Selective Enforcement and Revealed Preferences in International Security Organizations[*]

Jeffrey M. Kaplow[†]

Paper prepared for the ninth annual conference on the Political Economy of International Organizations (PEIO)

Abstract

International security institutions often lack formal enforcement mechanisms, but this does not mean that violations of international commitments always go unpunished. States that violate treaties routinely face pressure from others to change their behavior, including through sanctions and military attack. But the lack of formal enforcement measures probably does contribute to significant variation in which states are targeted for punishment—enforcement is nearly always at the discretion of the punishing state. Why do some states face punishment while the transgressions of others are overlooked? I argue that enforcing states look to the policy preferences of violators for signals about the likelihood that enforcement will change state behavior and the cost to the international community of allowing the violation to continue. Patterns of institutional membership within a larger regime help to credibly reveal the preferences of state parties. I use data on membership in the various agreements that make up the nuclear nonproliferation regime to derive a new measure of state preferences over nonproliferation policy issues, applying an item-response theory model of the type sometimes used to analyze political ideology in national legislatures. I show that a state's pattern of treaty memberships within the regime significantly affects the likelihood that the international community will pursue costly enforcement measures if the state seeks nuclear weapons. The findings in this paper suggest that state interests—as revealed by treaty adherence—can illuminate important new dynamics in international cooperation, with implications for existing theory on the role of international security institutions in constraining state behavior.

[*] This research was supported by the Nuclear Security Dissertation Fellowship at the University of California's Institute on Global Conflict and Cooperation.

[†] College of William & Mary, Department of Government; jkaplow@wm.edu

Most international security institutions lack formal mechanisms for punishing state violators, but states routinely act to exert pressure on cheaters to come back into compliance with their commitments and to protect the reputation of the institution against future, would-be violators. A lack of formal enforcement mechanisms, however, does put the decision to enforce a treaty's provisions almost entirely at the discretion of individual states. International security institutions thus enjoy selective enforcement—some violations are punished while others are overlooked.

What factors explain the variation we see in state decisions to engage in costly enforcement of international security institutions? Why are some states more likely than others to be punished for their violations? In deciding whether to punish or not, I argue that states consider the policy preferences of the violating state, as revealed by its pattern of treaty memberships within a larger security regime. Taken together, the constituent parts of a regime effectively screen for a state's preferences, sending valuable signals to others about the likely success of enforcement action and the ultimate cost of the violation to the punishing state.

I apply this theory to the case of the nuclear nonproliferation regime. Using new data on states' embeddedness within the regime, I employ an item response theory model to derive a measure of the latent nonproliferation policy space underlying membership in international treaties and organizations. I deploy this measure to examine how the information provided by the regime about a state's nonproliferation policy preferences affects the international community's response when a state chooses pursue nuclear weapons. Patterns of membership within the regime can reassure potential adversaries about the state's ultimate intentions. States with underlying preferences more strongly in

support of nonproliferation policy goals, as revealed by their treaty memberships, are less likely to be targeted for economic sanctions or for attacks against nuclear facilities. These results stand in contrast to the expectations of several important strands of IO scholarship, which see violation of treaty commitments as a greater sin than bad behavior outside of a treaty, and thus more deserving of costly enforcement measures.

This paper makes several contributions to the literatures on international security and international institutions. First, I theorize about an important information mechanism brought about by regime complexity—the system of nested, overlapping agreements and institutions that characterize some areas of international cooperation. This helps to explain why we see regime complexity in the first place. If complexity reduces uncertainty around state preferences and facilitates international cooperation, then a regime complex may well be rationally designed for that purpose. Second, I propose a mechanism by which international institutions matter, even if they are merely screening for state preferences. The IO literature often treats screening institutions as having no real effect on state behavior; this paper shows how multiple screening institutions, taken together, can provide information that affects outcomes of interest in international relations. Finally, I introduce a new measure of latent nonproliferation policy preferences—the nonproliferation score—that has broad utility in international security scholarship. Membership in the Treaty on the Non-Proliferation of Nuclear Weapons (NPT) is commonly used as an independent variable in the international security literature, but this dichotomous measure provides much less information than a state's pattern of institutional memberships within the wider nonproliferation regime.

This paper is divided into four parts. First, I present a theory of selective enforcement in international security institutions, deriving hypotheses about how the underlying preferences of states influence the use of costly enforcement measures when states violate international commitments. Second, I construct a new measure of states' latent nonproliferation policy preferences and explore the relationship between these nonproliferation scores and other measures of regime embeddedness. Third, I explain the results of several statistical models that test my hypotheses in the context of the nuclear nonproliferation regime. Finally, I conclude with a discussion of the implications of this study for the international security and international organizations literatures.

Selective enforcement and state preferences

When states fail to abide by their agreements, they may or may not incur some sort of punishment. Even in institutions with formal and well-specified enforcement mechanisms, punitive measures can be unevenly and sporadically applied. Stone (2004; 2008) finds that US intervention can compromise enforcement of International Monetary Fund conditionality, amounting to a kind of informal governance that exists in parallel to the Fund's formal enforcement mechanisms. Regional intergovernmental organizations, too, appear to selectively enforce democratic norms. They are more likely to act when election observers are involved, but less likely to target more influential states (Donno 2010; 2013).

When enforcement action falls to individual states, as it almost always does in international security institutions, the application of enforcement measures becomes even more varied. Two states committing the same offense may be subject to different enforcement measures, or one state may be punished while the other is not. The

underlying preferences of states may account for some of the variation we see in the enforcement of treaty provisions and the punishment of violators. Information about states' underlying preferences in a particular policy area can be revealed by their pattern of memberships in international agreements that are part of a larger regime.

The multiple institutions that make up a regime are both a curse and a blessing for international cooperation. Regime complexity is inefficient, forcing states to spread resources and foreign policy attention across organizations, and, as a result, increases the risk of agency slack and other organizational pathologies (Alter and Meunier 2009; Barnett and Finnemore 1999; Drezner 2009). Complexity introduces the possibility of forum shopping and makes it easier for states to shirk on international obligations, and to avoid punishment for doing so (Busch 2007; Hafner-Burton 2009; Helfer 1999; Kelley 2009; Raustiala and Victor 2004). Regime complexity dilutes the power of international law, complicating precedent-setting and making compliance with international institutions more difficult even for states that desire to comply (Alter and Meunier 2009).

All of this is balanced against an important but largely unrecognized benefit of regime complexity: a regime complex makes for a much more effective screening device than a single comprehensive institution, and can provide a useful signal of a state's underlying preferences in a particular policy domain. A single agreement within a policy space can tell observers only so much. Do states prefer some new policy to the status quo, or not? But multiple agreements in the same policy space can help distinguish between different gradations of state preferences.

The IO literature has long recognized that membership in an international institution can credibly commit states to comply with the treaty's rules. If joining the

institution is sufficiently costly *ex ante*, membership can function as a separating equilibrium, distinguishing the states that intend to comply from those that do not (Abbott and Snidal 2000; Simmons 2000, 2010; von Stein 2005). The discussion of screening institutions in the literature largely focuses on the implications for state compliance and the larger issue of treaty effectiveness (Downs, Rocke, and Barsoom 1996; Simmons and Hopkins 2005; von Stein 2005), but institutions that screen may also provide information that affects outcomes in international relations beyond treaty compliance. The effect of screening institutions may be particularly important in the context of the wider regime complex. If the agreements that make up a regime complex each carry some *ex ante* cost, a state's pattern of membership within the complex is likely to paint a more complete picture of the state's type, or, alternatively, of the state's interests within this area of international policy.

The literature on regime complexity is substantively diverse, examining the effects of institutions in issue areas such as the environment (Keohane and Victor 2011), energy (Colgan, Keohane, and Van de Graaf 2012), human rights (Hafner-Burton 2009), and international trade (Busch 2007; Davis 2009). Relatively understudied, however, is regime complexity in the realm of international security. Existing work on security regime complexes addresses the overlap between NATO and the European Security and Defense Policy (Hofmann 2009), but does not examine the much denser network of agreements, organizations, and institutions dealing with nuclear proliferation.[1]

[1] Regime complexity exists in other international security issue areas as well, including among arms control agreements and the various regional security organizations. My emphasis on the nuclear nonproliferation regime in this paper thus does not represent an exhaustive treatment of regime complexity in international security, but merely a first step in tying international security institutions more closely to the wider body of IO theory.

Screening for state policy preferences

The differing policy aims of the organizations and institutions that make up a regime complex can provide information about state preferences within that issue area. In some circumstances, a single treaty can screen states according to their interests: states that agree with the policy goals of a particular treaty will join, while those that disagree with the goals of the treaty will abstain. But a single treaty is capable of providing only a binary signal—whether or not the state supports the policy represented by the agreement. When multiple treaties exist within a policy space, however, much more information is potentially available. Each individual treaty can still only send a binary signal about state interests, but these separate pieces of information can be combined to provide a more precise understanding of a state's preferred position within the policy space.

