Online Appendix

Jane Vaynman*

November 2025

Supplementary materials for *Enemies in Agreement: Political Volatility and the Design of Arms Control*
Cambridge University Press, 2026

Contents

Part I: Game-Theoretic Model

1. Introduction A2
2. Model Setup A4
3. Analysis A8
4. Comparative Statics and Agreement Pathways A22

Part II: Statistical Analysis

1. Dataset Details B3
2. Explanatory Variables B5
3. Control Variables B9
4. Descriptive Statistics B14
5. Statistical Approach B18
6. Full Results: Hypothesis 3 . . . . . B19
7. Full Results: Hypotheses 1 and 2 . . . . . B19
8. Alternate Measures of Baseline Beliefs B26
9. Alternate Measures of Domestic Political Shifts B31
10. Socialization and Learning Results B34
11. Other Robustness Checks B34

Part III: List of Agreements

1. Adversarial Agreements Dataset: Names, Dates, and Information Provisions C1

*Assistant Professor, Johns Hopkins University School of Advanced International Studies (SAIS). Contact: jvaynman@jhu.edu. Replication materials available at https://doi.org/10.7910/DVN/EUFGGL or upon request from the author.

Online Appendix Part I
Game Theory for *Enemies in Agreement: Political Volatility and the Design of Arms Control*

Jane Vaynman

Contents

1 Introduction **A2**

2 Model Setup **A4**
2.1 Notation and Definitions A5
2.2 Stage Game A6
2.3 Monitoring Structure A7

3 Analysis **A8**
3.1 Proposition 1 . . . . . A8
3.2 Proposition 2 . . . . . A9
3.3 Proposition 3 . . . . . A18

4 Comparative Statics and Agreement Pathways **A22**

1 Introduction

This part of the online Appendix provides the full setup and evaluation of the treaty design game referenced in Chapter 2 of the book *Enemies in Agreement*. In the game, two states first select the level of monitoring in an agreement and then implement the agreement. The first stage is modeled as a take-it-or-leave-it game, and the implementation stage is modeled as a two-period Prisoner's Dilemma or Stag Hunt. The key comparative static of interest is the relationship between beliefs about the adversary's incentives for defection and whether an agreement with a certain monitoring level is possible in equilibrium, for fixed values of all other parameters.[1]

When states cooperate to restrain some military activity, they avoid spending resources on competition through arms racing or conflict. Investing in arms build-ups to maintain the balance of power, or even to improve it, can be avoided in principle through a deal that attains the same balance without the costs of military investments. This is similar to the premise that bargains can be struck to avoid costly conflict.[2] However, due to their opposing interests, adversaries face severe cooperation challenges.[3] In many cases, a state would be best off if its partner engaged in restraint, such as not building a new missile, while the state itself went ahead with the missile development and gained military superiority. States fear being on the losing end of these scenarios. By creating rules and commitments, formal agreements can allow states to maintain cooperation under the shadow of opponents that may have both the incentive and the capability to hide noncompliance.

For cooperation between states with incentives to cheat, provisions on information play a critical role in making agreements enforceable, but they are also a mixed blessing. On the one hand, monitoring provisions increase the probability that a violation will be detected and can facilitate timely responses, often with a reciprocal violation. On the other hand, greater information exchange, such as through inspections, also comes with increasing costs, including material costs of implementing monitoring as well as costs of potential security losses that arise when agreement monitoring also improves intelligence gathering.[4] Because of the potentially high costs of monitoring, states do not first choose to cooperate and then, conditional on that choice, select a level of information exchange. Rather, the monitoring design decision is an integral part of choosing any formal cooperation. States compare agreements with certain monitoring options directly with the non-cooperative status quo

[1] Throughout this analysis, the terms "defect" and "violate" are used synonymously, as are "cooperate" and "comply." Cooperate and defect follow standard game-theoretic terminology; comply and violate capture the substantive context of treaty implementation.

[2] James D. Fearon, "Rationalist Explanations for War," *International Organization* 49, no. 3 (Summer 1995): 379–414.

[3] Edward L. Miles et al., *Environmental Regime Effectiveness: Confronting Theory with Evidence* (Cambridge, MA: MIT Press, 2002).

