2. Breadth Versus Depth in International Security Institutions

Jeffrey M. Kaplow†

Paper prepared for the International Studies Association Annual Convention, Atlanta, March 2016

Abstract

Some scholarship in the international organizations literature has argued that there is an inherent tradeoff between the depth and breadth of international agreements: to achieve deep institutions with high levels of obligation, the pool of member states must be limited. International security institutions, however, often appear both broad and deep—they impose significant obligations on state parties but still enjoy a large membership. This paper examines this empirical puzzle, theorizing that in some cases, when it is costly both to abstain from and to be found in violation of an agreement, no broader versus deeper tradeoff exists. High obligation may actually be necessary for large membership, because states will only join institutions when they expect other member states to comply. Using a dataset of membership in the Treaty on the Non-Proliferation of Nuclear Weapons since 1968, I find that states are more likely to join when the level of obligation within the regime is seen to be high, but that this relationship is reversed among those states most likely to violate the treaty. The results in this paper point to a more dynamic view of state adherence, in which prospective members frequently reevaluate the costs of benefits of treaty membership in light of the expected behavior of other states.

† College of William and Mary; Department of Government; jkaplow@wm.edu
The nuclear nonproliferation regime, and international security institutions more generally, enjoy the surprising combination of a large number of member states and substantial legal obligation. Nearly all states have now joined the cornerstone of the regime, the Treaty on the Non-Proliferation of Nuclear Weapons (NPT)—only India, Israel, Pakistan and newly independent South Sudan have never joined, while North Korea withdrew from the treaty in 2003. But membership is extremely costly. The NPT requires non-nuclear weapons states to forgo a potentially valuable military capability and, in many cases, to submit to burdensome international inspections. Other elements of the nuclear nonproliferation regime share this seeming contradiction. For example, the International Atomic Energy Agency, the international body charged with verifying state compliance with the NPT, boasts 164 members but requires substantial commitments on the part of member states. International security institutions in general see wider initial adherence than treaties in most other issue areas. In one dataset, multilateral security agreements boast an average of 39 original signatories, compared to 27 for human rights treaties and 16 for environmental treaties (Koremenos 2013).

The wide membership in the nuclear nonproliferation regime and other international security institutions defies the conventional wisdom of much of the literature on international organizations (IOs), which expects to see a trade-off between the breadth and depth of treaties. Scholars have argued that treaties with higher levels of legal obligation or deeper cooperation will necessarily have a narrower membership. This idea has even turned into a kind of prescription for policymakers: crafters of international treaties that hope to achieve deep cooperation are advised to limit the number of states invited to join (Downs, Rocke, and Barsoom 1996; 1998).

---

1 Economic and trade agreements have a similar level of initial adherence to security institutions in these data, with 41 original signatories on average (Koremenos 2013).
This article asks why the nuclear nonproliferation regime and other international security institutions seem to achieve both depth and breadth, and in doing so addresses the more general but no less puzzling question of why states join security institutions in the first place. The IO literature has focused its attention on human rights and trade institutions, emphasizing the domestic legal and interest group mechanisms behind ratification decisions. Security treaties, however, usually lack the distributional consequences capable of mobilizing interest group politics, and often have little or no effect on domestic law. The international security literature, when it considers international institutions at all, tends to see state ratification decisions as expressions of international power, usually by the United States. The US role in prodding states to join the nuclear nonproliferation regime, however, may well be overstated, and in any case this theory struggles to account for the variation we see in the decisions of states to join the NPT over time.

In this paper, I argue that international security institutions like the NPT harbor no breadth versus depth tradeoff. Instead, a deeper institution actually encourages wider membership, as prospective member states seek to ensure that their commitment to forgo nuclear weapons will be verifiably reciprocated by others within the treaty. Shallow treaties provide no such assurance. I test this theory using data on the NPT adherence of all non-nuclear weapons states since the treaty opened for signature in 1968, finding that various indicators of the NPT’s depth—the recent history of member state compliance, the strength of verification measures, and the willingness of the international community to punish violators—are strongly associated with state decisions to join the treaty.

The paper proceeds in four parts. First, I discuss the existing literature on the breadth versus depth tradeoff and present a theory of depth and treaty adherence in international security institutions. Next, I examine how my theory applies to the case of the NPT. Third, I
describe a quantitative model of the decisions of states to join the NPT and explain my results.

Finally, I conclude with a brief discussion of the implications of this theory for the broader IO literature.

**Theorizing Depth in International Security Institutions**

Most of the IO literature uses the concept of depth as a short-hand for the level of cooperation engendered by an institution, but this broad idea harbors a number of definitions. For some, depth is a proxy for the overall constraining power of an agreement: “it is most useful to think of the treaty's depth of cooperation as the extent to which it requires states to depart from what they would have done in its absence” (Downs, Rocke, and Barsoom 1996: 383). But this view risks conflating cause and consequence—if a treaty’s depth is synonymous with its constraining power, then depth cannot help us understand why some treaties are effective and some are not. The constraining power of an agreement can also vary across members, while depth is usually thought of as an attribute of the institution as a whole.

The concept of depth is sometimes used to capture the ambition of a collective effort. In formal treatments, depth is frequently defined as the level of policy enacted by a particular institution—this translates to how much pollution is reined in by an environmental treaty, for example, or how many categories of weapons are prohibited by an arms control agreement (Downs, Rocke, and Barsoom 1998; Gilligan 2004). But this conception, too, seems to miss something of what we informally mean by depth. An ambitious agreement in terms of its policy goals could be either deep or shallow, depending on how it is designed. The Kyoto Protocol, for example, is not lacking in ambition—it seeks to “stabilize greenhouse gas concentrations at a level that would prevent dangerous anthropogenic (human induced) interference with the climate system” (United Nations 1992)—but is more often described as “ineffective” than “deep” (Victor 2004). Conflating policy level and depth poses particular problems for analysis of
international security institutions, where the policy level is often dichotomous—states are prohibited from developing nuclear weapons, with no middle ground. If depth is synonymous with the policy outcome espoused by a treaty, then an agreement to forgo nuclear weapons will carry a particular level of depth without regard to other design features of the institution.

A more satisfying understanding of depth would move beyond just the effectiveness of the institution or its policy ambition. In this paper, then, I follow Bernauer et al. (2013) in treating depth as a multidimensional concept that incorporates several design characteristics of international institutions, including obligation, monitoring, and enforcement. Fundamentally, deep institutions are those that ask something of their membership: to comply with significant obligations, to subject themselves to monitoring and verification measures, and to strongly enforce the provisions of the agreement. But institutions need not be uniformly deep across all these aspects of the agreements. A treaty may require intrusive monitoring measures, for example, but have only shallow enforcement.

An intuitive argument from the literature on international organizations suggests that institutions with broad memberships must necessarily sacrifice some level of constraining power over the international behavior of the states that join (Downs, Rocke, and Barsoom 1998). There are several reasons this broader versus deeper trade-off might exist. Large groups might fail to achieve widespread cooperation because of increased transaction costs, less credible threats of punishment against cheaters, and heterogeneous time horizons (Oye 1986). The formal structures necessary for broad multilateral action may introduce inefficiencies over more flexible, informal cooperative mechanisms, or the attendant bureaucracy may invite agency loss (Kahler 1992; Lake and McCubbins 2006). Preference heterogeneity in large institutions may limit states’ ability to engage in costly enforcement behavior, such as the imposition of economic sanctions (Early and Spice 2014). In organizations where depth is at least partly determined by member
state voting, the underlying distribution of state preferences may give rise to this trade-off. If depth is set according to the preferences of the median state, for example, then dropping the member state with the shallowest preferences will result in a deeper institution (Downs, Rocke, and Barsoom 1998).  