To see this, imagine three states, *A* through *C*, arrayed along a single policy dimension according to the maximum amount of a policy the states desire in some issue area (depicted in Figure 1). State *A* desires less of a policy (such as free trade or human rights enforcement) than state *B*, which desires less than state *C*. For simplicity, assume policies in this issue area are complementary, so there is no tradeoff between a policy and the status quo, and states prefer any agreement that enacts policy at their maximum level or lower. If institution I_1 is proposed, enacting policy p_1, each of the parties will join. This is analogous to a weak treaty with a large membership. Institution I_2, at policy p_2, entices *B* and *C* to join, but not *A*, since p_2 exceeds the maximum level of policy desired by *A*. Similarly, I_3 is joined by *C* only. In this example, a single agreement tells us little about

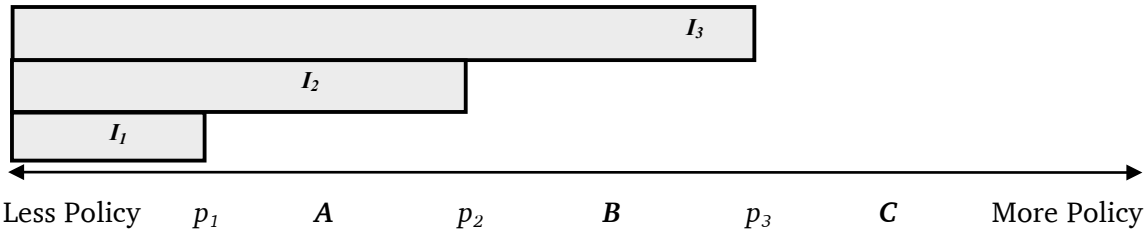

Figure 1: Regime complexity and state policy preferences

the preferences of the states. After I_1 is instituted, we know only that the maximum policy desired by all of the parties is greater than p_1, but we have no idea of their preference ordering or magnitude. When complexity is increased, and I_2 and I_3 are added, we have much more information about the maximum policy desired by each state.

In theory, then, regime complexity can screen states sufficiently to reveal state preferences. This reduced uncertainty about state interests in a particular issue area can lead to better interstate bargains and greater cooperation both within and outside the regime. In practice, of course, the situation is more complicated than the example above. States choose to join treaties for a variety of reasons, not all of them related to their views about the policy outcomes the institution represents, and the policies represented by an IO are often ambiguous. Still, the fact that there is complexity within a regime can be empirically revelatory. Even if sufficient information is not available to pin down the policy ideal point enacted by a particular component of a regime, states can learn something about the preferences of others by observing which states join which parts of which regimes.

This dynamic is illustrated by the set of nuclear safeguards agreements within the nonproliferation regime complex. States that have joined the NPT commit to reaching a safeguards agreement with the International Atomic Energy Agency (IAEA) that governs

monitoring and verification of domestic nuclear facilities. But member states have several options when it comes to signing safeguards agreements. States can, first, neglect to sign a safeguards agreement at all. Despite the requirement that a safeguards agreement enter into force within 18 months of signing the NPT, the average delay is nearly 10 years. As recently as 1995, only 55 percent of NPT member states had a comprehensive safeguards agreement in force, although total adoption had risen to 87 percent by 2010.

For those states that choose to sign, there are several types of safeguards agreements that vary in their level of obligation and the amount of access that is given to IAEA inspectors. A standard comprehensive safeguards agreement grants the IAEA access to nuclear facilities for the purposes of verifying state declarations about its nuclear activities. Beginning in 1997, as a response to undeclared nuclear work discovered in Iraq during the 1990–1991 Gulf War, member states were encouraged also to bring into force an Additional Protocol to their safeguards agreement that provides the IAEA with more complete access. The Additional Protocol includes the requirement to declare nuclear facilities and permit inspections there even when nuclear material is not present, and allows for the wider use of environmental sampling to provide assurances that nuclear material has not been introduced at undeclared sites (Hirsch 2004).

States with limited nuclear infrastructure may choose to sign the Small Quantities Protocol (SQP) instead of a standard comprehensive safeguards agreement. The SQP largely eliminates declaration requirements and inspector access to facilities within the state. Recognizing the potential loophole the SQP affords, the IAEA in 2005 created a

modified version of the SQP that allows the IAEA to verify that a state qualifies for this reduced level of scrutiny (International Atomic Energy Agency 2005).[2]

Because the Additional Protocol can be signed by states with either a standard comprehensive safeguards agreement or a small quantities protocol, there are a total of 7 combinations of safeguards arrangements that can be adopted by states.[3] The different configurations of these agreements help to distinguish the preferences of member states with regard to verification under the NPT and access by international inspectors, as shown in Figure 2. As the figure illustrates, the nuclear nonproliferation regime has added these agreements over time. The international community today thus has a great deal more information about the preferences of member states for monitoring and verification than it did 20 years ago, before the introductions of the Additional Protocol and the modified SQP.

Information and costly enforcement

If regime complexity provides information about the policy preferences of member states in a given domain, then we would expect this information to affect a wide range of strategic interactions in international relations. This information may be especially salient, however, when states are seen as violating their commitments within the regime;

[2] Then-IAEA Director-General Mohammed El-Baradei recommended eliminating this type of safeguards agreement altogether, but his proposal was not accepted by the IAEA's Board of Governors (Kerr 2005a).

[3] The options are, in increasing order of strength of verification: No safeguards agreement, SQP, modified SQP, SQP with Additional Protocol, modified SQP with Additional Protocol, comprehensive safeguards agreement, and comprehensive safeguards agreement with Additional Protocol. There are state adherents to each of these combinations.

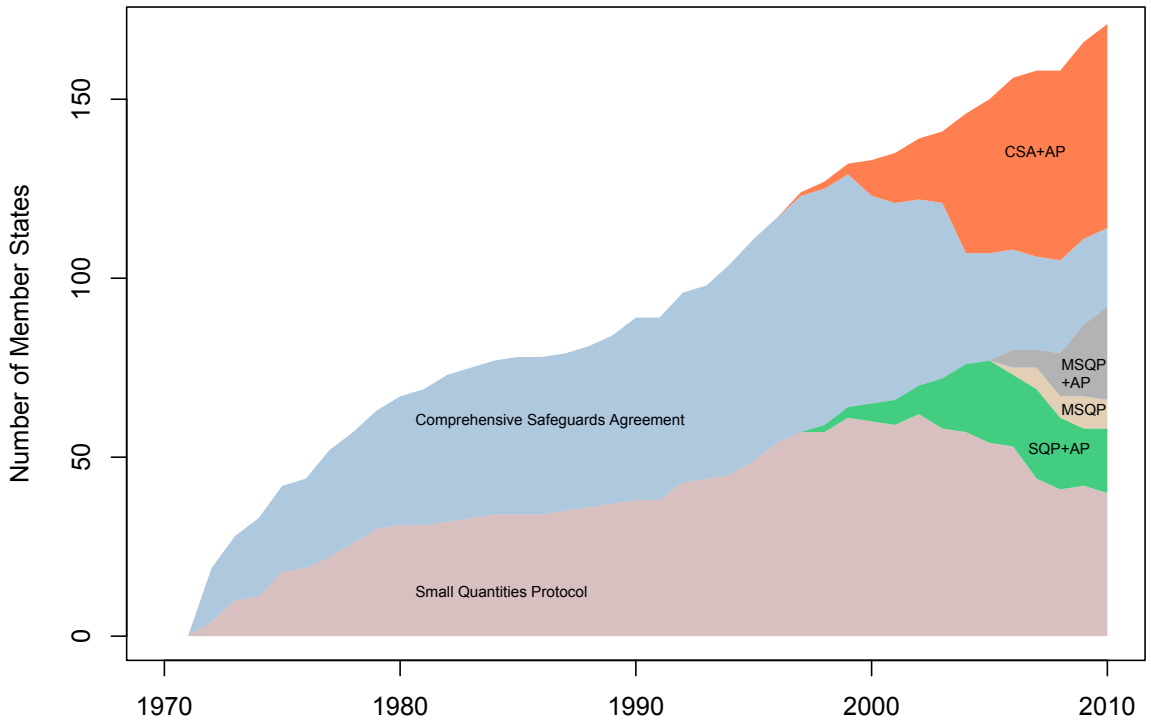

Figure 2: Safeguards agreements in the nuclear nonproliferation regime

the violator's revealed preferences in this area are particularly relevant for understanding its intentions and determining an appropriate foreign policy response.

The revealed preferences of the violating state are likely to affect the calculus of enforcement in two ways. First, patterns of regime membership can influence perceptions of the state's *resolve* in the violation of treaty provisions and thus the likelihood that enforcement efforts will succeed in changing the state's behavior. States that have been good nonproliferation citizens overall may be seen as more easily convinced to change their policies with less costly positive inducements or diplomatic pressure.

Second, membership in the regime complex may lead to particular judgments about the *intentions* of the violating state and thus the ultimate cost the international

community must bear for the state's treaty violation. In the international security context, the violation of treaties can lead to long-term changes in a state's military capabilities that may be seen as more or less threatening by others.[4] States that are pursuing nuclear weapons, for example, might be seen as less likely to pass sensitive technologies to a third party, or to use their weapons to challenge the prevailing international system, if they had been deeply embedded in the nuclear nonproliferation regime. A positive signal about the underlying preferences of treaty violators can help to convince other states that costly enforcement measures are not necessary.

Violators with only shallow regime membership, however, boast no such mitigating factor, and their patterns of international agreements within the regime complex reflect very different underlying policy preferences. Such states may require costly enforcement action before they are likely to change their policies and come back into compliance. If these states do persist in their noncompliance, they may incur greater costs for other regime members or the international community at large. If these states acquire nuclear weapons, for example, they may be more likely to act as revisionist powers, drawing on their newfound military capabilities to threaten the international status quo, or to share sensitive nuclear technology with other parties.

The effect of revealed preferences on perceptions of state resolve and intentions suggest the following hypothesis:

> *Preferences hypothesis:* States that violate an international regime are less likely to be targeted for costly enforcement action the more their underlying preferences support the policy goals of the regime.

[4] This dynamic is not exclusive to security regimes. Violations of trade, intellectual property, and environmental regimes may be more or less costly to other states depending on the intentions of the violator.

States also use information provided by the screening mechanism of the regime complex as a means of identifying similarities in underlying interests. The similarity of state preferences between the would-be enforcer state and the violating state may affect perceptions of both the likelihood of successful enforcement and the ultimate cost if enforcement fails. The violating state may be more receptive to lower cost enforcement strategies if they are initiated by a state with similar underlying interests. Diplomatic pressure or positive inducements are more effective when they come from a natural ally.