[4] Andrew J. Coe and Jane Vaynman, "Why Arms Control Is So Rare," *American Political Science Review* 114, no. 2 (2020); Ronald B. Mitchell, "Sources of Transparency: Information Systems in International Regimes," *International Studies Quarterly* 42, no. 1 (1998): 109–130; Kenneth W. Abbott, "Trust but Verify: The Production of Information in Arms Control Treaties and Other International Agreements," *Cornell International Law Journal* 26 (1993): 1–58.

and could very well prefer ongoing competition over an agreement with a monitoring level that will be too costly to implement because it leaves them more vulnerable or involves high resource expenditure.

Beliefs about the adversary's desire to cheat would reasonably play a role in any cooperation calculus. In designing an agreement, states face a dilemma: They want to reap the security and resource-saving benefits from cooperation with enough information to catch attempts to sucker them. But, if attempts at cheating are unlikely, states would rather not bear the costs of information provisions to observe violations that are not going to happen. While pointing to beliefs about the adversary as an important factor is intuitive, it is not immediately clear how beliefs about the other state's incentives affect which information-sharing provisions (if any) states not only prefer individually, but will also agree to jointly in a deal where neither side might get its optimal terms.

The model in this appendix illuminates the role that beliefs about the other side's incentives, and changes in beliefs, play in whether an agreement is possible and the level of monitoring it includes.[5] The setup follows a number of studies that use a two-stage approach in which the signing of a treaty depends on expectations on how it will be implemented.[6]

The model extends existing research by combining incomplete information about player type, imperfect monitoring of moves, and bargaining over the treaty form.[7] Each of these extensions serves a distinct purpose for investigating not only cooperation vs. competition outcomes but also agreement design tradeoffs. First, incomplete information about types allows for an evaluation of beliefs and uncertainty.[8] Second, most work on compliance with international agreements assumes that a state's compliance and violation behavior is

---

[5] I am grateful to Igor Vaynman for extensive collaboration in designing and analyzing this model.

[6] James D. Fearon, "Bargaining, Enforcement, and International Cooperation," *International Organization* 52, no. 2 (Spring 1998): 269–306; George W. Downs, David M. Rocke, and Peter N. Barsoom, "Is the Good News about Compliance Good News about Cooperation?," *International Organization* 50, no. 3 (Summer 1996): 379–406; Brett Ashley Leeds, "Domestic Political Institutions, Credible Commitments, and International Cooperation," *American Journal of Political Science* 43, no. 4 (October 1999): 979–1001.

[7] The setup builds most directly on Fearon, "Bargaining, Enforcement, and International Cooperation." Fearon's model is characterized by 1) bargaining over the distribution of benefits, 2) violations are perfectly observable, 3) imperfect information about other player's payoffs for the status quo (the "defect, defect" box).

[8] Here the model is most similar to Andrew Kydd's treatment of types in a number of his models. The outcome discussed below as the Full Visibility Threshold, is essentially the same as the minimum level of trust for cooperation without additional reassurance Kydd identifies in Andrew Kydd, "Trust, Reassurance, and Cooperation," *International Organization* 54, no. 2 (2000): 325–357, which is as expected given the setup. Uncertainty about preferences is conceptualized similarly to Andrew Kydd, "Trust Building, Trust Breaking: The Dilemma of NATO Enlargement," *International Organization* 55, no. 4 (2001): 801–828. Scott Wolford, "Neutrality Regimes," *Security Studies* 28, no. 4 (2019): 807–832 also treats player type as defined in terms of different payoffs for future behavior, in that case, attacking a neutral state. In addition, though it does not affect the formal model itself, it is important to note a substantive point in my use of "types." In my model, type is about incentives with specific opponents, not an immutable characteristic of the state itself. A state may see high incentives for cooperation with one potential partner and high incentives to violate a deal with another.

perfectly observable.[9] I relax this assumption in order to explicitly model the most pressing risk in arms control—failure to detect violations. Finally, the model assumes that the distribution of benefits from cooperation is fixed at evenly balanced, while the monitoring level is the variable parameter that actors must agree on. This change focuses attention on preferences over information provisions as an agreement design feature while holding constant distributional concerns that we would expect to matter in addition. As a result, the model can show that a constrained range for agreements exists even when states are not concerned about being locked into a disadvantageous distribution of benefits.[10]

2 Model Setup

There are two Players. At the start of the game, Nature assigns one of two Types to each player and these are observed as a private signal. Players can be either a Competitive type denoted as H or Cooperative type denoted as L. Players know their own type and have a belief about the other's type, where α represents the probability that the opponent is a Competitor (H). Beliefs are symmetric.[11]