This result is not limited, however, to institutions in which policy levels are determined by membership voting. The depth of international security institutions is more often set by some drafting state or coalition of states; taking the depth of an institution as given, other states make the decision to adhere or abstain. Even in these institutions, reducing depth might lead to increased membership by making the agreement palatable to more states. Figure 1 provides an illustration. Three states, A through C, vary in the maximum level of depth they will tolerate in an institution, and will join the institution if it achieves a depth at their maximum level or below. Institution $I_1$, with depth at $d_1$, will entice only state C to join, because $d_1$ is greater than the maximum depth tolerated by states A and B. If the drafting state proposes a lower level of depth, say $d_2$ as enacted by institution $I_2$, both states B and C will join, and all states will join institution $I_3$ at depth $d_3$. For a plausible set of decision rules and distribution of state preferences, then, the broader versus deeper trade-off can apply even to non-voting organizations like international security institutions.

Belief in this trade-off is by no means universal, of course. Kahler (1992) sees great power collaboration within institutions as an important mechanism for enabling substantial cooperation in multilateral institutions. Gilligan (2004) shows that the logic of the broader versus deeper trade-off breaks down when institutions assign different levels of policy across member states. If state policies are allowed to vary within an organization, as they often do, for

---

2 This is true as long as the median state prefers more depth than the state with the shallowest preference. See Downs, Rocke, and Barsoom (1998).
example, in international environmental or economic agreements, increasing the overall depth of an institution no longer leads to a narrower membership. But assuming a uniform policy across states does seem reasonable in the context of international security institutions. Here, treaties tend toward expressions of collective restraint—states agree not to develop some weapons capability, take over some territory, or cross some red line, as long as others do the same. Although there are exceptions, members of international security institutions are largely charged with enacting identical policies; militarizing outer space or occupying Antarctica is outlawed for all treaty members, not just a few.  

**Depth and the cost of IO membership**

The depth of an international agreement affects the decisions of states to join because depth is closely tied to the cost of membership. There are two broad categories of membership costs in international security institutions that are related to the institution’s depth. First, IO

---

3 Some security institutions, the NPT perhaps most prominently, have different tiers of membership that feature different levels of policy. In the NPT, the P-5 states (the United States, Russia, the United Kingdom, France, and China) are exempted from the general prohibition on the development of nuclear weapons. For analytic clarity, however, it can be useful to treat the NPT as two separate agreements: one which applies to the non-nuclear weapons states and one which applies to the P-5. Both of these sub-treaties can be seen as assigning a uniform policy level to its respective members.

4 There are also a number of costs and benefits of IO membership that are largely independent of the level of depth within the institution. Members of the NPT, for example, might see benefits in terms of international prestige (Walsh 2005), eligibility for development assistance (Brown and Kaplow 2014), side payments such as arms transfers (Erickson and Way 2011), and the approval of domestic constituencies (Cole 1997; Lantis 2008).
membership carries costs associated with the actual implementation of obligations—including verification, and enforcement—required by the agreement; such costs generally increase with institutional depth. Treaties with no verification measures, for example, are less expensive for member states to implement than those with intrusive inspections. In the nuclear nonproliferation regime, states collectively fund the IAEA’s inspection and verification efforts, and even states with no nuclear facilities contribute to the IAEA’s budget (International Atomic Energy Agency 2014). In addition to the cost of the inspectors themselves, a state’s nuclear power industry must make costly accommodations for international verification, adopting complicated accounting procedures and providing extensive training on international safeguards for its own personnel. Enforcement of an agreement’s provisions can also be quite costly. While the NPT lacks a formal enforcement mechanism, collective sanctions against treaty violators carries a substantial economic cost, not to mention the political and diplomatic costs of mobilizing the international community. A treaty with shallow enforcement, where no action is taken to punish violators, will carry lower costs for members than one in which collective enforcement is common or expected.

Costs due to the specific obligations of a treaty can be significant, but they are likely to be dwarfed for most states by a second category of membership costs—those that arise from the compliance behavior of others. These costs are analogous to the “sucker’s payoff” in the familiar two-person prisoner’s dilemma, which accrues to a player who defies his or her dominant strategy and cooperates, only to be faced with an opposing player who fails to reciprocate. The sucker’s payoff is the worst of the player’s possible outcomes in this game—it carries all the costs of cooperating and none of the benefits. In the n-player version of the prisoner’s dilemma that is often used to model international cooperation, the cooperating player’s payoff is transformed into a continuum of possible outcomes, where the payoff is usually expressed as a function of the
number of other players that choose to cooperate.$^5$ One end of that continuum is universal cooperation, offering the highest payoff for a cooperating player. This is the equivalent of the welfare-maximizing outcome in the two-player game, when both players choose to cooperate. But at the other end of the continuum, as fewer and fewer other players opt for cooperation, lies the $n$-player version of the sucker's payoff. There are substantial costs for cooperators whose choice is not reciprocated.

In the context of international security institutions like the NPT, the cost of the sucker's payoff can be quite pronounced.$^6$ Putting aside efforts to develop nuclear weapons and joining the NPT makes states more vulnerable to the threats of adversaries that have abstained from or cheated on the treaty, because NPT membership makes the development of nuclear weapons more costly.$^7$ Secret nuclear weapons programs are more likely to be discovered in NPT states, which must submit their domestic nuclear infrastructure to regular international inspection. And even if illicit weapons work can be kept hidden, member states will bear additional costs in this concealment effort. Once identified, nuclear weapons programs probably exact a greater toll on member states in terms of international standing and reputation effects, and the violation of a major international treaty might also carry domestic political consequences for the offending leaders. Finally, the revelation of nuclear weapons development is more likely to lead to

---

$^5$ For a discussion of the $n$-player iterated prisoner’s dilemma generally, see Yao and Darwen (1994). For an application of the linear version of the game to international cooperation, see Barrett (1999).

$^6$ While this underlying game structure is not unique to international security institutions, neither is it universal. In international human rights treaties, for example, the benefit of states to cooperation is largely independent of the cooperation of others, eliminating the sucker’s payoff and minimizing the costs of membership that arise from the noncompliance of others.

$^7$ In this paper, I use “violation,” “noncompliance,” and “cheating” interchangeably to mean the pursuit of nuclear weapons while an NPT member. There are, of course, other ways that states might fail to abide by their NPT commitments, but I do not consider them here. For a discussion of legal findings of noncompliance within the NPT, see Goldschmidt (2010).
diplomatic pressure, sanctions, or attack against an NPT member state than against a non-member.

