one relatively simple decision-making scenario, and

study the effect of only one basic, but important institutional feature: formal enforcement. We

also restricted ourselves to analyzing a limited choice set. It would not have been feasible to

experimentally study many of the potential responses to non-compliance risk, such as allowing

each subject to submit their own proposal for an institutional design.

Large *N* observational studies typically give up exogeneity of the treatment and generally

cannot examine how different individuals react to the exact same decision-making scenario. Yet

these studies typically gain the ability to examine a wider variety of institutions designed to deal

with numerous forms of uncertainty and non-compliance risk.[79] In sum, we use an experiment to

---

[79] Koremenos 2005; Koremenos, Lipson, and Snidal 2001; Kucik and Reinhardt 2008;
Rosendorff 2005; Rosendorff and Milner 2001.

examine an important question, but are limited to examining a single element of a vast literature on institutional design.

Conclusions and Implications

The debate over whether institutional enforcement is desirable for international cooperation has been running for decades with few clean, empirical tests of the competing theories. From one view, states are opportunistically tempted to violate their international commitments; cooperation emerges when commitments are self-enforcing or when the risk of defection is offset by threats of formal enforcement. From another view, formal enforcement plays a much more ambiguous role because states rarely violate agreements out of pure self-interest. Some violations arise and must be "managed." And violations, overall, are few because governments know that under conditions of repeated interaction where reputation matters and tit-for-tat retaliation is possible it is best only to join agreements that can be honored. While these contrasting logics are equally plausible, none is easily observed and distinguished from the other with crisp precision in the real world.

This paper makes several contributions to this debate, and thus to the study of international relations. First, it provides some of the first direct elite-level evidence that increasing the risk of non-compliance with international obligations decreases real policy makers' willingness to cooperate by joining agreements. In the eyes of actual decision makers, early steps in iterated assurance games, typical of many domains of international cooperation, are to be avoided unless the government, itself, can honor its commitments. Elite decision makers are reluctant to make false promises. We have focused on elites in just one country to avoid cultural and legal confounds that will surely arise in other national settings; whether our findings apply to elites in other countries is a matter for further investigation.

Second, and contrary to what a leading strand of the literature would expect, we provide unique evidence that this aversion effect is not driven by the presence of a formal enforcement mechanism. Instead, 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.

Third, we have demonstrated that not all elite decision makers react alike to non-compliance risk. Patient people 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 may depend not so much on the enforcement structure of the agreement but rather on how decision makers' themselves weight the future.

References

Achen, Christopher H. 1975. Mass Political Attitudes and the Survey Response. *American Political Science Review* 69:1218–31.

Anderson, Lisa R., and Jennifer M. Mellor. 2008. Predicting Health Behaviors with an Experimental Measure of Risk Preference. *Journal of Health Economics* 27:1260–74.

Ansolabehere, Stephen, Jonathan Rodden, and James M. Jr. Snyder. 2008. The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting. *American Political Science Review* 102:215–32.

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* 38:226–54.

Bendor, Jonathan, and Piotr Swistak. 1997. The Evolutionary Stability of Cooperation. *American Political Science Review* 91:290–307.

Dal Bó, Pedro. 2005. Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games. *American Economic Review* 95:1591–604.

Dal Bó, Pedro, and Guillaume R. Fréchette. 2011. The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence. *American Economic Review* 101:411–29.

Brenton, Tony. 1994. *The Greening of Machiavelli: The Evolution of International Environmental Politics*. London: Earthscan.

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: Princeton University Press.

Casey, B. J., Leah H. Somerville, Ian H. Gotlib, Ozlem Ayduk, Nicholas T. Franklin, Mary K. Askren, et al. 2011. Behavioral and Neural Correlates of Delay of Gratification 40 Years Later. *Proceedings of the National Academy of Sciences* 108:14998–5003.

Chayes, Abram, and Antonia Handler Chayes. 1993. On Compliance. *International Organization* 47:175–205.

Chayes, Abram, and Antonia Handler Chayes. 1998. *The New Sovereignty: Compliance With International Regulatory Agreements*. Cambridge: Harvard University Press.

Chilton, Adam, and Dustin Tingley. 2013. Why the Study of International Law Needs Experiments. *Columbia Journal of Transnational Law* 52:173–238.

Coller, Maribeth, and Melonie B. Williams. 1999. Eliciting Individual Discount Rates. *Experimental Economics* 2:107–27.

DellaVigna, Stefano. 2009. Psychology and Economics: Evidence from the Field. *Journal of Economic Literature* 47:315–72.

Downs, George W., and Michael A. Jones. 2002. Reputation, Compliance and International Law. *Journal of Legal Studies* 31:S95–114.