More importantly, perhaps, preference similarity probably leads to more favorable judgments about the intentions of violating states. Two states that have joined a similar portfolio of nonproliferation agreements, for example, may see each other as less threatening because they are likely to pursue similar foreign policy interests. This may be true even if a violating state's underlying preferences indicate that it largely rejects nonproliferation goals. This logic implies that the effects of regime complexity will be dyadic in nature, as in the following hypothesis:

> *Similarity hypothesis:* States that violate an international regime are less likely to be targeted for costly enforcement action the more similar their underlying preferences are to those of the enforcing state.

These hypotheses represent something of a hard case for the screening mechanism of the regime complex, in the sense that violation of a treaty is itself a strong signal of a state's type, providing a significant amount of information about the state's intentions and its stance toward the international status quo. Noncompliance with international agreements is generally seen as unequivocally bad from the standpoint of the regime's other members, and so the treatment of violators is a much less ambiguous situation than many encountered in international politics. But if the information provided by regime complexity matters in cases of treaty violations, where the state's actions

already send a strong signal about its policy preferences, then it is likely to matter elsewhere as well.

The above hypotheses are also notable for their contrast with more conventional views of violation and enforcement in international institutions. From the perspective of international legal scholarship, committing to but then violating an international agreement is a greater sin than merely behaving badly outside the agreement (Guzman 2002). Only in the former case has the state actually broken the law, and thus we would expect stronger enforcement action against violators than we would against abstainers that pursue the same policy. Similarly, mechanisms of compliance that rely on reciprocal enforcement theorize that punishment will be more likely when a state has joined but then violated a treaty. Morrow, for example, argues that "states should be more likely to respond to violations that breach these legal 'bright lines' than to acts that are not clearly illegal" (Morrow 2007, 561). These perspectives suggest that a deeper level of embeddedness in a regime will make violators more likely to be targeted for costly enforcement or punishment.

The nuclear nonproliferation regime complex

I evaluate my theory using the case of the nuclear nonproliferation regime. Though infrequently examined in the IO literature, the nuclear nonproliferation regime is a full-blown regime complex, consisting of a series of institutions, treaties, agreements, protocols, and conventions, with varying levels of institutionalization, monitoring and verification, and state membership. Agreements within the regime govern the pursuit of nuclear weapons, both globally and within particular regions; the safeguarding of civilian nuclear infrastructure; the commercial exchange of nuclear technologies; and the

physical protection of nuclear material generally. Nearly all states participate in this regime complex in some way, although the level of embeddedness in the regime varies significantly by state and over time.

Examining my hypotheses requires some measure of states' nonproliferation preferences. One option is a simple count of the number of treaties and institutions within the regime that a state has joined. In the nonproliferation regime complex, however, the opportunity to join treaties is unevenly distributed. A number of agreements are limited to the P-5 nuclear weapons states, for example, and others are restricted by region, nuclear supplier status, or ratification of a particular treaty. Simply tallying a state's agreements within the regime penalizes those states for which fewer institutions are available to join.

A more reasonable measure is the share of eligible agreements joined; that is, the number of nonproliferation treaties of which a state is a member, divided by the number of nonproliferation treaties the state is eligible to join. Figure 3 uses this measure to track the average level of state membership in the regime, using data on 22 nonproliferation and nuclear security agreements from Carcelli et al. (2014). The solid line represents the mean regime membership for all states, measured as the share of eligible agreements joined (left axis). The dashed line is the average number of agreements that states were eligible to join in each year, while the dotted line is the average number of agreements actually joined (right axis). As the figure shows, the share of agreements within the regime to which states are a party has remained relatively constant over time, even as the total number of agreements has increased dramatically. The rate of joining has largely kept pace with the rate at which new agreements are created.

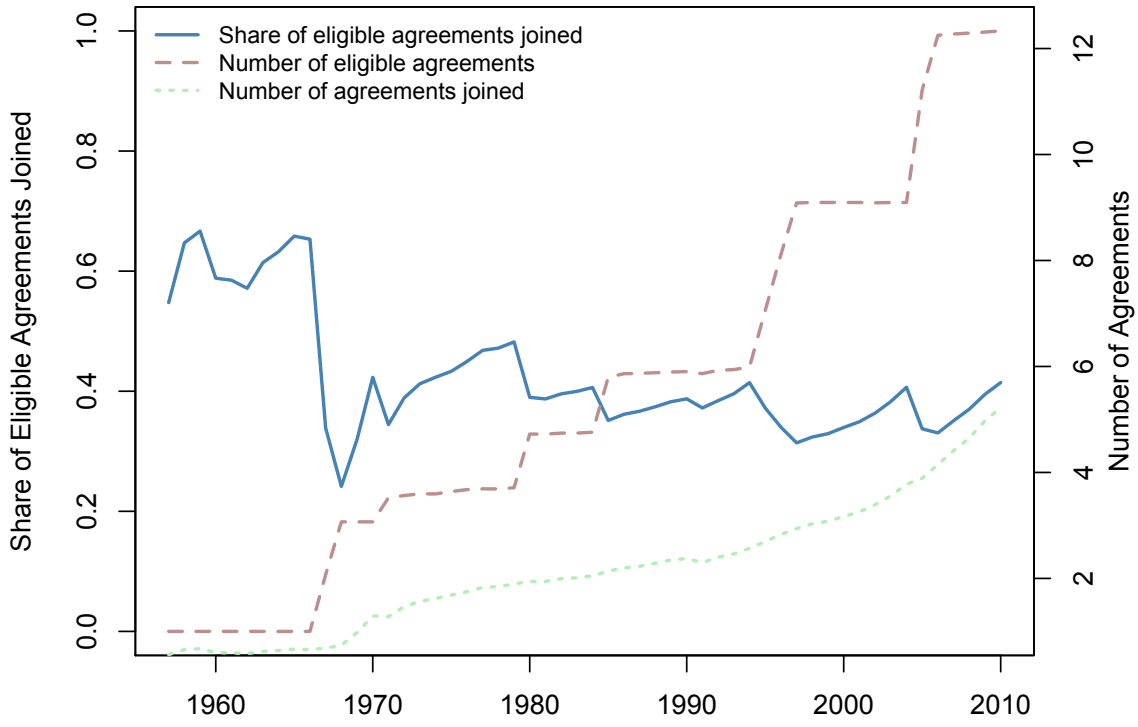

Figure 3: Mean membership in the nuclear nonproliferation regime over time

This way of thinking about regime embeddedness undoubtedly captures an important relationship between treaty membership and policy preferences; more treaties joined roughly equates to stronger support for nonproliferation policy goals. But as a proxy for the underlying nonproliferation preferences of member states, this simple additive measure has some shortcomings. This measure assumes, for example, that the same amount of information is provided by each decision to join or abstain from a treaty—that no one treaty is more informative than others. This assumption is questionable on its face. The NPT is widely seen as the cornerstone of the regime, while other agreements—Protocol I of the Treaty of Rarotonga, for example, which outlaws the manufacture, stationing, or testing of nuclear weapons by the P-5 states on their

territories within the South Pacific Nuclear Weapons Free Zone—are little known in most capitals. If not all treaties provide the same level of information about member state preferences, then a simple additive measure of regime embeddedness will overstate the importance of some agreements and understate the importance of others.

This measure also is censored at 0 and 1 in a way that makes it difficult to distinguish between states that merely lack the opportunity to join treaties, and those that would prefer to abstain. Consider two states that have not ratified any agreements in the nonproliferation regime complex. One of these—Costa Rica in 1960, for example—is only eligible for a single treaty, while another, such as India in 2000, is eligible to join 7 treaties. These states have the same share of eligible treaties joined, but their circumstances raise the possibility that there are differences in their underlying preferences.

To address these problems, I construct a latent variable model of the nuclear nonproliferation policy space. Ideal point models have long been used to estimate legislator ideology from roll call votes (Poole and Rosenthal 1991). More recently, Bayesian item response theory (IRT) models have emerged as an important tool in the analysis of roll call voting (Clinton, Jackman, and Rivers 2004).[5] To recover states' latent nonproliferation policy preferences, I use a dynamic binary IRT model that allows state preferences in a given year to depend on preferences in a previous year (Martin and Quinn 2002; Schnakenberg and Fariss 2014). The IRT model addresses the shortcomings

[5] Applications in international relations are still relatively rare. For recent examples, see Fariss (2014); Hug and Lukács (2014); Schnakenberg and Fariss (2014); and Treier and Jackman (2008).

of additive measures discussed above: the model allows the importance of individual treaties to vary and does not throw out information at the extremes.

This approach theorizes that the decision to join a particular treaty is a function of the state's underlying nonproliferation policy preferences—the latent variable we are attempting to measure. Of course, there are a number of factors that might lead states to join a particular agreement, but there is some reason to believe that policy preferences play an important role. First, most of the treaties within the nonproliferation regime carry some *ex ante* cost for a state to join and come into compliance, in the form of reporting requirements, verification measures, or representation within the institution. For many states, in addition, ratification itself is a costly exercise (Martin 2005). These *ex ante* costs mean that states are unlikely to sign onto nonproliferation agreements merely as a matter of course. Second, many of the treaties within the regime have relatively low stakes. The NPT, of course, is the focal point of international attention within the regime.[6] For every NPT, however, there are several Joint Conventions on the Safety of Spent Fuel Management and on the Safety of Radioactive Waste Management. Member states are unlikely to face strong outside pressure to adhere to this kind of nonproliferation agreement. For most elements of the regime, the best explanation for the decision of a state to join is that it prefers the policy outcome embodied in the institution.

[6] I investigate the NPT's importance in signaling state preferences in the empirical models to follow. Deriving the latent nonproliferation policy space using a measurement model that excludes the NPT does not affect the results of my analysis. Furthermore, the item-discrimination parameter (β) from the primary IRT model indicates that the NPT ranks in the middle of the pack in terms of the information it provides about underlying policy preferences.