These additional costs for developing nuclear weapons under the NPT hamper members’ ability to respond to the nuclear provocations of others. The vulnerability of member states is exacerbated by the relatively long timeline for nuclear weapons development, and uncertainty about the technical progress of others’ programs. Consider a hypothetical state that receives some indication that an adversary has an early nuclear weapons effort. If the state is not within the treaty, it is free to begin low-level nuclear weapons work of its own. Such a weapons program would have a relatively low chance of being discovered, and provides the state with a low-cost hedge against the risk that the adversary’s weapons effort actually exists and ultimately bears fruit. If the state is an NPT member, however, launching a weapons program of its own—given the early state of the adversary’s program and uncertainty over its progress—becomes a tremendously risky move. The member state thus has a strong incentive to delay action until the true extent of the adversary’s nuclear effort becomes more clear. But waiting just makes the state more vulnerable. By the time the adversary’s program advances to the point where it justifies the costs associated with the decision to violate the NPT, the member state will have fallen significantly behind.

---

8 On the ability of the NPT to lengthen the timeline for nuclear weapons development, see Quester (1972).
9 It is possible to adopt a nuclear hedging strategy that does not involve clearly violating the NPT. In recent years, member states—most notably Iran—have sought to build up civilian nuclear infrastructure that is permitted by the NPT but applicable to future weapons work. These states reduce the time it would take to develop a nuclear weapon should they make the decision to pursue that capability. This strategy can be seen as a means of reducing the cost of membership and the risk of earning a sucker’s payoff. On nuclear hedging strategies, see Wohlstetter (1976) and Levite (2003). For a discussion of nuclear hedging in the context of Iran’s nuclear program, see Kaplow and Gibbons (2015).
Fundamentally, then, states that join and comply with IOs are gambling that others will reciprocate—a gamble that is made more precarious in the realm of international security by the fact that the true level of compliance in the institution is often quite uncertain. The depth of the international institution, however, can play an important role in resolving compliance uncertainty and reducing the expected cost of membership. Depth functions as a signal of the overall effectiveness of the IO, reassuring states that they will not be taken advantage of, that their cooperation is likely to be reciprocated. Deep institutions back up their policy ambitions with real obligations, holding members to their commitments by monitoring and verifying compliance and by punishing violators. As institutions increase in depth, the sucker’s payoff becomes increasingly unlikely, and the attendant cost of membership due to the noncompliance of others decreases.

This is not to say that only members of the institution benefit from increasing depth. To the extent that deeper international security institutions are more successful in constraining state behavior, they improve the security even of states that have abstained. But while security benefits accrue to members and non-members alike, only for member states does depth help to mitigate the significant costs associated with forgoing nuclear weapons or some other security policy. Member states simply have more skin in the game than do abstaining states. A winning hand of poker would benefit any player, but the benefits are greater for those that have placed larger bets.
Depth and treaty adherence

States choose to adhere to international institutions when the cost of membership is less than the cost of abstaining from the treaty.\(^\text{10}\) It seems reasonable to assume that there is some cost associated with the decision not to join a treaty—states may face international or domestic pressure to adhere to IOs, or may forgo international prestige and more tangible benefits by electing not to join. In the context of the NPT, for example, the United States and the Soviet Union exerted some pressure on states within their sphere to adhere to the treaty (Gibbons 2015), and joining was popular with the public in some states (Cole 1997; Lantis 2008). More convincing, perhaps, is the fact that nearly all states have chosen to join the NPT, despite its evident cost. If there were no cost to remaining outside the institution, then we would have expected more states to have taken that course.

Figure 2 illustrates the calculus states face in determining whether or not they will join an international institution. The figure depicts cost curves for states A through C. The cost of membership in the institution (on the y axis) is high when the institution is shallow, as states fear making themselves vulnerable with little expectation that others will reciprocate. Costs generally decrease, however, as depth increases (along the x axis); states can be more confident that others will abide by the treaty when depth is higher, so their expected cost for joining drops accordingly. For clarity, the cost of abstention (\(C_{Abs}\)) here is assumed to be common to all states. Its negative slope reflects the fact that increasing depth provides a security benefit to non-members as well as members.

\(^\text{10}\) Adherence is the act of joining an institution; it encompasses ratification (when the state is a signatory to the treaty) as well as accession (joining without having already signed the treaty). Abstention is the decision not to join an institution.
Figure 2 highlights the relationship between depth and treaty membership. For any given depth, states with a cost of membership below the cost of abstention will elect to join, while states with costs of membership higher than the cost of abstention will choose adherence. In the shallow institution $I_1$, for example, the cost of abstention exceeds the cost of membership for only state $A$. State $A$, then, will elect to join the treaty, while states $B$ and $C$ abstain. But increasing the depth of the institution—to the level of $I_2$, for example—prompts more states to join. At the greater depth of $I_2$, the cost of abstention exceeds the cost of membership for states $A$ and $B$, leaving only state $C$ out of the institution. Increasing breadth, then, can be a direct consequence of the increasing depth of the IO. This is the central hypothesis of this paper.

*Depth hypothesis: States are more likely to join international security institutions of greater depth.*
Note that the cost of joining creeps upward again at high levels of depth. This is because the marginal benefit of depth for cooperation is decreasing, while the marginal costs associated with the obligations of the treaty are increasing. With the low-hanging fruit gone, moving from a strict inspection regime to near-perfect verification, for example, can be extremely costly. When depth is high, therefore, the breadth versus depth tradeoff can reassert itself. In some sense, the nature of the relationship between breadth and depth is a function of where on this continuum an institution is located. When an institution is shallow, adding depth is likely to spur more adherence; when an institution is already deep, increasing depth further may push states away. In the realm of international security institutions, we are probably much closer to the shallow end of the spectrum than to the deep, but other types of institutions—economic treaties, for example, or political unions such as the EU—may well see the type of breadth versus depth tradeoff theorized for highly constraining institutions.

States can have the general cost functions shown in Figure 2 and still have strong preferences over the specific obligations embodied in a treaty; that is, some mechanisms for achieving depth are likely to be costlier for a particular state than others. It is thus not surprising that states bargain to reduce the depth of the institution in some cases, while still preferring a deeper treaty overall. The effect of specific concessions on the cooperation of others—and the risk of a sucker's payoff—puts a kind of limit on the natural tendency of states to seek loopholes that benefit their particular situation. States will still seek concessions that provide them with some benefit, but only to the extent that the expected benefit exceeds the cost in terms of the increased danger that others will exploit the concession to violate the treaty.
In early negotiations over the NPT, for example, several states sought to exempt from international safeguards nuclear material intended for use in nuclear submarines.\textsuperscript{11} The United States initially objected, concerned that states would take advantage of this exemption to divert nuclear material to secret weapons programs, but ultimately allowed the exemption into the final treaty. Notably, the two states leading the charge for a naval propulsion loophole to the NPT were the Netherlands and Italy (Moltz 1998: 109; Fischer 1997: 272; U.S. Arms Control and Disarmament Agency 1965). Both states hoped to deploy nuclear-powered submarines in the future, which made restrictions on naval applications of nuclear material particularly costly for them, but these states were otherwise strong supporters of nonproliferation.\textsuperscript{12}

The analysis above assumes that the expected cost of joining and violating an agreement is always greater than the cost of abstention. That is, states elect not to become members of the institution in the first place if they anticipate having to violate their obligations; they do not join with the intent to cheat. This seems like a reasonable assumption for most states. Noncompliance generally carries additional costs over and above adopting the same course of action outside the treaty, as discussed above.

The history of the NPT supports this assumption, implying that most states join the treaty in good faith.\textsuperscript{13} Relatively few states have violated the NPT by seeking nuclear weapons while members, and of those that have, most did not have weapons programs when they joined. But

\textsuperscript{11} For a detailed discussion of the naval nuclear propulsion loophole to the NPT, see Kaplow 2015.