Downs, George W., and David M. Rocke. 1995. *Optimal Imperfection?: Domestic Uncertainty and Institutions in International Relations*. Princeton: Princeton University Press.

Downs, George W., David M. Rocke, and Peter N. Barsoom. 1996. Is the Good News About Compliance Good News About Cooperation? *International Organization* 50:379–406.

Edelstein, David M. 2002. Managing Uncertainty: Beliefs about Intentions and the Rise of Great Powers. *Security Studies* 12:1–40.

Ely, Jeffrey C. 2002. Local Conventions. *Advances in Theoretical Economics* 2:1534–5963.

Evans, Peter B., Harold K. Jacobson, and Robert D. Putnam, eds. 1993. *Double-Edged Diplomacy: International Bargaining and Domestic Politics*. Berkeley: University of California Press.

Fearon, James D. 1998. Bargaining, Enforcement, and International Cooperation. *International Organization* 52:269–305.

Finnemore, Martha, and Stephen J. Toope. 2001. Alternatives to 'Legalization': Richer Views of Law and Politics. *International Organization* 55:743–58.

Fowler, James H., and Cindy D. Kam. 2006. Patience as a Political Virtue: Delayed Gratification and Turnout. *Political Behavior* 28:113–28.

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* 58:459–84.

Gilligan, Michael J., and Leslie Johns. 2012. Formal Models of International Institutions. *Annual Review of Political Science* 15:221–43.

Greif, Avner, Paul Milgrom, and Barry R. Weingast. 1994. Coordination, Commitment, and Enforcement: The Case of the Merchant Guild. *Journal of Political Economy* 102:745–76.

Guisinger, Alexandra, and Alastair Smith. 2002. Honest Threats: The Interaction of Reputation and Political Institutions in International Crises. *Journal of Conflict Resolution* 46:175–200.

Guzman, Andrew T. 2002. A Compliance-Based Theory of International Law. *California Law Review* 90:1823–88.

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

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* 11:368–86.

Hafner-Burton, Emilie M., Brad L. LeVeck, David G. Victor, and James H. Fowler. 2014. Decision Maker Preferences for International Legal Cooperation. *International Organization* 68:845–76.

Harrison, Glenn W., Morten I. Lau, and Melonie B. Williams. 2002. Estimating Individual Discount Rates in Denmark: A Field Experiment. *American Economic Review* 92:1606–17.

Hart, Michael. 1994. *Decision at Midnight Inside the Canada-US Free Trade Negotiations*. Vancouver B.C.: UBC Press.

Hauert, Christoph, Silvia De Monte, Josef Hofbauer, and Karl Sigmund. 2002. Volunteering as Red Queen Mechanism for Cooperation in Public Goods Games. *Science* 296:1129–32.

Holt, Charles A., and Susan K. Laury. 2002. Risk Aversion and Incentive Effects. *American Economic Review* 92:1644–55.

Iida, Keisuke. 1993. Analytic Uncertainty and International Cooperation: Theory and Application to International Economic Policy Coordination. *International Studies Quarterly* 37:431–57.

Jarvis, Darryl S. L. 2011. Theorising Risk and Uncertainty in International Relations The Contributions of Frank Knight. *International Relations* 25:296–312.

Kertzer, Joshua D. 2013. Making Sense of Isolationism: Foreign Policy Mood as a Multilevel Phenomenon. *The Journal of Politics* 75:225–40.

Kertzer, Joshua D., and Kathleen M. McGraw. 2012. Folk Realism: Testing the Microfoundations of Realism in Ordinary Citizens. *International Studies Quarterly* 56:245–58.

Koh, Harold. 1997. 'Why Do Nations Obey International Law?'. Faculty Scholarship Series Paper 2101, Yale Law School.

Koremenos, Barbara. 2005. Contracting around International Uncertainty. *American Political Science Review* 99:549–65.

Koremenos, Barbara, Charles Lipson, and Duncan Snidal. 2001. The Rational Design of International Institutions. *International Organization* 55:761–99.

Kreps, David M., and Robert Wilson. 1982. Sequential Equilibria. *Econometrica* 50:863–94.

Kucik, Jeffrey, and Eric Reinhardt. 2008. Does Flexibility Promote Cooperation? An Application to the Global Trade Regime. *International Organization* 62:477–505.

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* 111:18536–41.

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* 63:425–57.

Martin, Lisa L. 1994. Heterogeneity, Linkage and Commons Problems. *Journal of Theoretical Politics* 6:473–93.

Martin, Lisa L. 2013. Against Compliance. Pp. 591–612 in *Interdisciplinary Perspectives on International Law and International Relations: The State of the Art*, edited by Jeffrey L. Dunoff and Mark A. Pollack. New York: Cambridge University Press.