Types are defined by their different payoffs for unilateral defection in the Implementation game, which is a modified Prisoner's Dilemma. For player type H, the payoff from unilateral defection t is the highest possible payoff, or the "temptation payoff." For player type L, the unilateral defection payoff g is less than the payoff from mutual compliance r. Both types receive $-l$ for unilateral compliance (sucker's payoff) and $-c$ for mutual defection (status quo). Assume that the relationship between payoffs is $t > r > g > -c > -l$. The payoff structure for the stage game for each implementation period is therefore either a classical Stag Hunt (two L types), Prisoner's Dilemma (two H types), or non-standard asymmetric game (one of each).

The order of moves in the game is as follows:

(1) At the start of the game, Nature assigns the player types. Each player knows its own type and has a belief about the type of the adversary. (2) States bargain over a monitoring level for an agreement. One state makes a take-it-or-leave-it offer of a monitoring level between 0 and 1. (3) The offer is either accepted or rejected. (4) If rejected, the

[9] For both this implicit assumption and a review of compliance studies see: James D. Morrow, "When Do States Follow the Laws of War?," *American Political Science Review* 101, no. 3 (2007): 559–572.

[10] Like related work, my approach incorporates expectations about the near future but does not delve into questions on time horizons or future interactions. Variation in how long each state expects an agreement to be implemented, or whether states expect one agreement to affect the prospects for cooperation on other issues, could be a factor in current negotiations. However, these are also likely second-order concerns after addressing the imminent cheating risks for the current agreement. The model here also differs from treatments that focus on commitment problems and uncertainty about the future conditions, both in conflict and institutional design contexts. Commitment problem games are often modeled with complete information and either uncertainty about some future condition, such as the distribution of power or the size of benefits, or an element of randomness about the future values of these variables.

[11] Similar results would follow if this assumption were relaxed, but making it simplifies the analysis and illustration of the results by eliminating the need to consider asymmetric levels of monitoring.

game ends with status quo payoffs for two periods. If accepted, the game moves on to the implementation stage.

The implementation stage is two rounds. At the start, (5) states privately and simultaneously decide to comply or violate the agreement.[12] (6) A public signal is then generated according to the monitoring structure, and the level of monitoring chosen in bargaining. The signal $s \in \{\bar{s}, \underline{s}\}$ is generated after the first period and observed by both players. The signal $\underline{s}$ represents an observed defection. The distribution of the signals conditional on the actions in period 1 is: $P(\bar{s}|a) = 1$ if $a = CC$ and $(1-p)$ if $a \in \{CD, DC, DD\}$; $P(\underline{s}|a) = 0$ if $a = CC$ and p if $a \in \{CD, DC, DD\}$. Players do not observe the outcome in step 5 separately from the signal. (7) After observing the signal, states again simultaneously decide to comply or violate the agreement. (8) The game ends and states get their payoffs. Final payoffs depend on the payoffs in each round, minus monitoring costs.

The analysis starts with the implementation game. The equilibria conditions identified in the implementation game will then affect what states do in the prior bargaining step. In bargaining, states select a monitoring level and pay a cost for that monitoring. In implementation, states comply or defect at a fixed monitoring level.

2.1 Notation and Definitions

Formally, let $\Theta_i = \{L, H\}$ denote the types of player i, with generic element θ_i. We can interpret θ_i as a private signal for player i observed at the start of the game. Let $A_i = \{C, D\}$ denote the set of pure actions available to player i in the stage game, and a_i an element of the set. The stage game payoff for player i is the function $u_i : \Theta_i \times A_i \times A_{-i} \to \mathbb{R}$, and written as $u^{\theta_i}(a_i, a_{-i})$.

Let $S = \{\bar{s}, \underline{s}\}$ denote the public signal space, with generic element s. Players receive no information on the opponent's play other than the observed signal s, and only receive payoffs at the end of the final stage game.

A history for player i in period $t = 1$ is $h_i^1 = \theta_i$, with the set of histories denoted by $\mathscr{H}_i^1 = \Theta_i$. A history for player i in period $t = 2$ is $h_i^2 = (\theta_i, a_i, s)$, with the set of histories denoted by $\mathscr{H}_i^2 = \Theta_i \times A_i \times S$. The set of all histories for player i is $\mathscr{H} = \mathscr{H}_i^1 \cup \mathscr{H}_i^2$.