The results of the IRT model are shown in Figure 4.[7] Each dot represents an estimate of the latent nuclear nonproliferation preferences—the nonproliferation score—in a given country-year.[8] Variation in nonproliferation scores increases over time, illustrating the additional information about policy preferences that may be revealed as the regime increases in complexity. The average nonproliferation score also increases with time, shifting from about 0.25 in 1960 to about 2.10 in 2010; this suggests a general increase in support for nonproliferation policy goals.

As Figure 4 shows, the United States has generally been a solid proponent of nonproliferation, but it is far from the strongest adherent to the regime. The US nonproliferation score—depicted in the figure by a triangle—has remained largely unchanged since the early 1980s, but other states have steadily grown more supportive of nonproliferation goals. As a result, US standing has fallen. Its nonproliferation score ranked 27th among all states in 1990, but by 2010 had fallen to 111th. By contrast, India, shown as a square in the figure, had long been on the periphery of the nonproliferation regime. In 1990, only four states had lower nonproliferation scores; in 2010, however, its nonproliferation score was greater than 36 other states.

Nonproliferation scores are strongly correlated (0.60) with the additive measure of regime embeddedness discussed above, the share of eligible treaties joined. Figure 5 plots individual country-years using these two measures. Triangles represent non-P-5 states with active nuclear weapons programs; these country-years score lower, on

[7] The IRT model is estimated using Markov chain Monte Carlo methods with the JAGS software (Plummer 2010). After running 30,000 iterations, the first 5,000 are discarded and the rest used for inference. Diagnostics indicate convergence (Gill 2007).

[8] Dots are plotted with vertical jitter for clarity.

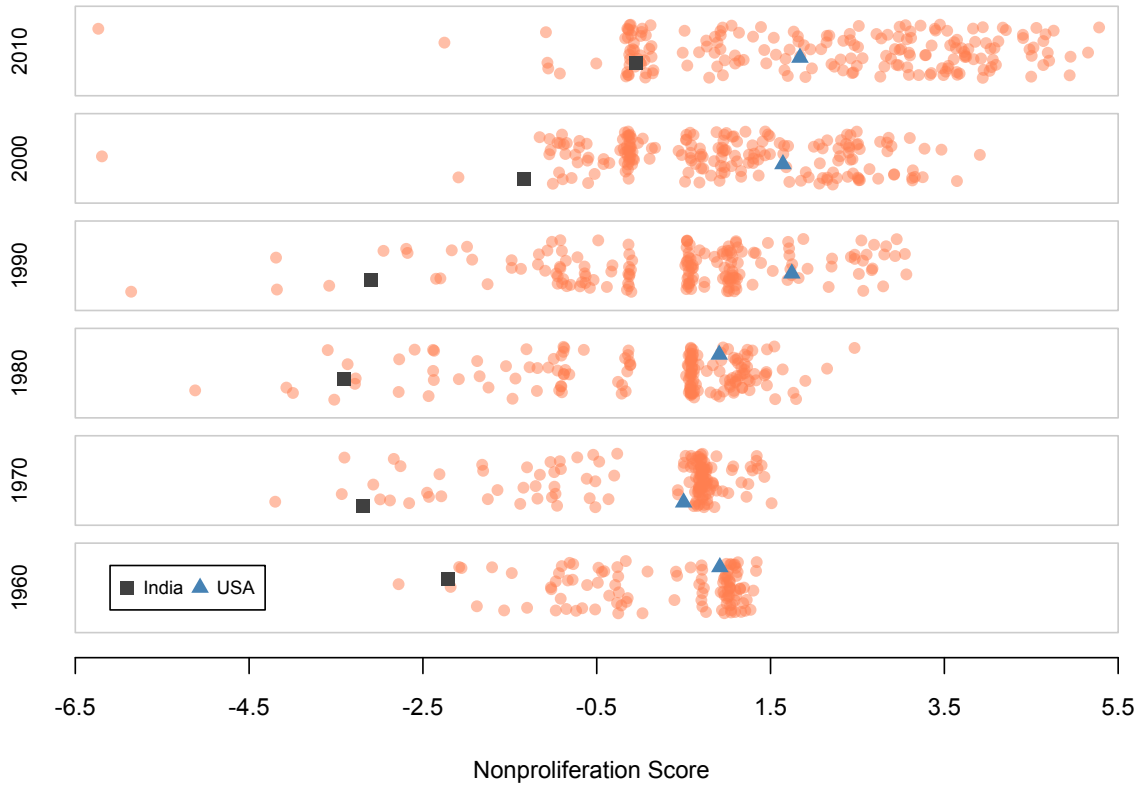

Figure 4: Distribution of nonproliferation scores over time

average, along both measures. Note that while the additive measure is left-censored at 0, the latent measure derives substantial variation in underlying preferences among even those states that had not yet signed onto a nonproliferation agreement.

Sanctions, preemptive attack, and nonproliferation policy preferences

Noncompliance within the nuclear nonproliferation regime can take many lesser forms, but the most clear-cut violation of the regime's tenets is a state's pursuit of nuclear weapons. Such violations are not infrequent. Eighteen states, not including the P-5 weapons states recognized by the NPT, have actively sought nuclear weapons since 1940,

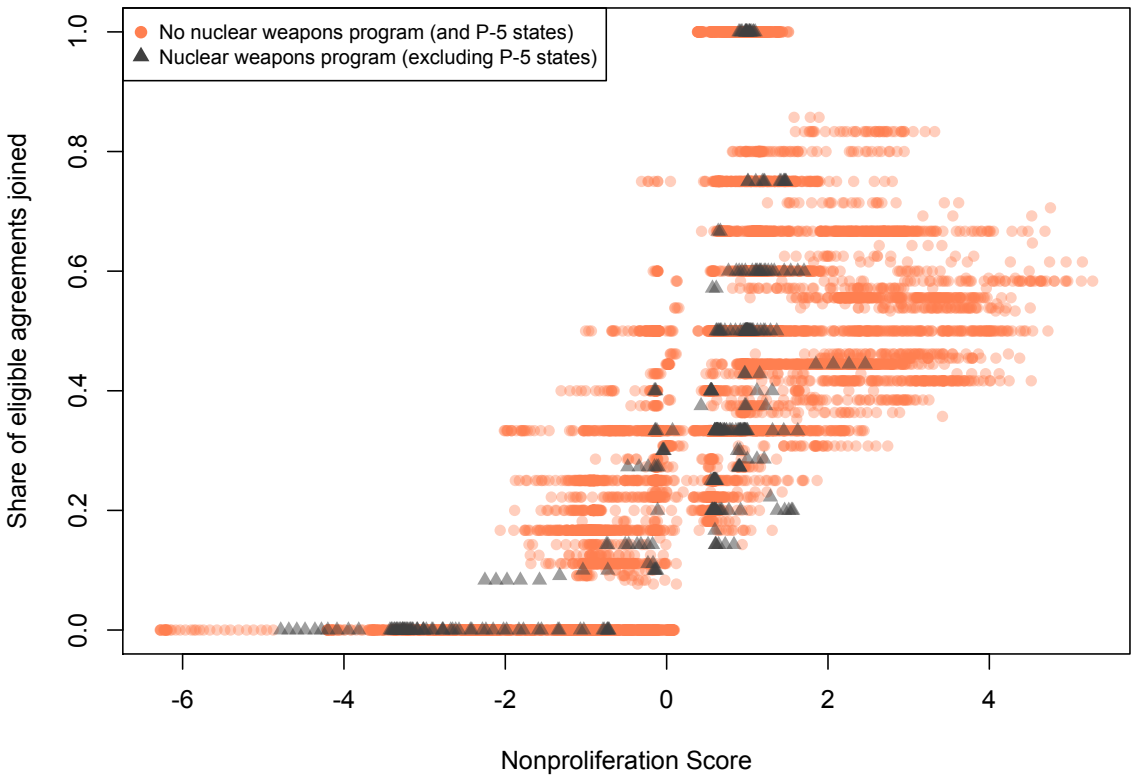

Figure 5: Additive and latent measures of nonproliferation policy preferences

with some engaging in multiple efforts.[9] Most of these states were members of at least some elements of the nonproliferation regime while they maintained a nuclear weapons program; nine states have sought weapons while members of the NPT.

Does a state's position within the nonproliferation regime complex affect how others respond if it chooses to pursue nuclear weapons? My theory suggests that states are less likely to be targeted with costly enforcement when their underlying policy preferences support nonproliferation goals or are similar to the preferences of the enforcing state. To examine this question empirically, I draw on a pooled time series

[9] Nuclear weapons program dates are updated from Jo and Gartzke (2007).

dataset of directed dyad-years from 1957—when the first institution within the regime, the IAEA, opened for membership—to 2001. Each observation consists of a potential enforcing state and a potential target state in a given year.[10] Because our interest is in how the international community responds to proliferating states, I limit the sample to those dyad-years in which the target state has a nuclear weapons program or an actual nuclear weapons capability.[11]

Restricting the data to cases of nuclear weapons pursuit raises questions about the possible confounding effects of nuclear secrecy. My coding of nuclear weapons programs has the benefit of hindsight, but states at the time may well have been unaware of some nuclear weapons efforts; the fact that a state is developing nuclear weapons is frequently one of its most closely held secrets.[12] Nuclear secrecy is unlikely to bias my results, however, unless a state's ability to disguise its nuclear efforts is systematically related to its embeddedness in the regime. Because several of the agreements within the regime are focused on the verification and monitoring of state nuclear efforts, this seems unlikely. If anything, states with stronger links to the regime are at greater risk of exposing secret

[10] While regime membership data from Carcelli et al. (2014), along with the corresponding nonproliferation scores, are available through 2014, I limit the scope of the analysis here to match the availability of covariates.