\textsuperscript{12} Italy's broad support for nonproliferation goals did not always translate into direct support for the NPT. The country “had been quite difficult” in negotiating the treaty, according to the Director of the U.S. Arms Control and Disarmament Agency (U.S. National Security Council 1968). On Italy's stance toward NPT ratification more broadly, see U.S. Department of State (1968).

\textsuperscript{13} By “good faith,” I mean merely that the state joins the treaty with the intent to comply, not that it will remain in compliance if circumstances change. “Bad faith” states are those that join the treaty while in noncompliance, or with the intent to violate.
the remaining few—the “bad faith” states—suggest that it is not correct to assume that all states join security institutions with the intent to comply. Bad faith states are likely to have a different cost function associated with adherence than states that plan to join and comply. Prospective cheaters need not worry about making themselves more vulnerable to exploitation by others—earning the sucker’s payoff—because they intend to pursue the proscribed behavior regardless of whether or not they join the institution. Instead, bad faith states must concern themselves with the expected cost of noncompliance—the chances that violations will be discovered and punished, and the type of punishment they will suffer. Deeper treaties are costlier to cheaters for the same reason they are attractive to compliers—they promise a higher expected cost of noncompliance.

The cost of joining and violating the treaty, then, unlike the cost of joining and complying, is increasing in depth. Figure 3 illustrates the cost curves of bad-faith states. As the depth of an institution increases (along the x axis), the cost of membership to states A through C increases. The cost of abstention ($C_{Abs}$) remains the same as in Figure 2, although we would expect this cost to be absolutely higher for states that see joining and violating as preferable to staying out of the treaty altogether. As before, when the cost of abstention is greater the cost of membership, the state chooses adherence. But bad-faith states become less likely to join as depth increases. In Figure 3, a shallow institution, $I_1$, garners two members (states A and B). Increasing the IO's depth, to the level of institution $I_2$, discourages membership. At $I_2$, only state A elects to join. The breadth versus depth tradeoff thus reemerges for states that plan to violate the treaty, suggesting the following hypothesis.

---

14 In the Jo and Gartzke (2007) data, Libya, Taiwan, North Korea, and South Korea joined the NPT in the same year that they maintained active nuclear weapons program. South Korea, however, is a questionable case of bad-faith adherence, as South Korea also is coded as ending its program in the year it joined the NPT.
**Bad faith hypothesis:** States that intend noncompliance are less likely to join international security institutions of greater depth.

**Depth and the Decision to Join the NPT**

The history of the NPT suggests that states considered the depth of the treaty—and the risk of a sucker’s payoff—when negotiating the treaty and deciding whether to join. In the run-up to the negotiations over the NPT, it was commonplace for states to express a preference for depth in the new treaty. Egypt’s representative to disarmament talks in 1965, for example, warned the assembled international delegations:

“[A]ny international agreement on non-proliferation should not have what are known as ‘escape clauses,’ drafted in a way which weakens the agreement’s importance from the very beginning, even before it is signed; otherwise it would not be a real agreement but an artificial facade to deceive world public opinion.”

(U.S. Arms Control and Disarmament Agency 1966: 487)
This general sentiment, a call for an NPT with substantial depth, was echoed as a main principle for future negotiation in a UN General Assembly resolution that same year: “The treaty should be void of any loop-holes which might permit nuclear or non-nuclear Powers to proliferate, directly or indirectly, nuclear weapons in any form” (U.S. Arms Control and Disarmament Agency 1966: 533).

In private, states expressed a more specific concern about the risk associated with giving up their right to nuclear weapons while others maintained freedom of action. In high level meetings with US counterparts, Japanese diplomats repeatedly emphasized the difficulty Tokyo would face in joining the treaty, if other nuclear-capable states were not similarly constrained (U.S. Department of State 1967). Others aired similar worries. In meetings with Henry Kissinger in 1966, for example, West German leaders asked “if [West Germany] signed the treaty, where would it be if countries like India and Israel acquired nuclear weapons” (Kissinger 1966).

In treaty negotiations, non-nuclear weapons states frequently expressed concern about the willingness of the P-5 to reduce their nuclear arsenals, and this too can be seen in the context of their potential vulnerability as NPT members. Because the P-5’s weapons programs were not restricted by the treaty, other member states were that much further from the cooperator’s ideal payoff and that much closer to the sucker’s payoff. Even if all other states reciprocated in their cooperation, the special status of the P-5 would put other members at some risk of nuclear aggression. This was particularly true for states like West Germany and India, key adversaries of whom were not subject to the constraints of the treaty. It is thus not surprising that non-weapons states pushed hard for stronger disarmament provisions in the treaty, for the obligations of non-
weapons states to extend to P-5 states as much as possible, and for security assurances from P-5 states as a condition of adherence.\textsuperscript{15}

When negotiations on the NPT concluded and it opened for signature in 1968, its provisions established a particular level of depth for the treaty—the NPT requires states to place their nuclear facilities under safeguards, for example, and lacks any formal enforcement mechanism. But the design of the treaty is just a starting point; its actual depth is dynamic, changing over time as the cooperation of states ebbs and flows. Further, the extent of these changes varies across the different dimensions of depth.

The monitoring and verification regime within the NPT, for example, has moved steadily deeper since the early days of the treaty. Significant changes have occurred in three areas: the fundamental approach to nuclear safeguards, the technologies and procedures employed to verify compliance, and the scope of coverage of member states’ nuclear facilities and materials.

The NPT requires member states to put in place comprehensive safeguards, providing IAEA access to all of a country's nuclear material. The need for these “full-scope” safeguards was emphasized by India's 1974 nuclear test, which used nuclear material that had been diverted from civilian facilities provided by the United States and Canada (Tape and Pilat 2008). The resulting safeguards approach of the 1970s and 1980s focused on the verification of state declarations and ensuring that no nuclear material had been diverted from a safeguarded facility (Carlson et al. 1999; Gruemm 1983). The underlying safeguards philosophy shifted in 1991 with

\textsuperscript{15} Japan is a particularly influential example of the advocacy of non-weapons states, successfully arguing to move treaty language about the requirement of states to work toward disarmament from the preamble—where it was seen as mostly symbolic—to the body of the treaty. Japan also succeeded in winning an automatic review of the treaty every five years and an initial 25-year term for the NPT, both intended to allow non-weapons states to evaluate the progress of the P-5 in working toward disarmament (U.S. Department of State 1967). On the origins of the NPT's sunset clause, see Sokolski (2010) and Koremenos (2001). Efforts to convince the P-5 to provide security assurances to NPT members were less successful. For a full discussion of security guarantees under the NPT, see Simpson (2012).
the discovery that Iraq had harbored an extensive nuclear weapons program despite having full-scope IAEA safeguards in force. Rather than focus solely on the verification of state declarations to the IAEA, the mission of safeguards inspectors was expanded to address the completeness of declarations and the possibility of undeclared nuclear activities (Carlson et al. 1999; Goldschmidt 2001; Pellaud 2000; Tape and Pilat 2008).