A pure strategy for player i is the mapping from possible histories into pure actions: $\sigma_i : \mathscr{H}_i \to A_i$. We say a pure strategy is "action-free" if it is invariant over player i's action in period $t = 1$: $\sigma_i(\theta_i, a_i, s) = \sigma_i(\theta_i, \tilde{a}_i, s)$, $\forall a_i, \tilde{a}_i \in A_i$. For notational simplicity, for action free pure strategies we write $\sigma_i(\theta_i, a_i, s)$ as $\sigma_i^{\theta_i}(s)$ since the action a_i is redundant. Similarly, for period $t = 1$, $\sigma_i(\theta_i)$ is simply written as $\sigma_i^{\theta_i}$. For example, $\sigma_1^L(\bar{s})$ is the period $t = 2$ action for player 1, type L, after observing signal $\bar{s}$. And σ_2^H is the period 1 action of player 2, type H.

[12] Note that both types of players can choose either action, i.e., a Cooperator L can play violate and a Competitor H can choose to comply.

A belief system assigns a probability of the opponent being a Competitive type at each history in the second period. Specifically, $\mu : \mathscr{H}_2 \to [0,1]$, where $\mu_i^{\theta_i}(a_i, s) = P(\theta_{-i} = H|\theta_i, a_i, s)$. For clarity we will drop the subscript i and simply write $\mu^{\theta}(a, s)$. Intuitively, α represents the initial belief that the opponent is a Competitor, which may or may not be updated in the second period.

Let p represent the probability of observing a defection if one has occurred, in accordance with the monitoring structure discussed below. The cost of monitoring is given by the function $c(p)$, which is increasing in p. The cost of monitoring is incurred in the Bargaining Game, and p is set at the start of the Implementation Game.

2.2 Stage Game

The stage game is either a classical Stag Hunt or Prisoner's Dilemma game, depending on the type of the players, summarized in Figure 1.

Figure 1: Stag Hunt: Cooperative Type vs. Cooperative Type

		B (Type L) C	B (Type L) D
A (Type L)	C	r, r	$-l, g$
	D	$g, -l$	$-c, -c$

Figure 2: Prisoner's Dilemma: Competitive Type vs. Competitive Type

		B (Type H) C	B (Type H) D
A (Type H)	C	r, r	$-l, t$
	D	$t, -l$	$-c, -c$

Figure 3: Mixed Defection Prisoner's Dilemma: Cooperative Type vs. Competitive Type

		B (Type H) C	B (Type H) D
A (Type L)	C	r, r	$-l, t$
	D	$g, -l$	$-c, -c$

The two player types receive the same payoffs except in the case of unilateral defection: $(a_i, a_{-i}) = (D, C)$. For the Cooperative type (L), the payoff for unilateral defection

is less than the payoff from full cooperation. However, for the Competitive type (H), the payoff from unilateral defection is the highest possible payoff, and hence we refer to it as the "temptation payoff," reflecting its usual interpretation in the setting of the Prisoner's Dilemma.

The full set of stage game payoff parameters are:

r	benefit from mutual cooperation
g	benefit to the Cooperative type from unilateral defection
t	benefit to the Competitive type from unilateral defection (temptation payoff)
$-l$	cost of unilateral compliance (sucker's payoff)
$-c$	cost of mutual defection (status quo)

Assume that the relationship between payoffs is, $t > r > g > -c > -l$. When both types are Cooperative, then the stage game is a classical Stag Hunt game since $r > g$. When both types are Competitive, then the stage game is a Prisoner's Dilemma since $t > r$. When one player is Competitive and the other is Cooperative, then we are in a non-standard asymmetric game. *Notation note:* Payoffs $-l$ and $-c$ represent losses, while l and c denote the positive costs used in algebraic expressions (e.g., $l - c$ is the difference between these costs).

2.3 Monitoring Structure

In the first period, players choose to comply or defect. The action is not observed publicly and a signal $s \in \{\bar{s}, \underline{s}\}$ is generated and observed by both players. In this monitoring structure, the signal $\underline{s}$ represents an observed defection.