[11] I include in the main sample the P-5 states whose nuclear weapons programs are recognized under the NPT because these states too may be subject to costly enforcement measures; both the United States and Taiwan, for example, considered attacks against Chinese nuclear infrastructure before the NPT came into force (Fuhrmann and Kreps 2010). Removing the P-5 states from the sample after the advent of the NPT, or removing them from the sample completely, does not change my findings.

[12] This uncertainty is emphasized by the fact that there are several alternative codings of nuclear weapons pursuit currently in wide use in the international relations literature. For alternative nuclear weapons pursuit dates, see Bleek and Lorber (2014) and Singh and Way (2004). On the pitfalls of coding and analyzing nuclear weapons pursuit and acquisition, see Montgomery and Sagan (2009).

nuclear work. As an extra check, I conducted additional analyses in which I limited the nuclear weapons programs in my data to those that had been in progress for 1, 2, and 5 years, and so would have a better chance of being widely known to other member states. These robustness tests did not change my findings.

Dependent variables

The treaties and agreements within the nuclear nonproliferation regime lack enforcement measures of their own, but the international community has a number of tools with which to punish violations, including positive inducements such as foreign aid, military sales, security guarantees, and alliance ties.[13] Two policy tools are notable, however, for their potential cost to both the violator and the punisher: economic sanctions and military attack. The cost of these options is such that the enforcing state is likely to consider all available information that bears on the behavior of the violators—including state preferences revealed by patterns of membership within the regime complex—before launching an attack or imposing sanctions.

The dependent variables in my analysis are these two costly enforcement strategies. I first create a dichotomous variable that is set to 1 if a state threatened or imposed sanctions on the target in a given year, using data from the Threat and Imposition of Sanctions (TIES) dataset (Morgan, Bapat, and Kobayashi 2013). I include in the measure only those sanctions intended to deny strategic materials to the target state, which leaves sanctions threatened or imposed against 11 states with nuclear

[13] The IAEA's Board of Governors may refer cases of noncompliance with the NPT to the United Nations for action by the Security Council. For detailed discussion of the positive and negative inducements available to individual states in limiting proliferation, and their applications in specific cases, see the contributions to Solingen (2012).

weapons programs since 1957.[14] Including both threatened and imposed sanctions in the analysis helps to avoid the potential selection problem that results from excluding cases in which costly enforcement would have occurred if the target had maintained their bad behavior. Robustness checks using only imposed sanctions yield similar results for the explanatory variables of interest.

To test hypotheses about the military targeting of nuclear weapons programs, I use data from Fuhrmann and Kreps (2010), which identifies 15 dyads since 1957 in which attacks against nuclear facilities were considered or actually occurred. This dependent variable takes on the value of 1 if the would-be attacker considered or actually initiated an attack against the potential target state in a given year.[15] There is very little overlap between those dyads in which sanctions are threatened and those in which military action against a nuclear program is considered. Employing both measures thus allows for a broader test of my hypotheses.

Regime membership and nonproliferation preferences

I use several variables to represent states' nonproliferation policy interests. To test the preference hypothesis—that underlying support for nonproliferation policy goals reduces the risk of costly enforcement—I use both the share of eligible treaties the target state chooses to join and the target state's nonproliferation score. The similarity hypothesis posits that states will be less likely to target for costly enforcement those states

[14] This category does not include sanctions that were threatened or imposed with the goal of preventing states from sharing sensitive technology with third-party countries.

[15] An alternative variable that counts only attacks that were actually initiated yields no significant results for the explanatory variables of interest. This is not surprising—the Fuhrmann and Kreps (2010) data includes only 12 dyad years in which attacks against nuclear weapons facilities have occurred since 1957, probably too few cases to identify any effect.

that have similar underlying nonproliferation preferences. I test this hypothesis using two measures of preference similarity: the absolute difference between the share of eligible treaties joined by the states in the dyad, and the absolute difference between the two states' nonproliferation scores. Each of these variables is derived from the nonproliferation and nuclear security agreements in the Carcelli et al. (2014) nuclear regimes dataset.

I attempt to distinguish in my empirical tests between the information provided through regime complexity and the signal sent by membership in a single treaty or institution. In the nuclear nonproliferation regime, the NPT is by far the most prominent international agreement. It may be that NPT membership is all that is necessary to send a signal to the international community about a state's type, and that the regime complex provides little or no additional information. To test this possibility, I use a dichotomous control variable that takes on the value of 1 if a state is an NPT member in a given year.

Control variables

I control for a variety of other factors that may affect both the likelihood of costly enforcement and a state's nonproliferation policy preferences or pattern of regime membership. Sanctions or military action may be more likely in conflict dyads, so I use a dichotomous variable to control for ongoing conflict between the two states (N. P. Gleditsch et al. 2002). I address power dynamics between the two states using the ratio of the CINC scores in the dyad (Singer 1988). We would expect that greater power disparities in a dyad would make sanctions threats or consideration of an attack more likely. International alliances may also play a role in determining both the use of costly enforcement mechanisms and a state's regime embeddedness. I include a dichotomous

measure that takes on the value of 1 if a target state has an alliance with a nuclear weapons state (Gibler and Sarkees 2004). Such states may face more pressure to join the nonproliferation regime, and may also be less subject to sanction because of their powerful protectors.

More advanced nuclear weapons programs are likely to draw stronger enforcement action. States with fledgling efforts might respond to lower cost diplomatic pressure or positive inducements. Programs that are further along in nuclear development also are perceived as more threatening to other states, for the simple reason that they are more likely to result in the successful acquisition of weapons. And of course it may take the international community some time to reach a consensus on whether costly enforcement is necessary. I control for both the target state's latent nuclear capacity—using a 7-point index of nuclear capability from Jo and Gartzke (2007)—and for the number of years that a nuclear program has been in progress.

I also control for whether or not the target state has actually acquired nuclear weapons, using the consensus nuclear acquisition dates from Gartzke and Kroenig (2009). This variable may cut both ways. The successful development of weapons may galvanize an enforcement response; the international community may consider it necessary to show that violators will face some punishment, even if the targeted state seems unlikely to relinquish its weapons. At the same time, enforcement efforts against nuclear states are costly and have little chance of success. Potential enforcers may be reluctant to antagonize the most recent member of the nuclear club.

Joint democracy, too, may influence the chance of sanctions or military action. We might expect that democracies are less likely at least to consider attacking other

democracies. The relationship between joint democracy and sanctions is less straightforward, because democracies may be both better positioned to punish states with trade sanctions and more vulnerable to the having sanctions imposed upon them. I control for joint democracy by creating a dichotomous variable that takes on the value of 1 if both states in the dyad have Polity scores greater than 6 (Marshall, Jaggers, and Gurr 2010).

Proliferation by neighbors may be more salient than proliferation taking place continents away; I control for geographic distance with a dummy variable representing contiguity between the states in the dyad (Weidmann, Kuse, and Gleditsch 2010).[16] Finally, to address the possibility that it is general foreign policy similarity within the dyad, and not the information provided by the regime complex, that affects the likelihood of sanctions, I control for the affinity of the two states using *S* scores derived from UN general assembly voting data (Strezhnev and Voeten 2013; Gartzke 2006).

Modeling approach

Costly enforcement action is very rare. Sanctions are threatened or imposed in about 0.6 percent of dyad-years in my dataset, and military attack is considered in only 0.1 percent of observations. To correct for rare event and finite sample biases, I employ penalized likelihood logistic regression; this approach has the added benefit of addressing problems of quasi-separation in the data (Firth 1993; Heinze and Schemper 2002; King

[16] Robustness checks using a measure of the minimum distance between states in the dyad yield similar results.

and Zeng 2001; Zorn 2005).[17] I address temporal dependence with a cubic polynomial of the count of the number of years since the previous costly enforcement in the dyad—threatened or imposed sanctions or the consideration of a military attack, depending on the model (Carter and Signorino 2010). I report robust standard errors clustered on the dyad; using two-way clustering—on each of the countries in the dyad—yields the same results for the independent variables of interest (Cameron, Gelbach, and Miller 2011).

The nonproliferation scores used in some models are themselves the output of a statistical model that has been estimated with uncertainty. To take into account the additional error that attends to this variable, I follow the recommendation of Schnakenberg and Fariss (2014) and combine multiple nonproliferation scores in the same way one would treat multiply imputed data. For models that make use of nonproliferation scores, I model costly enforcement using 10 separate datasets, each with a separate draw from the posterior distribution of the IRT model, and combine the estimates from each model using Rubin's calculation for combined standard errors (King et al. 2001; Rubin 1987).

Findings

My findings, overall, support the idea that the information gleaned from regime complexity can affect important outcomes in international relations, such as the use of costly enforcement in response to treaty violations. Table 1 shows the results of four models that examine the relationship between regime membership and sanctions

[17] Robustness checks using rare-events logit yield the same results for the variables of interest (King and Zeng 2001). Some of these models suffer from quasi-separation in the data, however, so the results presented below are estimated using penalized likelihood logistic regression.