The Iraq revelations also drove the introduction of new technologies and procedures into safeguards practice. Monitoring systems, for example, saw substantial improvements. The IAEA in the 1970s monitored spent-fuel pools using a pair of cameras set to take a photograph every twenty minutes. Inspectors would attempt to review thousands of black-and-white images at every visit, looking for any sign that the spent fuel had been moved. In the last twenty years, however, the IAEA has deployed more advanced unattended monitoring systems. Today’s systems can capture radiation and other measurements as well as images, and in some cases are monitored remotely (Schanfein 2008). While the IAEA has always benefited from satellite imagery provided directly from member states, only recently has it gained the capability to independently order and analyze commercial satellite imagery in support of safeguards goals (Chitumbo, Robb, and Hilliard 2002).

Environmental sampling techniques were introduced to IAEA safeguards in the mid-1990s, allowing IAEA inspectors to measure the isotopic composition of nuclear materials, potentially contradicting a state’s declarations (Donohue 1998; Donohue, Deron, and Kuhn 1994). Environmental sampling played an important role in building the noncompliance case against Iran. Agency inspectors in 2003 requested access to a suspected, but undeclared, centrifuge workshop. Iran eventually allowed inspections, but only after an attempt to decontaminate the facility—replacing the floor, repainting walls, and moving equipment to other locations. Still, environmental sampling there detected enriched uranium particles that could not
be explained by Iran’s existing declarations and led to Iran’s acknowledgment later that year that it had indeed conducted undeclared centrifuge testing at the facility using nuclear material (Samore 2005).

Safeguards technologies like environmental sampling are only useful if the IAEA has access to nuclear facilities, granted through a comprehensive safeguards agreement between the non-nuclear weapons state and the IAEA. A less stringent level of safeguards—the Small Quantities Protocol—was designed for states with little or no nuclear material and generally precludes international inspections. Since 1997, states also have had the option to voluntarily implement an Additional Protocol to their safeguards agreement, which carries further reporting requirements and enhanced access for IAEA inspectors. (Hirsch 2004). As of June 2015, 126 states had an Additional Protocol in force (International Atomic Energy Agency 2015). This increased access, combined with a new approach to inspections that makes use of new technologies and procedures, suggests that IAEA safeguards have become significantly more formidable over time.

Unlike monitoring and verification, enforcement of state commitments under the NPT seems to come in phases, corresponding somewhat, although not perfectly, with the ebb and flow of nuclear proliferation itself. While the treaty’s formal power to punish violators is limited—a finding of noncompliance merits merely a referral to the UN Security Council, for further action at its discretion—the international community does frequently act in a variety of ways to pressure states to comply.

\[ \text{16 The P-5 states have safeguards at some facilities as well, but these arrangements are voluntary and do not cover all nuclear material within the country.} \]

\[ \text{17 On the importance of the Additional Protocol for limiting proliferation generally, see Schulte (2010). The Additional Protocol and Small Quantities Protocol are not mutually exclusive, but the Small Quantities Protocol places limits on IAEA action even when an Additional Protocol is also present, for example by waiving the requirement that states provide initial declarations of nuclear materials and activities.} \]
The enforcement tools available to states range from unilateral attempts at persuasion to global sanctions regimes like those arrayed against Iran and North Korea in recent years. In the early years of the NPT, for example, pressure from the United States was instrumental in shutting down nuclear weapons programs in Taiwan and South Korea. In both cases, the implicit American threat to rethink security assurances if the weapons efforts continued probably played an important role.\(^{18}\) The mid- to late-1970s saw new U.S. legislation to mandate sanctions under particular nonproliferation circumstances, but legally mandated sanctions have always existed side-by-side with discretionary policy decisions that would have similar impact, and indeed with the mere threat of such sanctions (Speier, Chow, and Starr 2001). Multilateral sanctions aimed at nonproliferation—as in North Korea and Iran—are a more recent phenomenon.\(^{19}\)

Empirical studies of nuclear sanctions point to no steady pattern, but rather a series of highs and lows. Reynolds and Wan (2012) track 454 sanctions and positive inducements leveled against Iraq, Iran, Libya, and North Korea since 1990, finding that sanctions have largely leveled off since the 1990–1994 period, while positive inducements spiked in the late 1990s and have declined since.\(^{20}\) Miller (2014) takes a longer view, but focuses on nuclear sanctions threatened or imposed by the US. In his list of cases, sanctions are most prevalent when they first emerge as a prominent tool of US foreign policy; he identifies 15 sanctions episodes in the 1970s, but only 13 since.\(^{21}\)

\(^{18}\) On Taiwan’s nuclear weapons effort, see Albright and Gay (1998) and Burr (2007). On South Korea’s nuclear program, see Pollack and Reiss (2004).

\(^{19}\) For a discussion of the role of sanctions in nonproliferation policy toward North Korea, see Haggard and Nolan (2012). On sanctions against Iran, see Nader (2012).

\(^{20}\) As Reynolds and Wan (2012) point out, some of this trend can be explained by the removal of Iraq and Libya as sanctions targets by the early 2000s.

\(^{21}\) Most of the cases identified by Miller (2014) are not truly instances of enforcement or punishment, because they involve US sanctions (threatened or imposed) against states that are outside of the NPT.
The treaty’s depth of enforcement is necessarily related to the underlying depth of cooperation; only when states violate the NPT can their noncompliance be punished. Like enforcement activity, cooperation has varied over the lifetime of the nonproliferation regime. Measured in terms of the number of state violators of the NPT, cooperation was at its shallowest in the mid-1970s, mid-1980s, and early 2000s. Cooperation was at its deepest point—with only one violating state—prior to 1973 and after 2007.22

These three dimensions of depth, then, do not move in lock-step. The depth of verification has been steadily increasing, while enforcement and cooperation move in fits and starts. If states are more likely to join when depth is high, as my theory suggests, then changes in the NPT’s depth over time may help to explain variation in adherence to the treaty. I turn now to a more systematic examination of the relationship between depth and membership in the NPT.

A Quantitative Analysis of NPT Adherence

To better understand the role that depth plays in the decisions of states to join the treaty, I construct several quantitative models of treaty adherence using data on NPT membership between 1968—when the treaty opened for signature—and 2006.23 These data are structured as a pooled time series, with each observation representing a single country in a single year. I exclude from the analysis states that have acquired nuclear weapons, for the years after they have a nuclear weapons capability. The five nuclear weapons states recognized by the treaty are likely to face dramatically different costs and benefits for joining the NPT than the non-nuclear weapons states. And for the nuclear weapons states outside the treaty—Israel, India, Pakistan,

22 Violators are NPT member states pursuing nuclear weapons in a given year. Nuclear pursuit data is updated from Jo and Gartzke (2007).

23 Beginning the dataset in 1970, when the NPT entered into force, yields similar results. After 2006, only South Sudan (about whom full covariate data is unavailable) and states possessing nuclear weapons remain outside the treaty.
South Africa (for a time), and North Korea—the presumed price for joining is the dismantling of their existing nuclear capability. This toll seems sufficiently different from the calculus faced by other non-members that it makes little sense to lump them together in the model. Still, my findings do not depend on this modeling decision; I see largely the same results when excluding only the P-5 states, and when including all states in the analysis.