The distribution of the signals conditional on the actions in period 1 is,

$$P(\bar{s}|a) = \begin{cases} 1, & \text{if } a = CC \\ 1-p, & \text{if } a \in \{CD, DC, DD\} \end{cases} \quad (1)$$

$$P(\underline{s}|a) = \begin{cases} 0, & \text{if } a = CC \\ p, & \text{if } a \in \{CD, DC, DD\} \end{cases} \quad (2)$$

In this monitoring structure, the signal $\underline{s}$ represents an observed defection, and there are no false positives of defection: $P(CC|\underline{s}) = 0$. Therefore if the signal $\underline{s}$ is observed, it is known with certainty that at least one player defected.[13] There are false negatives, however.

[13] The lack of false positives is an important simplification to make the model tractable and has the benefit of focusing substantive attention on the most important threat - an intentional violation. However, in real-world arms control cases, states are also often interested in avoiding mistaken information that there has been a violation. Such situations could exacerbate tensions between adversaries who start to blame one another for abrogating the agreement. See for example Schelling (1985) on false positives and false

If there is no defect signal, $\underline{s}$, that does not mean that no one cheated; it only means that no player was observed cheating. In this monitoring structure, the probability of observing a defection signal is the same if one player or both players defect.

3 Analysis

The key results are presented first in plain language as Propositions 1, 2, and 3. Each is then restated formally, followed by corresponding theorems and proofs. This section also includes some additional material that supports the propositions.

3.1 Proposition 1

If both sides are confident enough that the other is a Cooperator, then cooperation is possible without any monitoring.

If states are sufficiently sure that the adversary has low incentives to cheat, then they know they are playing the Stag Hunt game in both rounds with benefits from mutual cooperation for both states. Spending on monitoring is unnecessary because they expect one another to comply regardless. States do not have to be completely certain that the adversary is a Cooperator, they just need to be certain enough that the risks of cheating are small, with beliefs (α) falling below a "no visibility threshold."

This first key result comes from an analysis of the stage game as a one-shot game, solving for the pure strategy equilibria. In any equilibrium of the one-shot stage game, the Competitive type must be playing D. The Cooperative type can play C in equilibrium if there is a sufficiently small likelihood that the opponent is a Competitive type. The belief threshold for cooperation is determined by the gain from cooperation relative to unilateral defection ($r - g$) and the mitigation of loss due to the status quo relative to unilateral compliance ($l - c$). We refer to this belief threshold as the No Visibility Threshold because it defines the sufficient level of confidence for cooperation to occur without any ability to observe the opponent's compliance or defection behavior.

3.1.1 No Visibility Threshold

Theorem 1 (No Visibility Threshold)
Let β denote the symmetric belief that opponent is the Competitive type: $\beta = P(\theta_{-i} = H)$, and define

$$\mu^* = \frac{r - g}{r - g + l - c} \tag{3}$$

Then,

negatives in information provisions. A model can be constructed to account for this idea of "mistakes" in observation, but this would require a considerably more complex monitoring structure and would be best explored in future research. Thomas C. Schelling and Morton H. Halperin, *Strategy and Arms Control* (Washington, DC: Pergamon-Brassey, 1985), 9.

1. *The action profile (D, D), is an equilibrium for any belief $\beta \in [0, 1]$.*

2. *There exists another pure strategy equilibrium in which Cooperative types play C while Competitive types play D if and only if $\beta \leq \mu^*$.*

Proof: That (D, D) is an equilibrium is trivial, so we focus on the second equilibrium. Since $a_i^H = D$ is always a best response, it is sufficient to show $a_i^L = C$ is a best response to actions $a_{-i}^L = C$ and $a_{-i}^H = D$. The expected payoff from $a_i^L = C$ is,

$$\begin{aligned} E[u_i(C, a_{-i}, L)] &= u_i(C, D, L)P(\theta_{-i} = H) + u_i(C, C, L)P(\theta_{-i} = L) \\ &= -l\beta + r(1 - \beta) \end{aligned} \tag{4}$$

The expected payoff from $a_i^L = D$ is,

$$\begin{aligned} E[u_i(D, a_{-i}, L)] &= u_i(D, D, L)P(\theta_{-i} = H) + u_i(D, C, L)P(\theta_{-i} = L) \\ &= -c\beta + g(1 - \beta) \end{aligned} \tag{5}$$

Therefore $a_i^L = C$ is a best response if,

$$E[u_i(C, a_{-i}, L)] \geq E[u_i(D, a_{-i}, L)] \quad \Rightarrow \quad \beta \leq \frac{r - g}{r - g + l - c} = \mu^\star \tag{6}$$

Conversely, if $\beta > \mu^\star$ then $a_i^L = C$ is not a best response, and hence the only equilibrium is mutual defection, (D, D).