Table 1: Sanctions threatened or imposed against regime violators

		Model 1	Model 2	Model 3	Model 4
		Sanctions	Sanctions	Sanctions	Sanctions
Regime complex	Share of eligible treaties	**-1.334** **(0.272)**			
	Difference in share of eligible treaties		**1.359** **(0.242)**		
	Nonproliferation score			**-0.518** **(0.058)**	
	Difference in nonproliferation scores				**0.858** **(0.066)**
	NPT membership	**1.427** **(0.216)**	**1.537** **(0.226)**	**2.068** **(0.242)**	**1.606** **(0.242)**
Conflict	Conflict dyad	**1.021** **(0.442)**	**1.130** **(0.413)**	**1.001** **(0.478)**	**1.141** **(0.458)**
Power	CINC ratio	**0.029** **(0.006)**	**0.028** **(0.006)**	**0.030** **(0.006)**	**0.027** **(0.006)**
Alliances	Nuclear ally	-0.259 (0.238)	0.170 (0.234)	**-0.525** **(0.242)**	-0.215 (0.245)
Nuclear capability	Length of nuclear program	**0.150** **(0.014)**	**0.141** **(0.014)**	**0.150** **(0.014)**	**0.171** **(0.016)**
	Latent nuclear capacity	**0.200** **(0.070)**	0.134 (0.069)	0.106 (0.073)	0.057 (0.077)
	Nuclear weapons state	**1.554** **(0.417)**	**1.356** **(0.374)**	**2.046** **(0.413)**	**1.588** **(0.442)**
Democracy	Both democracies	**1.855** **(0.271)**	**1.915** **(0.282)**	**1.835** **(0.276)**	**2.486** **(0.314)**
Geography	Contiguity	**1.471** **(0.277)**	**1.906** **(0.281)**	**1.268** **(0.293)**	**2.670** **(0.310)**
Affinity	UN affinity	**-0.505** **(0.147)**	**-0.514** **(0.146)**	**-0.511** **(0.150)**	**-0.922** **(0.166)**
	Constant	**-5.832** **(0.530)**	**-6.847** **(0.580)**	**-6.108** **(0.564)**	**-7.987** **(0.660)**
	N	30,652	30,652	30,652	30,652

Penalized likelihood logistic regression coefficients with robust standard errors, clustered on dyad, in parentheses values are statistically significant ($p<0.05$). Cubic polynomials of time since the previous consideration of attack included but not shown.

threatened or imposed against states with nuclear weapons programs. In Model 1, the coefficient on the independent variable of interest, the share of eligible treaties joined by the target state, is negative and significant. Deeper regime membership is associated with fewer sanctions on states seeking nuclear weapons. Model 2 investigates how the difference between two states' regime embeddedness affects the likelihood that one will threaten or impose sanctions on the other. A larger difference between the levels of regime membership within the dyad significantly increases the likelihood of sanctions. Models 3 and 4 come to the same result using nonproliferation scores—stronger preferences for nonproliferation and more similar preferences within the dyad are strongly associated with a reduced likelihood of threatened or imposed sanctions. These findings are consistent with my theory, and support both the preference and similarity hypotheses. A state's pattern of treaty membership seems to affect the way it is treated by other states when it chooses to violate a central tenet of the regime.

Table 2 gives the results for models of military attack against nuclear weapons programs. Again, the findings are largely consistent with my theory. The target state's level of regime membership and its nonproliferation score (Models 5 and 7) are strongly associated with a reduced likelihood of a nuclear weapons program being targeted for attack. The difference in nonproliferation scores within the dyad, in Model 8, is associated with an increased likelihood of an attack being considered. The results of Model 6, however, show that the difference in the share of eligible treaties joined by the two states in the dyad is not a statistically significant driver of military attack.

Table 2: Military attack considered or implemented against regime violators

		Model 5 Attack considered	Model 6 Attack considered	Model 7 Attack considered	Model 8 Attack considered
Regime complex	Share of eligible treaties	**-3.1991** **(1.255)**			
	Difference in share of eligible treaties		-0.768 (0.768)		
	Nonproliferation score			**-0.596** **(0.167)**	
	Difference in nonproliferation scores				**0.472** **(0.198)**
	NPT membership	1.030 (0.714)	0.898 (0.547)	**1.222** **(0.616)**	0.731 (0.558)
Conflict	Conflict dyad	**4.226** **(0.983)**	**4.399** **(0.935)**	**4.646** **(1.016)**	**4.665** **(0.963)**
Power	CINC ratio	0.020 (0.011)	0.019 (0.011)	0.021 (0.011)	**0.025** **(0.011)**
Alliances	Nuclear ally	-1.240 (0.639)	-0.564 (0.588)	**-1.379** **(0.665)**	-0.779 (0.599)
Nuclear capability	Length of nuclear program	**0.115** **(0.041)**	**0.112** **(0.039)**	**0.142** **(0.044)**	**0.121** **(0.040)**
	Latent nuclear capacity	0.283 (0.157)	0.239 (0.152)	0.213 (0.165)	0.212 (0.161)
	Nuclear weapons state	1.267 (1.219)	0.875 (1.223)	1.928 (1.266)	1.253 (1.170)
Democracy	Both democracies	-0.260 (1.456)	-0.800 (1.455)	-0.282 (1.426)	-0.647 (1.434)
Geography	Contiguity	**2.730** **(0.666)**	**2.728** **(0.645)**	**2.777** **(0.698)**	**2.996** **(0.649)**
Affinity	UN affinity	**-1.283** **(0.451)**	**-1.006** **(0.429)**	**-1.501** **(0.479)**	**-1.243** **(0.443)**
	Constant	**-6.975** **(1.440)**	**-7.589** **(1.393)**	**-8.302** **(1.560)**	**-8.764** **(1.571)**
	N	29,493	29,493	29,493	29,493

Penalized likelihood logistic regression coefficients with robust standard errors, clustered on dyad, in parentheses. Bold values are statistically significant ($p < 0.05$). Cubic polynomials of time since the previous consideration of attack are included but not shown.

These results persist even when controlling for NPT membership, suggesting that there is more to the nonproliferation regime complex than a single high-profile treaty. The coefficient on NPT membership is positive and significant in Models 1–4 and 7, and fails to reach significance in the remaining models. Sanctions, then, are more likely when the target state is an NPT member.[18] This finding seems to lend some support to a more conventional understanding of enforcement behavior in international institutions based on, for example, ideas about reciprocal enforcement or the primacy of international law. While overall regime embeddedness seems to militate against costly enforcement, NPT membership points in the other direction. This may not be surprising, given the intense international focus on the NPT. One way in which the NPT may work to constrain state behavior is by mobilizing the international community against violators (Kaplow 2012).

The control variables in these models largely behave as expected. Conflict dyads are more likely to see both sanctions and the consideration of military action against nuclear weapons programs. Sanctions are significantly associated with power disparities within the dyad; the coefficient on the ratio of CINC scores also reached significance in Model 8, where military action is more likely to be considered when the enforcing state is much more powerful than the target state. The presence of a nuclear ally makes sanctions or military action less likely in Models 3 and 7. The various measures of nuclear capability are generally significant drivers of threatened or imposed sanctions. Only the length of a nuclear program, however, seems to affect the likelihood of being targeted for attack. Contiguity and affinity are both important determinants of costly enforcement.

[18] Results for the independent variables of interest do not change when NPT membership is omitted from the model.

States are more likely to punish their neighbors for violations of the regime, and they are less likely to sanction or attack states with which they share common foreign policy interests.

One surprise is the joint democracy variable, which is associated with threatened or imposed sanctions in Models 1–4. This result is driven in part by the strong international response to the Indian and Pakistani nuclear tests in 1998, in which these two democratic states were the target of sanctions threats from a number of other democratic states. If dyads involving India and Pakistan are excluded from the dataset, the joint democracy variable loses its significant association with threatened or imposed sanctions.

Substantive effects and predictive validity

The role of nonproliferation policy preferences in moderating costly enforcement is both statistically and substantively significant. Both sanctions and military action are rare events, so the raw predicted probabilities for these outcomes are quite low across the range of nonproliferation scores or levels of regime membership, when other variables are held at their global mean. These predicted probabilities, however, are skewed by the presence of dyads in which the chances of sanctions or military action are exceedingly unlikely; as a result, they probably understate substantive effects for the cases about which we are most interested. If our interest is primarily in the dyads that have at least some reasonable risk of costly enforcement, we may more usefully estimate the substantive effects of nonproliferation policy preferences by holding all other variables at

the mean among those observations predicted by the model to have at least a small chance of costly enforcement.[19]

Figure 6 illustrates these substantive effects. The range of nonproliferation scores is plotted on the horizontal axis, with the likelihood of costly enforcement on the vertical axis. The solid line shows the predicted probability of a threatened or imposed sanction for each level of nonproliferation score, with all other variables held at their mean among those dyads predicted to have a greater than 1 percent chance of threatened sanctions.[20] Among high-risk dyads, then, a shift in nonproliferation score from one standard deviation below its mean (about -1.7) to one standard deviation above its mean (about 1.7) is associated with a roughly 9-percentage point drop in the chance of threatened or imposed sanctions.[21] The dashed line shows predicted probabilities of considered military action. Here, the same increase in nonproliferation score is associated with about a 10-percentage point reduction in the risk that military action will be considered.[22]

Another approach to examining the substantive importance of these findings is through prediction. Does incorporating the underlying nonproliferation preferences of states into our analysis contribute to our ability to predict costly enforcement? Out-of-sample prediction, in particular, helps us look beyond statistical significance to identify variables that have real predictive power; strong out-of-sample performance should give

[19] A similar approach was used by Beck, King, and Zeng (2000) to examine those dyads at higher risk of international conflict.

[20] The predicted probabilities in Figure 6 are calculated using Models 3 and 7.

[21] The standard deviation (1.7) and mean (0) computed here are for the population of proliferating states, not the high-risk subsample.

[22] The preference similarity variables tested in other models have substantive effects of similar magnitude.