The dependent variable in these models is a dichotomous measure set to one if a state joins the NPT in a given year, and zero otherwise, using NPT membership data from Carcelli et al. (2014). Many states signed the NPT well before officially becoming a member through ratification. While becoming a signatory to a treaty does pose some limits on state behavior as a matter of international law, only states that have ratified or acceded to the NPT can properly be considered members, and so I use that standard here. Because states almost never withdraw from the NPT, I drop country-year observations beginning the year after states elect to join the treaty.24

My explanatory variable of interest is the depth of the NPT. I employ three different measures of depth, in an attempt to capture different dimensions of the underlying concept. First, to measure the depth of monitoring and verification within the treaty, I count the number of member states that allow inspectors access to their nuclear facilities in a given year—that is, those with a comprehensive safeguards agreement in force (Carcelli et al. 2014).25 Second, I measure the depth of enforcement as the number of member states that have had nuclear-related sanctions threatened or imposed against them in a given year (Morgan, Bapat, and Kobayashi

24 North Korea, the lone state to leave the NPT, reenters the dataset briefly after its withdrawal from the treaty in 2003. It exits the dataset again in 2006 following its acquisition of nuclear weapons.

25 Excluded from this count are states that have enacted a Small Quantities Protocol to their safeguards agreement, because the IAEA cannot inspect nuclear facilities in those states.
Third, to capture the underlying depth of cooperation in the NPT, I include the treaty’s recent history of compliance—its track record—measured as the three-year trend in NPT violators. Each of these variables is lagged by one year to avoid problems of mutual causality in the model, in which the decision of a particular state to join the treaty in turn affects the NPT’s depth.

These measures of treaty depth operate at the system level and do not vary by state; in a given year, all states in the dataset will have the same value for the variables representing depth of verification, enforcement, and cooperation, respectively. One concern, then, is that these variables merely proxy for some underlying cause associated with the time trend in NPT adherence. But, as discussed above, the different dimensions of depth do not move together over time. The variation among these three variables should provide some reassurance that depth is not simply standing in for another driver of NPT membership in my quantitative models.

Figure 4 summarizes these measures of the depth of the NPT, along with the pattern of treaty adherence over time. The top line of rectangles represents the number of states joining the NPT in a given year (darker rectangles signify a higher number of new members in that year). The next three lines depict the depth of verification, enforcement, and cooperation, respectively, with darker rectangles representing greater depth. As the figure shows, there is no clear relationship among the three variables.

---

26 I count only sanctions episodes that are intended to deny strategic materials to the target state, and do not include sanctions intended to prevent states from sharing technology or materials with third parties.

27 The three-year trend in NPT violators for year $t$ is the number of NPT member states with nuclear weapons programs in year $t$, minus the number of violators in year $t-3$. 
Omitting variables from the model risks biasing the results if those variables affect both the dependent variable (adherence) and the key explanatory variable of interest (depth). Other determinants of state membership probably fall into this category, because the membership decisions of particular states are in turn likely to influence the treaty's overall level of cooperation, verification, and enforcement. I thus include in the model explanatory variables to represent several other possible drivers of state adherence.

The NPT differs from many security institutions in that it has some distributive economic consequences. One implication of the treaty, although it has not always functioned this way in practice, is that member states ought not to share nuclear technology with non-nuclear weapons states that are not party to the NPT. States with commercial nuclear interests, then, have an
incentive to support adherence to the NPT out of concern that abstaining from the treaty will limit their access to foreign nuclear materials and technical assistance. I thus include in the models a measure of the percentage of domestic energy production that comes from nuclear power (World Bank 2015).

A state's regime type may also influence its willingness to join the NPT (and its willingness to comply once it adheres to the treaty). Democracies might be more susceptible to domestic pressure to become a member of a high-profile international agreement; particularly so for states in which peace or environmental movements hold some political power. It is also possible that democratic leaders—because they are likely to face domestic opposition to treaty violations—would be more cautious about joining a treaty in the first place. To capture the effect of regime type on NPT adherence, I include in my models the Polity project's measure, ranging from -10 (most autocratic) to 10 (most democratic) (Marshall, Jaggers, and Gurr 2010).

Some states simply are not capable of developing nuclear weapons. Such states lack the financial, material, and human resources for a successful nuclear program, or the means or opportunity to acquire them from another state. For such states, the cost of joining the NPT is quite low (Jensen 1974; Way and Sasikumar 2004). While nuclear-capable states that join an effective treaty are forgoing the option of nuclear weapons in the future, states that lack the capability to develop weapons are not forgoing anything. While there is always the possibility that a state will acquire the necessary resources for a weapons effort in the future, the present benefits of the treaty may swamp the discounted chance of future nuclear flexibility. We thus would expect higher levels of nuclear capability to be associated with a reduced likelihood of NPT adherence. I include two variables to identify the effect of supply-side constraints on membership in the treaty. To capture capability broadly, I employ real GDP per capita, using data
International pressure, particularly from the United States, has undoubtedly encouraged many states to join the NPT. The US role in driving adherence, however, is sometimes overstated, particularly in reference to the early years of the treaty. President Nixon was openly ambivalent about the fledgling NPT. In a 1969 national security council meeting, for example, in the context of discussions about convincing allies like West Germany to adhere to the NPT, the president “wanted it understood that there was to be no arm twisting of other states on the NPT issue, that it is completely up to them as to whether or not they follow [the] U.S. lead [by joining the treaty]” (National Security Council 1969). By 1972, US diplomats were ramping up pressure on Japan to join the NPT. But when Japanese Prime Minister Sato asked President Nixon whether Tokyo should quickly ratify the treaty, Nixon replied:

> “Each nation should handle this problem in the light of its own circumstances. It is not a matter for us to decide and we respect the right of each nation to decide for itself in the light of its own desires. The United States...is not exerting pressure. In fact...Japan might take its time and thus keep any potential enemy concerned.”

(“Conversation Between Nixon and Kissinger” 1972)

Despite the sometimes spotty record of the United States in advocating for the treaty, we would expect states that share US policy goals to be more likely to adhere to the NPT, both because nonproliferation is more likely to be in the interests of such states and because the

---

28 Jo and Gartzke’s (2007) nuclear capability data is available only through 2002. This measure changes very slowly over time, however, as we would expect for a measure based on such factors as the presence of uranium deposits and the ability to produce certain chemicals and industrial products. I thus simply repeat the 2002 data for all abstaining states through 2006. My results, however, are robust to conducting the full analysis through only 2002, or to omitting this variable altogether.

29 On United States leadership in the nonproliferation regime more broadly, see Gibbons (2015).

United States may be in a better position to exert pressure on those states in its sphere. I thus include in the models a measure of a state’s foreign policy affinity with the United States, based on a spatial model of UN General Assembly voting (Bailey, Strezhnev, and Voeten Forthcoming). Of course, the United States is not the only state exerting pressure on others to join. States may be more susceptible to such pressure from their allies, and the presence of a military alliance may make states feel more secure in forgoing weapons, making NPT adherence correspondingly less costly. To capture this dynamic, I create a dichotomous variable that takes on the value of one when a state has a defense pact with an NPT member, using alliance data from Gibler (2009).

There is also a temporal dimension to international pressure. The NPT’s quinquennial review conferences focus international attention on the extent of treaty adherence and may also help to activate domestic constituencies and civil society in support of membership. Each model thus includes a dichotomous variable set to one if an NPT review conference occurs in that year.