It is useful to consider the threshold $\mu^* = (r - g)/(r - g + l - c)$ in a bit more detail. We can see that μ^* decreases as $r - g$ decreases or as $l - c$ increases. The difference $r - g$ represents the additional benefit to the Cooperative type from coordinating on (C, C). When the benefits of cooperation are higher, then the players will be willing to cooperate even while having a higher initial belief that the opponent is a Competitive type. The difference $l - c$ measures the costs of being taken advantage of versus the status quo. As this cost increases, the Cooperative type needs to be more sure that she is not playing an opponent Competitive type that will take advantage of her, and so μ^* threshold decreases. In other words, when costs of being suckered are high, then the Cooperative type will need to be fairly sure she is playing another Cooperative type. Both these relationships – the relative benefits of cooperation and the relative costs of being taken advantage of – should sound fairly intuitive as being important factors from the substantive perspective of arms control treaties.

We refer to the (D, D) equilibrium as mutual defection. The equilibrium in which Cooperative types play C is the cooperative equilibrium. Since $r - g > 0$ and $l - c > 0$, then $\mu^* \in (0, 1)$, and hence for any set of payoff parameters there will be a non-empty region $[0, \mu^*]$ in which the cooperative equilibrium exists.

3.2 Proposition 2

If each side is confident enough that the other is a Competitor, then there is no agreement even if monitoring is perfect.

Competitive types will always defect in the second round of the game and would prefer to defect undetected in the first round as well. When states are sure that this is the likely scenario, they do not want to set themselves up to be duped, so they avoid cooperation in the first place. At a sufficiently high level of confidence, with beliefs (α) falling above a "full visibility threshold," even perfect monitoring would not resolve the dilemma—monitoring would only reveal what states already believe is likely, that defection occurred.[14]

Theorem 8 below formalizes this result, so readers can skip ahead to that section if desired. The result for Proposition 2 and in part Proposition 3 is based on analysis of the full Treaty Implementation Game, so we discuss this first.

3.2.1 Treaty Implementation Game

We consider strategy profiles to evaluate the conditions under which compliance with an agreement can happen in equilibrium. The equilibrium concept used is the Perfect Bayesian Nash Equilibrium (PBE). A profile of strategies is a PBE when (i) players are sequentially rational given a system of beliefs and (ii) beliefs are updated correctly via Bayes Rule when possible.

Since the overall Treaty Implementation Game consists of two periods, the final period is simply a one-shot stage game. Theorem 1 therefore shows that if the belief at the start of the second period that the opponent is a Competitive type is below $\mu^\star$, then the cooperative equilibrium is possible in the final period. For a Competitive type, it is always rational to violate in the second round regardless of the signal observed from monitoring.

Lemma 2
A strategy profile $\sigma = (\sigma_1, \sigma_2)$ is sequentially rational if and only if $\sigma_i^H(s) = D$.

Proof: This result is trivial since D strictly dominates C for the Competitive types in the final period.

We can therefore restrict attention to action free pure strategies where $\sigma_i^H(s) = D$. This restriction leaves only the following mappings to be defined for any strategy: σ_i^H, σ_i^L, $\sigma_i^L(\bar{s})$, and $\sigma_i^L(\underline{s})$. We can describe any strategy compactly as a 4-tuple,

$$\sigma_i = \begin{pmatrix} H \text{ type action in period 1,} \\ L \text{ type action in period 1,} \\ L \text{ type action in period 2 after observing } \bar{s} \text{ signal,} \\ L \text{ type action in period 2 after observing } \underline{s} \text{ signal} \end{pmatrix} \tag{7}$$

For example, the strategy $\sigma_i = (D, C, C, D)$ means that type H plays D in period 1, type L plays C in period 1, C in period 2 after observing a $\bar{s}$ signal, and D in period 2 after observing a $\underline{s}$ signal. Of course, type H plays D in period 2 regardless of the signal.

[14] See Appendix Theorem 2 on the Full Visibility Threshold, which defines the range of beliefs (α) beyond which agreements are not possible.