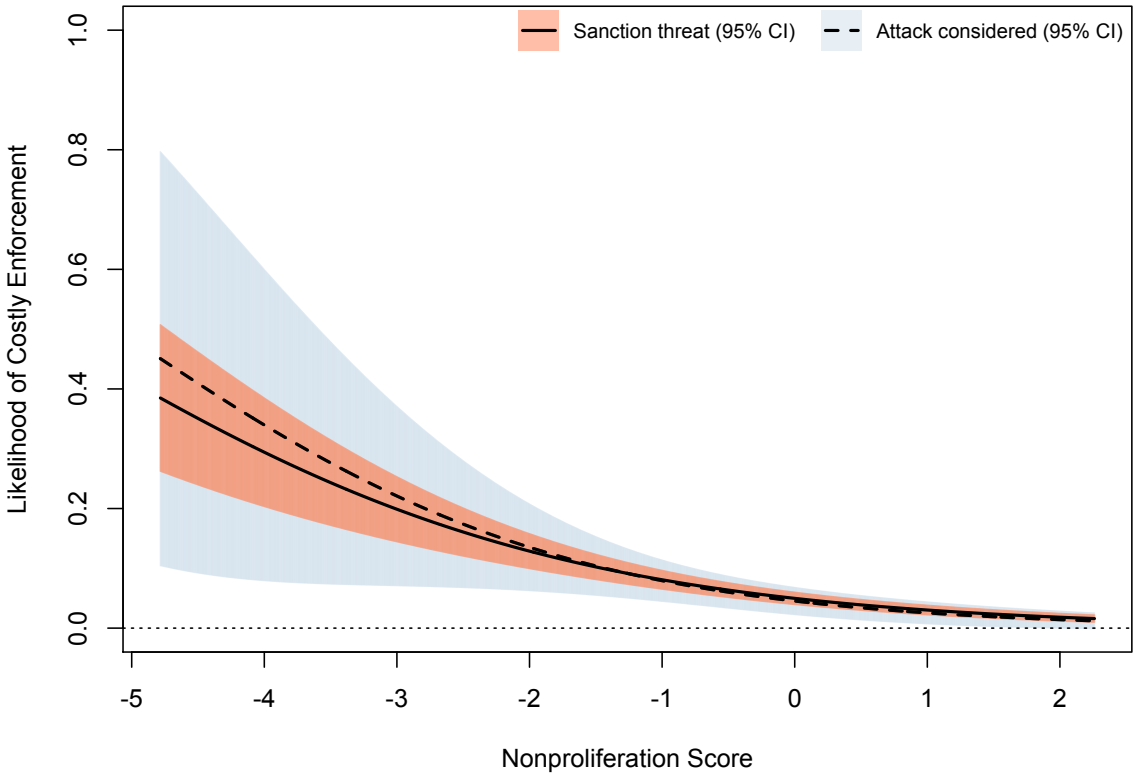

Figure 6: Substantive effect of nonproliferation preferences in high-risk dyads

us more confidence that these statistical models are successfully capturing real relationships between nonproliferation preferences and costly enforcement (Beck, King, and Zeng 2000; Ward, Greenhill, and Bakke 2010).

To test the predictive power of regime complexity, I employ a 3-fold cross-validation procedure. The data are randomly divided into three equal parts, two of which are used to train the model and one of which is reserved for out-of-sample testing.[23] This process is repeated three times so that each third of the dataset serves once as the test

[23] Splitting the data into two equal parts rather than three gives similar results. Because costly enforcement is quite rare, creating additional subsamples frequently leads to validation datasets with no incidence of sanctions or considered attack.

data. To be sure that my findings do not depend on the initial random division of the data, I execute the entire 3-fold validation procedure 10 times, using different random subsamples in each attempt, and average the results. The modeling approach used here is identical to that presented above; I employ penalized likelihood logistic regression, with 10 separate draws from the posterior distribution of the IRT model which are combined after estimation.

There are many available measures for evaluating the predictive performance of a statistical model. One obvious option is the overall accuracy of the prediction, but this measure is uninformative when predicting rare events; a model that guesses "no" for every dyad-year would correctly predict threatened or imposed sanctions 99.4 percent of the time. While there is no consensus in the literature on the preferred evaluation metric with data of this type, a common strategy is to focus on the model's predictions in the category of rare events, rather than the model's predictions overall.[24] I evaluate the predictive performance of my models using the precision metric, also known as positive predictive value. Precision is the number of true positives divided by the number of positive predictions; that is, the percent of "yes" guesses that the model got right.[25]

The predictive testing procedure described above finds that a model that uses the nonproliferation score of the target state to predict threatened or imposed sanctions (Model 3) has an out-of-sample precision of 0.63. Put another way, when this model

[24] Prediction in the presence of skewed data or class imbalance is a familiar problem in the machine learning literature. For a discussion of evaluation metrics in this context, see He and Garcia (2009); Joshi (2002); and Kotsiantis, Kanellopoulos, and Pintelas (2006).

[25] This measure is particularly appropriate when false positives are costly. This might be the case, for example, if the predictive model was being used to screen data for additional qualitative analysis.

predicts that sanctions will be threatened or imposed in the validation data, it is correct 63 percent of the time on average. This performance compares favorably to other models of costly enforcement. Figure 7 illustrates the precision metric for several alternatives ways of understanding threatened or imposed sanctions, compared to a baseline model that omits measures of regime or NPT membership and nonproliferation preferences. The vertical solid gray lines in the figure give a point estimate of the change in precision when adding the variable in question to the baseline model; the length of the horizontal gray bars represents the 95 percent bootstrapped confidence interval.

When nonproliferation scores are added to the baseline model, precision increases by about 13 percentage points. The share of eligible treaties joined is also an improvement over the baseline, increasing precision by about 10 percentage points. NPT status, however, actually hurts prediction by this evaluative measure. A statistical model with nonproliferation scores improves precision by about 19 percentage points over a model with NPT membership. Incorporating nonproliferation policy preferences into our analysis significantly enhances our ability to make out-of-sample predictions about the threatened or imposed sanctions, compared to analysis that looks just at NPT membership or ignores regime membership altogether.

Conclusion

In this paper, I present a novel explanation for selective enforcement in international security institutions. Would-be enforcing states look to the underlying policy preferences of violators, as revealed by their patterns of adherence within a larger security regime. These preferences provide useful signals both about the likelihood of

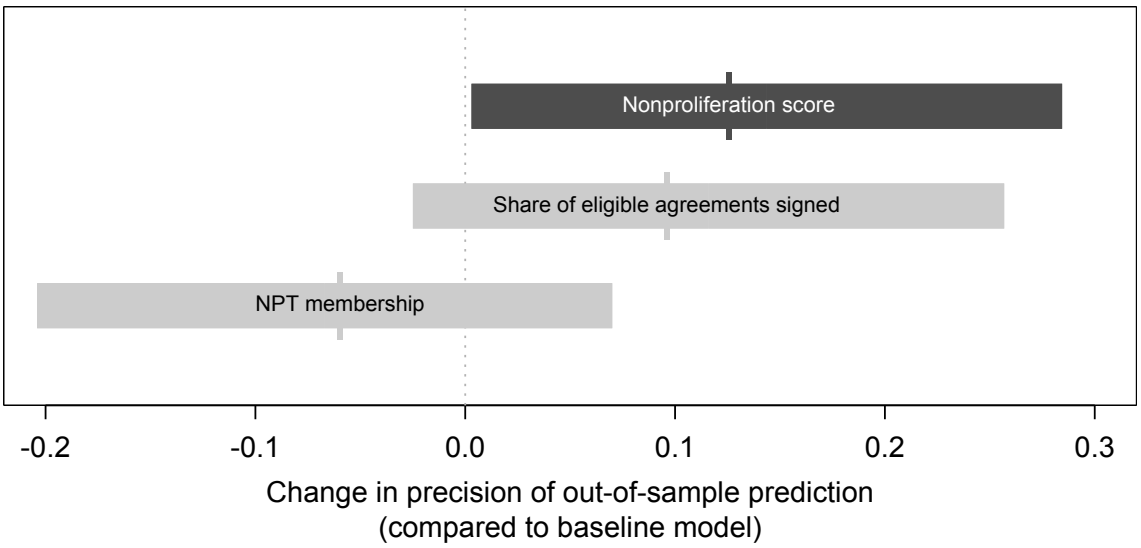

Figure 7: Out-of-sample prediction of threatened or imposed sanctions

enforcement action succeeding and about the potential cost to the international community if the violation continues.

This argument highlights a positive side effect of regime complexity: by more effectively screening states, the regime complex is able to generate valuable information about the underlying policy preferences of its members. This information may affect international outcomes generally, and in particular is likely to influence the way the international community responds to member states that have violated their treaty commitments. In the case of the nuclear nonproliferation regime, states may perceive underlying preferences over nonproliferation policy—information generated by the regime complex—as a mitigating factor when considering costly enforcement measures against proliferant states. States may also discount threats from others based on the similarities between their underlying policy preferences or regime portfolios—again, information that would not be available absent the regime complex.

I test this theory with new data on state membership in 22 constituent agreements of the nuclear nonproliferation regime. Using an IRT model, I derive state positions in the latent nuclear nonproliferation policy space. I then examine whether these nonproliferation scores affect how states are treated by the international community if they choose to pursue a nuclear weapons program in violation of a central obligation of the regime. States with stronger support for nonproliferation policy goals are significantly less likely to be targeted for international sanctions or considered for preemptive attack. States pursuing nuclear weapons also are less likely to be sanctioned or considered for attack by states that have similar underlying policy preferences.

These findings suggest that scholars of international organizations could benefit from looking beyond a single treaty or set of treaties when studying the effect of institutions on state behavior. By examining the role played by the regime complex itself, the literature may capture additional constraining mechanisms or pathologies not visible at the treaty level.

These findings also have important implications for the large body of work on nuclear proliferation, which avoids, for the most part, serious consideration of the ability of international institutions to limit the spread of nuclear weapons. What work has been done to understand the nuclear nonproliferation regime has focused almost exclusively on the NPT. As my findings show, examining the NPT alone risks an incomplete analysis at best, and an incorrect inference at worst. In one quantitative model described above, a proliferant state's membership in the NPT was associated with an increased likelihood of nuclear sanctions, while deeper membership in the wider regime made sanctions less likely. The NPT is clearly important, but in some sense it is a victim of its own success.