Where nuclear suspicion runs high, states may see accession to the NPT as a means of reducing tensions and building confidence between rivals. Pakistan, for example, for many years held out the prospect of joining the NPT if only India would do the same (Solingen 1994). Brazil and Argentina, while resisting the NPT until the mid-1990s, joined a regional nuclear security regime at the same time to help build confidence (Reiss 1995). Even Israel has discussed the possibility of a nuclear weapons free zone in the Middle East as part of a larger peace agreement (Solingen 1994). Even if states join the treaty for other reasons, their rivals may feel more comfortable signing on to the NPT if they are not doing so unilaterally. To capture this possibility, I construct a dichotomous variable that takes on a value of one if any of a state’s rivals has joined the NPT in the preceding three years, using rivalry data from Thompson and Dreyer (2012).
Finally, the decision to forgo nuclear weapons in the future becomes costlier as the perceived usefulness of nuclear weapons increases. States facing significant external threats may be less willing to bear the cost of limited flexibility down the road and may even see the decision not to join the NPT as creating a kind of strategic ambiguity around their nuclear intentions. States that face few significant external threats accept much lower costs in forgoing weapons development. To capture the dynamic of reduced demand for nuclear weapons, and thus the reduced cost of adherence, I create a dichotomous variable that takes the value of one if a state has engaged in an interstate armed conflict in the previous five years, using data from Gleditsch et al. (2002).

I conduct my analysis using logistic regression models, reporting robust standard errors clustered by country. Temporal dependence is a concern in time-series cross-section data of this kind. To address this, I include a cubic polynomial of the number of years since the state was first able to join the treaty or, in the case of North Korea after its withdrawal from the NPT, the number of years since it was last a member (Carter and Signorino 2010).

**Quantitative results**

Results from this analysis are shown in Table 1. Each model focuses on a different dimension of treaty depth, and increasing depth in each is significantly associated with state adherence, consistent with my theory. In model 1, as the number of states with robust
### Table 1: Logit Analysis of NPT Adherence 1968 – 2006

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Depth</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verification (robust safeguards)</td>
<td>0.021</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enforcement (sanctioned states)</td>
<td></td>
<td>0.293</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.124)</td>
<td></td>
</tr>
<tr>
<td>Cooperation (trend in violators)</td>
<td></td>
<td></td>
<td>-0.161</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.068)</td>
</tr>
<tr>
<td><strong>Domestic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuclear power</td>
<td>0.022</td>
<td>0.018</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.025)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Regime type</td>
<td>0.038</td>
<td>0.036</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Real GDP per capita</td>
<td>-0.002</td>
<td>-0.003</td>
<td>-0.028</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Nuclear capability</td>
<td>-0.142</td>
<td>-0.130</td>
<td>-0.138</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.051)</td>
<td>(0.047)</td>
</tr>
<tr>
<td><strong>International</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>US affinity</td>
<td>0.337</td>
<td>0.344</td>
<td>0.331</td>
</tr>
<tr>
<td></td>
<td>(0.160)</td>
<td>(0.164)</td>
<td>(0.155)</td>
</tr>
<tr>
<td>Defense pact with NPT member</td>
<td>0.687</td>
<td>0.443</td>
<td>0.699</td>
</tr>
<tr>
<td></td>
<td>(0.203)</td>
<td>(0.200)</td>
<td>(0.201)</td>
</tr>
<tr>
<td>Review conference year</td>
<td>0.551</td>
<td>0.636</td>
<td>0.518</td>
</tr>
<tr>
<td></td>
<td>(0.272)</td>
<td>(0.277)</td>
<td>(0.272)</td>
</tr>
<tr>
<td>Rival membership</td>
<td>1.023</td>
<td>0.857</td>
<td>0.964</td>
</tr>
<tr>
<td></td>
<td>(0.290)</td>
<td>(0.285)</td>
<td>(0.281)</td>
</tr>
<tr>
<td>Conflict history</td>
<td>-0.723</td>
<td>-0.637</td>
<td>-0.778</td>
</tr>
<tr>
<td></td>
<td>(0.382)</td>
<td>(0.391)</td>
<td>(0.369)</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.573</td>
<td>-0.775</td>
<td>-1.620</td>
</tr>
<tr>
<td></td>
<td>(0.173)</td>
<td>(0.231)</td>
<td>(0.156)</td>
</tr>
<tr>
<td>N</td>
<td>1238</td>
<td>1238</td>
<td>1238</td>
</tr>
</tbody>
</table>

Logit coefficients with robust standard errors, clustered on country, in parentheses. Bold values are statistically significant ($p < 0.05$). A cubic polynomial of the years since the treaty opened or since a state’s withdrawal are included in all models but not shown.
safeguards arrangements increases, states become more likely to join the treaty. In model 2, an increase in the number of member states threatened with or actually targeted for nuclear-related sanctions is associated with an increased probability that a state will adhere to the NPT. And in model 3, the negative coefficient on the three-year trend in violators indicates that movement toward greater cooperation within the treaty is associated with increased adherence.\textsuperscript{31}

The variables representing other explanations for the decisions of states to join the NPT are largely significant and behave as we would expect. Among the domestic drivers of adherence, the percentage of a state’s power production from nuclear sources is significant only in model 3, but there it suggests that states with civilian nuclear capabilities are more likely to join the treaty. This finding is interesting in light of the results for other variables that proxy for state capability. The coefficients on real GDP per capita and a composite index of nuclear capacity are negative and significant in all models. A state that is more capable of violating the treaty—and that thus faces a higher cost of membership—is less likely to become a member. Regime type also is a significant driver of adherence; more democratic states are more likely to join.

International factors, too, are consistent with theory. Affinity with US foreign policy positions and defense pacts with members of the NPT are significantly associated with an increased propensity to join the treaty. State adherence is significantly more likely to occur in concert with one of the NPT’s review conferences—at least in models 1 and 2—probably reflecting increased pressure or stronger inducements for states to join at these key times. States also appear more likely to join the NPT when a rival has recently joined, suggesting the treaty can function as a kind of confidence building measure between rivals—or at least that rival states tend to join the treaty in pairs. The coefficient on conflict history is negative and statistically

\textsuperscript{31} Higher values for the three-year trend in violators indicate shallow cooperation, while lower values indicate greater depth.
significant in model 3, lending some support to the idea that states with a higher demand for nuclear weapons will bear a higher cost for treaty membership.

These models suggest that treaty depth also has a substantively important effect on state adherence. Shifting the measure of verification strength from its shallowest to deepest value is associated with a 12 percentage point increase in the likelihood of a state joining the NPT, with all other variables held at their mean.\textsuperscript{32} For depth of enforcement, moving from the minimum to maximum number of member states threatened or targeted for sanctions is associated with a 15 percentage point increase in the propensity of states to join.\textsuperscript{33} Finally, the strongest track record of compliance is associated with a probability of adherence about 9 percentage points greater than the weakest, when all other variables are held at their mean.\textsuperscript{34}

Another way of examining the substantive importance of these empirical findings is by assessing their predictive validity; to what extent does considering a particular variable or factor improve our ability to make predictions outside of our data sample? Out-of-sample prediction also helps to combat model over-fitting—the risk that my results depend on idiosyncrasies in the particular data being used, rather than true relationships in the outside world (Beck, King, and Zeng 2000; Ward, Greenhill, and Bakke 2010).