Further restricting attention to symmetric strategy profiles where $\sigma_1 = \sigma_2$, we end up with 16 unique strategy profiles to consider.

Lemma 3 and **Lemma 5** below identify two strategies that are PBE under a set of conditions. Then, **Theorem 8** considers both strategies in formalizing the logic of Proposition 2. We refer to this as the **Full Visibility Threshold** because it represents a level of confidence beyond which even a perfect ability to observe the other side's actions will not support cooperation, and the only equilibrium strategy is mutual defection.

3.2.2 Strategy 1

Under the **Repeated Violation** strategy, $\sigma_i = (D, C, C, D)$, the Competitive type plays violate in the first round, Cooperative type complies, and in the second round, the Cooperative type complies if the comply signal was observed and violates if a defect signal was observed. This is an equilibrium strategy for implementing an agreement and is conditional on initial beliefs and a monitoring level that is sufficiently informative to allow states to update that belief.

Lemma 3 (Strategy 1: Repeated Violation)
The symmetric strategy profile $\sigma_i = (D, C, C, D)$ is a PBE under the belief system $\mu^{L,H}(\emptyset) = \alpha$, $\mu^L(\bar{s}) = (1-p)\alpha/(1-p\alpha)$, $\mu^L(\underline{s}) = 1$, and $\mu^H(\bar{s}) = \mu^H(\underline{s}) = \alpha$ if, the monitoring level p and initial belief α satisfy,

$$\left(\frac{\alpha}{1-\alpha}\right)\left(\frac{l-c}{t+c}\right) + \left(\frac{t-r}{t+c}\right) \geq p \geq \max\left\{0, \frac{\alpha-\mu^*}{\alpha-\alpha\mu^*}, \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{l-c}{r+c}\right) - \left(\frac{r-g}{r+c}\right)\right\}$$

Proof: We begin by showing that beliefs are consistent with Bayes Rule when possible. Both signals $\bar{s}$ and $\underline{s}$ occur on the equilibrium path, and there are well-defined updated beliefs given each signal. Based on the monitoring structure, the probability of observing a defection when the opponent is the Competitive type and plays D in the first round is $P(\underline{s}|\theta_{-i} = H, \sigma) = p$, and the probability of not observing the defection is,

$$P(\bar{s}|\theta_{-i} = H, a_i = C, a_{-i}) = P(\bar{s}|CD) = 1 - p$$

If the opponent is a Cooperative type, then lack of false positives in the monitoring implies,

$$P(\bar{s}|\theta_{-i} = L, a_i = C, a_{-i}) = P(\bar{s}|CC) = 1$$

Then by Bayes Rule,

$$\begin{aligned}
P(\theta_{-i} = H|\bar{s}, a_i = C, a_i) &= \frac{P(\bar{s}|CD)P(\theta_{-i} = H)}{P(\bar{s}|CD)P(\theta_{-i} = H) + P(\bar{s}|CC)P(\theta_{-i} = L)} && (8)\\
&= \frac{(1-p)\alpha}{(1-p)\alpha + (1-\alpha)} && (9)\\
&= \frac{(1-p)\alpha}{1-p\alpha} && (10)\\
&= \mu^L(C, \bar{s}) && (11)
\end{aligned}$$

To show $\mu^L(C, \underline{s}) = 1$, note that if a defection signal $\underline{s}$ is observed, then the opponent must be a Competitive type since there are no false positives in the monitoring structure:

$$\begin{aligned} P(\theta_{-i} = H|\underline{s}, a_i = C, a_i) &= \frac{P(\underline{s}|CD)P(\theta_{-i} = H)}{P(\underline{s}|CD)P(\theta_{-i} = H) + P(\underline{s}|CC)P(\theta_{-i} = L)} \\ &= \frac{p\alpha}{p\alpha + 0 \times (1-\alpha)} \\ &= 1 \end{aligned}$$

For the Competitive type, the belief entering the second period has no implication on the equilibrium since the dominant action is D, but for completeness, we verify that the beliefs are consistent with Bayes Rule. Given that the Competitive type plays D the first period, the distribution of signals is the same for a Competitive and Cooperative opponent playing the equilibrium strategy in period 1:

$$\begin{aligned} P(\bar{s}|\theta_{-i} = H, a_i = D, a_{-i}) = P(\bar{s}|DD) &= 1 - p \\ P(\bar{s}|\theta_{-i} = L, a_i = D, a_{-i}) = P(\bar{s}|DC) &= 1 - p \end{aligned}$$