Because nearly every state is an NPT member, the treaty provides little information for those seeking to understand the intentions and interests of member states. Policymakers already know to look to the broader regime for answers; scholars should as well.

References

Abbott, Kenneth W., and Duncan Snidal. 2000. "Hard and Soft Law in International Governance." *International Organization* 54(3): 421–56.

Alter, Karen J., and Sophie Meunier. 2009. "The Politics of International Regime Complexity." *Perspectives on Politics* 7(1): 13–24.

Barnett, Michael N., and Martha Finnemore. 1999. "The Politics, Power, and Pathologies of International Organizations." *International Organization* 53(4): 699–732.

Beck, Nathaniel, Gary King, and Langche Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review* 94(1): 21–35.

Bleek, Philipp C., and Eric B. Lorber. 2014. "Security Guarantees and Allied Nuclear Proliferation." *Journal of Conflict Resolution* 58(3): 429–54.

Busch, Marc L. 2007. "Overlapping Institutions, Forum Shopping, and Dispute Settlement in International Trade." *International Organization* 61(4): 735–61.

Carcelli, Shannon, Erik Gartzke, Rebecca Gibbons, and Jeffrey M. Kaplow. 2014. "The Nuclear Regime Complex: A New Dataset." Paper prepared for the 2014 International Studies Associate Annual Convention, Toronto, Canada.

Carter, David B., and Curtis S. Signorino. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18(3): 271–292.

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98(2): 355–70.

Colgan, Jeff D., Robert O. Keohane, and Thijs Van de Graaf. 2012. "Punctuated Equilibrium in the Energy Regime Complex." *The Review of International Organizations* 7(2): 117–43.

Davis, Christina L. 2009. "Overlapping Institutions in Trade Policy." *Perspectives on Politics* 7(1): 25–31.

Donno, Daniela. 2010. “Who Is Punished? Regional Intergovernmental Organizations and the Enforcement of Democratic Norms.” *International Organization* 64(4): 593–625.

———. 2013. *Defending Democratic Norms: International Actors and the Politics of Electoral Misconduct*. Oxford, UK: Oxford 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(3): 379–406.

Drezner, Daniel W. 2009. “The Power and Peril of International Regime Complexity.” *Perspectives on Politics* 7(1): 65–70.

Fariss, Christopher J. 2014. “Respect for Human Rights Has Improved Over Time: Modeling the Changing Standard of Accountability.” *American Political Science Review* 108(2): 297–318.

Firth, David. 1993. “Bias Reduction of Maximum Likelihood Estimates.” *Biometrika* 80(1): 27–38.

Fuhrmann, Matthew, and Sarah E. Kreps. 2010. “Targeting Nuclear Programs in War and Peace: A Quantitative Empirical Analysis, 1941-2000.” *Journal of Conflict Resolution* 54(6): 831 –859.

Gartzke, Erik. 2006. “The Affinity of Nations Index, 1946-2002.”

Gartzke, Erik, and Matthew Kroenig. 2009. “A Strategic Approach to Nuclear Proliferation.” *Journal of Conflict Resolution* 53(2): 151–60.

Gibler, Douglas M, and Meredith Reid Sarkees. 2004. “Measuring Alliances: The Correlates of War Formal Interstate Alliance Dataset, 1816–2000.” *Journal of Peace Research* 41(2): 211–22.

Gill, Jeff. 2007. *Bayesian Methods: A Social and Behavioral Sciences Approach, Second Edition*. Boca Raton, FL: CRC Press.

Gleditsch, Nils Petter et al. 2002. “Armed Conflict 1946-2001: A New Dataset.” *Journal of Peace Research* 39(5): 615–37.

Guzman, Andrew T. 2002. “A Compliance-Based Theory of International Law.” *California Law Review* 90(6): 1823–87.

Hafner-Burton, Emilie M. 2009. “The Power Politics of Regime Complexity: Human Rights Trade Conditionality in Europe.” *Perspectives on Politics* 7(1): 33–37.

He, Haibo, and E.A Garcia. 2009. “Learning from Imbalanced Data.” *IEEE Transactions on Knowledge and Data Engineering* 21(9): 1263–84.

Heinze, Georg, and Michael Schemper. 2002. “A Solution to the Problem of Separation in Logistic Regression.” *Statistics in Medicine* 21(16): 2409–19.

Helfer, Laurence R. 1999. “Forum Shopping for Human Rights.” *University of Pennsylvania Law Review* 148(2): 285–400.

Hirsch, Theodore. 2004. “The IAEA Additional Protocol: What It Is and Why It Matters.” *The Nonproliferation Review* 11(3): 140–66.

Hofmann, Stephanie C. 2009. “Overlapping Institutions in the Realm of International Security: The Case of NATO and ESDP.” *Perspectives on Politics* 7(1): 45–52.

Hug, Simon, and Richard Lukács. 2014. “Preferences or Blocs? Voting in the United Nations Human Rights Council.” *The Review of International Organizations* 9(1): 83–106.

International Atomic Energy Agency. 2005. *Strengthening the Effectiveness and Improving the Efficiency of the Safeguards System Including Implementation of Additional Protocols (GC(49)/1)*. Report to the IAEA General Conference.

Jo, Dong-Joon, and Erik Gartzke. 2007. “Determinants of Nuclear Weapons Proliferation.” *Journal of Conflict Resolution* 51(1): 167–94.

Joshi, Mahesh V. 2002. “On Evaluating Performance of Classifiers for Rare Classes.” *IEEE International Conference on Data Mining, 2002,* 641–44.

Kaplow, Jeffrey M. 2012. “Does the Nuclear Non-Proliferation Treaty Matter?” Paper prepared for the 2012 Annual Conference of the International Studies Association, San Diego, CA.

Kelley, Judith. 2009. “The More the Merrier? The Effects of Having Multiple International Election Monitoring Organizations.” *Perspectives on Politics* 7(1): 59–64.

Keohane, Robert O., and David G. Victor. 2011. “The Regime Complex for Climate Change.” *Perspectives on Politics* 9(1): 7–23.

Kerr, Paul. 2005. “IAEA Board Closes Safeguards Loophole.” *Arms Control Today*.

King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. 2001. “Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation.” *American Political Science Review* 95(1): 49–69.

King, Gary, and Langche Zeng. 2001. “Logistic Regression in Rare Events Data.” *Political Analysis* 9(2): 137–163.

Kotsiantis, Sotiris, Dimitris Kanellopoulos, and Panayiotis Pintelas. “Handling Imbalanced Datasets: A Review.” *GESTS International Transactions on Computer Science and Engineering* 30(1): 25–36.

Marshall, Monty G., Keith Jaggers, and Ted Robert Gurr. 2010. *Polity IV Project: Political Regime Characteristics and Transitions, 1800-2010*. Vienna, VA: Center for Systemic Peace.

Martin, Andrew D., and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999." *Political Analysis* 10(2): 134–53.

Martin, Lisa L. 2005. "The President and International Commitments: Treaties as Signaling Devices." *Presidential Studies Quarterly* 35(3): 440–65.

Montgomery, Alexander H., and Scott D. Sagan. 2009. "The Perils of Predicting Proliferation." *Journal of Conflict Resolution* 53(2): 302–28.

Morrow, James D. 2007. "When Do States Follow the Laws of War?" *American Political Science Review* 101(3): 559–72.

Plummer, Martyn. 2010. "JAGS (Just Another Gibbs Sampler) 1.0.3 Universal." http://www-fis.iarc.fr/martyn/software/jags/

Poole, Keith T., and Howard Rosenthal. 1991. "Patterns of Congressional Voting." *American Journal of Political Science* 35(1): 228.

Raustiala, Kal, and David G. Victor. 2004. "The Regime Complex for Plant Genetic Resources." *International Organization* 58(2): 277–309.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York, NY: J. Wiley & Sons.

Schnakenberg, Keith E., and Christopher J. Fariss. 2014. "Dynamic Patterns of Human Rights Practices." *Political Science Research and Methods* 2(1): 1–31.

Simmons, Beth A. 2000. "International Law and State Behavior: Commitment and Compliance in International Monetary Affairs." *American Political Science Review* 94(4): 819–35.

———. 2010. "Treaty Compliance and Violation." *Annual Review of Political Science* 13(1): 273–96.

Simmons, Beth A., and Daniel J. Hopkins. 2005. "The Constraining Power of International Treaties: Theory and Methods." *American Political Science Review* 99(4): 623–31.

Singer, J. David. 1988. "Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816–1985." *International Interactions* 14(2): 115–32.

Singh, Sonali, and Christopher R. Way. 2004. "The Correlates of Nuclear Proliferation: A Quantitative Test." *Journal of Conflict Resolution* 48(6): 859–85.

Solingen, Etel, ed. 2012. *Sanctions, Statecraft, and Nuclear Proliferation*. Cambridge, UK: Cambridge University Press.

Stone, Randall W. 2004. "The Political Economy of IMF Lending in Africa." *American Political Science Review* 98(4): 577–91.

———. 2008. "The Scope of IMF Conditionality." *International Organization* 62(04): 589–620.

Von Stein, Jana. 2005. "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance." *American Political Science Review* 99(4): 611–22.

Treier, Shawn, and Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52(1): 201–17.

Ward, Michael D., Brian D. Greenhill, and Kristin M. Bakke. 2010. "The Perils of Policy by P-Value: Predicting Civil Conflicts." *Journal of Peace Research* 47(4): 363–75.

Weidmann, Nils B., Doreen Kuse, and Kristian Skrede Gleditsch. 2010. "The Geography of the International System: The CShapes Dataset." *International Interactions* 36(1): 86–106.

Zorn, Christopher. 2005. "A Solution to Separation in Binary Response Models." *Political Analysis* 13(2): 157–170.