To test the predictive validity of my results, I use a 3-fold cross-validation procedure of the kind commonly used in the computer science and machine learning literatures (Arlot and Celisse 2010). I divide my dataset—the same data used in the logistic regression models above—randomly into three parts.\textsuperscript{35} Two parts of the dataset are used as training data to construct the model, while the third part is used to test the predictive power of the model. The divided data is

\textsuperscript{32} Calculated using model 1; 95% CI: [1.9, 23.9]
\textsuperscript{33} Calculated using model 2; 95% CI: [0.7, 35.3]
\textsuperscript{34} Calculated using model 3; 95% CI: [0.5, 19.5]
\textsuperscript{35} 2-fold or 4-fold cross validation yield similar results.
then shuffled and the process repeated, so that each piece of the original dataset serves once as the test data. To avoid any bias introduced by the initial division of the data, I repeat this entire process 10 times using different random subsamples and average the results. The predictive models incorporating treaty depth are identical to those shown in Table 1, above, except I employ penalized likelihood logistic regression in lieu of standard logistic regression to avoid problems with quasi-separation of the data in smaller subsamples (Firth 1993; Heinze and Schepner 2002; Zorn 2005).

To measure the predictive power of these models, I use the area under the ROC curve (AUC) (Swets 1988). On one axis of the ROC curve is the rate of false positives—the number of cases in which the model incorrectly predicted state adherence, divided by the total number of abstentions. On the other axis is the true-positive rate—the number of correct predictions of adherence divided by the total number of cases of states joining the treaty. A perfect model, one that correctly predicts all cases, will have an AUC of 1.

Figure 5 shows the models' performance in out-of-sample prediction with bootstrapped 95 percent confidence intervals. The AUC for the three models that include variables representing treaty depth was very similar: 0.64 for verification, 0.66 for enforcement, and 0.65 for cooperation. The out-of-sample predictions of these models compare favorably to a model that does not account for the depth of the institution, but is otherwise identical (0.56). By way of comparison, a naïve model, comprised of only a cubic polynomial of time since the NPT opened for signature, has an AUC of 0.50. One way of thinking about the substantive significance of treaty depth, then, is that incorporating an understanding of depth into our analysis more than doubles the predictive power of all the other substantive factors considered in my model.
The empirical analysis up to this point largely assumes that treaty depth has the same impact on the adherence decisions of all states. My theory, however, predicts that the effect of depth will differ depending on the likelihood that the state will decide to cheat on its international commitments. Bad-faith states—those that join with the intention of violating the treaty—see depth as increasing the cost of membership, and so should be less likely to join the NPT when depth is high.

I use a two-stage approach to test the bad-faith hypothesis. First, I construct a model of state violations of the NPT, using a dataset restricted to treaty members. The dependent variable in this model is whether or not a member state has a nuclear weapons program in a given year; that is, whether or not the state is cheating on its NPT commitments. Explanatory variables
include the treaty’s track record, strength of verification measures, multilateral nuclear aid, conflict history, the presence of a weapons program in a rival state, GDP, and the presence of a defense pact with a weapons state, along with cubic polynomials to control for temporal dependence. This model correctly predicts whether a state violates the NPT in 99.5 percent of cases within the sample.\textsuperscript{36} I then apply this model to the full range of cases in my dataset of NPT adherence, capturing a predicted probability of cheating even for states that have not joined the treaty. In the second-stage of this modeling approach, I include the predicted probability of violation as an independent variable in models 1 through 3 above, interacting it with my three measures of treaty depth.

The results are illustrated in Figure 6. For each panel in the figure, the $y$ axis depicts the change in the predicted probability of state adherence associated with a shift from the shallowest to deepest level of a particular dimension of depth. The horizontal axis shows the range of predicted probabilities of violating the NPT. Each curve—surrounded by a 95-percent confidence interval—shows how the effect of depth on adherence is moderated or amplified by the state's likelihood of cheating.\textsuperscript{37}

\textsuperscript{36} The total in-sample accuracy of the model is somewhat deceptive as a measure of predictive power. Because NPT violation is an unusual event, merely guessing “no” in every case produces a very high level of accuracy. The model of NPT violation, however, also performs well by other metrics that are better suited to evaluating the prediction of rare events (Joshi 2002; Sun, Wong, and Kamel 2009). The model has a positive predictive value (or precision) of 0.979 and a true positive rate (or sensitivity) of 0.968, both measured within the sample.

\textsuperscript{37} Results are calculated holding all variables other than depth and the predicted probability of violation at their mean value in the data. To capture the additional uncertainty associated with the predicted probability of violation variable, which is itself a product of regression analysis, I use a cluster bootstrap procedure (resampling with replacement within each state) to calculate confidence intervals (Cameron, Gelbach, and Miller 2008).
The downward sloping pattern is similar for each dimension of depth, and the second difference of the interaction is statistically significant for each (Berry, DeMeritt, and Esaray 2010), lending support to the theory. For states that are unlikely to violate the treaty, depth is a major determinant of adherence. But for the minority of states that can reasonably anticipate the need to cheat, treaty depth either has little effect on the decision to join or actually discourages

---

38 This min-max second difference is calculated at the minimum and maximum of depth and the predicted probability of violation.
states from joining.\textsuperscript{39} Put another way, the breadth versus depth tradeoff does exist for those states that are likely to violate the central commitment of the NPT.

Conclusion

This article theorizes that the depth versus breadth tradeoff—widely advanced in the IO literature—does not apply to international security institutions like the NPT. Instead, I argue that greater depth actually makes security institutions more appealing to most prospective members by reducing the vulnerability of member states to the opportunistic behavior of those that refuse to comply. Deep treaties are more effective in holding members to their commitments, and so indicators of depth in an IO—stronger verification measures, effective enforcement, or a history of cooperation, for example—spur greater adherence by reassuring states that they will not be subject to a sucker's payoff. The theory makes the opposite prediction, however, with regard to the minority of states that join a treaty in bad faith, intending to violate their commitments. For these states, depth signals an increased cost of membership—because their noncompliance is more likely to be discovered and punished—and so bad-faith states are less likely to become members of the institution as its depth increases.

An empirical analysis of NPT membership supports this theory. Greater depth—measured across several dimensions—is both a statistically and substantively significant driver of state adherence. Further, incorporating the concept of depth into statistical models results in substantially better out-of-sample prediction of whether or not states will join the NPT. Models that interact measures of depth with a state's predicted probability of cheating reveal that the effect of depth is dependent on the likelihood of compliance. For most states, which enter the

\textsuperscript{39} States that are likely to comply substantially outnumber bad-faith states. Of the country years in my dataset, only about 1 in 6 are given a predicted probability of cheating of greater than 5 percent.
treaty intending to comply, greater depth is likely to encourage adherence. But for those states with a high probability of cheating, greater depth can have a negative effect on adherence—these states are less likely to join the treaty when depth is high.

Theorizing about the depth of international institutions helps us deal with two important limitations of the existing literature. First, prevailing explanations for treaty membership in the realm of international security—pressure from great powers, domestic politics, regime type—are largely static, and so have trouble explaining patterns of IO adherence over time. The depth of a treaty, however, often varies dramatically over the life the institution, with corresponding changes in the decisions of states to join or abstain.

Second, most theories of institutional membership and compliance do not take into account the behavior of the full population of member states. But collective restraint is the primary function of many international security institutions—without assurances that others are abiding by their commitments, states are unlikely to comply themselves. Our existing explanations for adherence, focusing as they do on state-level factors or the actions of a single influential state, miss this important part of the story. Reconsidering the depth versus breadth tradeoff, then, introduces a powerful driver of adherence that changes the way we understand the composition of international institutions.

References