McDermott, Rose. 1992. Prospect Theory in International Relations: The Iranian Hostage Rescue Mission. *Political Psychology* 13:237–63.

McDermott, Rose. 1998. *Risk-Taking in International Politics: Prospect Theory in American Foreign Policy*. Ann Arbor: University of Michigan Press.

McDermott, Rose. 2004. *Political Psychology in International Relations*. Ann Arbor: University of Michigan Press.

McDermott, Rose. 2011. New Directions for Experimental Work in International Relations. *International Studies Quarterly* 55:503–20.

McDermott, Rose, James H. Fowler, and Oleg Smirnov. 2008. On the Evolutionary Origin of Prospect Theory Preferences. *The Journal of Politics* 70:335–50.

McDermott, Rose, and Yoshiko Herrera. 2010. Psychological Versus Rational Models of Human Decision-Making. *APSA Comparative Politics Newsletter* 21:21–3.

Meier, Stephan, and Charles Sprenger. 2010. 'Stability of Time Preferences'. Discussion Paper No. 4756, Bonn, Germany: IZA.

Milner, Helen V., and Dustin Tingley. 2013a. Public Opinion and Foreign Aid: A Review Essay. *International Interactions* 39:389–401.

Milner, Helen V., and Dustin Tingley. 2013b. The Choice for Multilateralism: Foreign Aid and American Foreign Policy. *Review of International Organizations* 8:313–41.

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* 50:757–76.

Mintz, Alex, Yi Yang, and Rose Mc Dermott. 2011. Experimental Approaches to International Relations. *International Studies Quarterly* 55:493–501.

Mischel, Walter. 1974. Processes in Delay of Gratification. Pp. 249–92 in *Advances in Experimental Social Psychology*, vol. 7, edited by Leonard Berkowitz. New York: Academic Press.

Nowak, Martin A. 2006. Five Rules for the Evolution of Cooperation. *Science* 314:1560–3.

O'Neill, Barry. 2001. Risk Aversion in International Relations Theory. *International Studies Quarterly* 45:617–40.

Putnam, Robert. 1988. Diplomacy and Domestic Politics: The Logic of Two-Level Games. *International Organization* 41:427–60.

Rand, David G., and Martin A. Nowak. 2013. Human Cooperation. *Trends in Cognitive Sciences* 17:413–25.

Rathbun, Brian C. 2007. Uncertain about Uncertainty: Understanding the Multiple Meanings of a Crucial Concept in International Relations Theory. *International Studies Quarterly* 51:533–57.

Renshon, Jonathan, Julia Lee, and Dustin Tingley. 2014. Physiological Arousal and Political Beliefs. *Political Psychology*.

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.

Rho, Sungmin, and Michael Tomz. 2015. 'Why Don't Trade Preferences Reflect Economic Self-Interest?'. Working Paper, Stanford University.

Rosendorff, B. Peter. 2005. Stability and Rigidity: Politics and the Design of the WTO's Dispute Resolution Procedure. *American Political Science Review* 99:389–400.

Rosendorff, B. Peter, and Helen V. Milner. 2001. The Optimal Design of International Trade Institutions: Uncertainty and Escape. *International Organization* 55:829–57.

Shamir, Jacob, and Khalil Shikaki. 2005. Public Opinion in the Israeli-Palestinian Two-Level Game. *Journal of Peace Research* 42:311–28.

Simmons, Beth A. 2000. International Law and State Behavior: Commitment and Compliance in International Monetary Affairs. *American Political Science Review* 94:819–35.

Simmons, Beth A., and Daniel J. Hopkins. 2005. The Constraining Power of International Treaties: Theories and Methods. *American Political Science Review* 99:623–31.

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: Princeton University Press.

Thompson, Alexander. 2009. The Rational Enforcement of International Law: Solving the Sanctioners' Dilemma. *International Theory* 1:307–21.

Tingley, Dustin. 2011. The Dark Side of the Future: An Experimental Test of Commitment Problems in Bargaining. *International Studies Quarterly* 55:521–44.

Tingley, Dustin, and Michael Tomz. 2012. 'How Does the UN Security Council Influence Public Opinion'. Working Paper, Stanford University and Harvard University.

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

Tomz, Michael. 2009. The Foundations of Domestic Audience Costs: Attitudes, Expectations and Institutions. Pp. 85–97 in *Kitai, Seido, Gurobaru-Shakai (Expectations, Institutions, and Global Society)*, edited by Masaru Kohno and Aiji Tanaka. Tokyo: Keiso-Shobo.

 Von Stein, Jana. 2005. Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance. *American Political Science Review* 99:611–22.