Then by Bayes Rule,

$$\begin{aligned} P(\theta_{-i} = H|\bar{s}, a_i = D, a_i) &= \frac{P(\bar{s}|DD)P(\theta_{-i} = H)}{P(\bar{s}|DD)P(\theta_{-i} = H) + P(\bar{s}|DC)P(\theta_{-i} = L)} \\ &= \frac{(1-p)\alpha}{(1-p)\alpha + (1-p)(1-\alpha)} \\ &= \alpha \end{aligned}$$

We can get an identical result for $s = \underline{s}$. Hence the beliefs of the Competitive type playing D in the first period after observing any signal are, $\mu^H(D, \bar{s}) = \mu^H(D, \underline{s}) = \alpha$

The information sets off the equilibrium path are, $(\theta_i = L, a_i = D, s)$ and $(\theta_i = H, a_i = C, s)$. For these information sets, the beliefs are not pinned down by Bayes Rule, and we simply set the beliefs in the second period to be $P(\theta_{-i} = H|s) = 1$ and all players play D.

We next verify sequential rationality in the second period. Lemma 2 trivially assures sequential rationality at information sets $(\theta_i = H, a_i, s)$. Now consider information sets $(\theta_i = L, a_i, s)$. Sequential rationality after observing signal $\bar{s}$ occurs if the belief that the opponent is the Competitive type is less than or equal to μ^*:

$$\mu^L(C, \bar{s}) = \frac{(1-p)\alpha}{1-p\alpha} \leq \mu^* \quad \Longrightarrow \quad p \geq \frac{\alpha - \mu^*}{\alpha - \alpha\mu^*}$$

After observing the defection signal $\underline{s}$, the opponent is revealed to be Competitive $\mu^L(C, \underline{s}) = 1$. In this case, D is clearly the best response.

The term $p^* = (\alpha - \mu^*)/(\alpha - \alpha\mu^*)$ therefore represents the minimum monitoring necessary for this strategy profile to be sequentially rational. As expected, if $\alpha \leq \mu^*$, then

$p = 0$ satisfies the condition: the Cooperative player is already sufficiently sure that the other player is not Competitive so he is willing to play C in period 2 even without updating and monitoring is not necessary. On the other hand, if $1 > \alpha > \mu^*$, then $p^* \in (0, 1)$, and so a necessary condition for this strategy profile to be sequentially rational is that there is a minimum level $p^* > 0$.

Next, we verify sequential rationality in the first period. Begin with the Cooperative type. The payoff from the equilibrium strategy σ is,

$$\begin{aligned} E[U_i^L(\sigma,\sigma)] &= (1-\alpha)2r + \alpha(-l + p(-c) + (1-p)(-l)) && (12)\\ &= (1-\alpha)2r - 2\alpha l + p\alpha(l-c) && (13)\end{aligned}$$

Let $\tilde{\sigma}$ be the strategy where the Cooperative type plays D in the first period instead of C, and then follows σ in the remaining information sets. The expected full game payoff from $\tilde{\sigma}$ is,

$$\begin{aligned} E[U_i^L(\tilde{\sigma},\sigma)] &= (1-\alpha)(g + p(-c) + (1-p)r) + \alpha(-c + p(-c) + (1-p)(-l))\\ &= p\left[(1-\alpha)(-c-r) + \alpha(l-c)\right] + \alpha(-c-l) + (1-\alpha)(g+r)\end{aligned}$$

Sequential rationality for Cooperative type is satisfied if,

$$\begin{aligned} E[U_i^L(\sigma,\sigma)] &\geq E[U_i^L(\tilde{\sigma},\sigma)]\\ (1-\alpha)2r - 2\alpha l + p\alpha(l-c) &\geq p\left[(1-\alpha)(-c-r) + \alpha(l-c)\right] + \alpha(-c-l) + (1-\alpha)(g+r)\\ p(1-\alpha)(r+c) &\geq \alpha(l-c) - (1-\alpha)(r-g)\\ p &\geq \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{l-c}{r+c}\right) - \left(\frac{r-g}{r+c}\right)\end{aligned}$$

Define the minimum level of monitoring,

$$p^\dagger = \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{l-c}{r+c}\right) - \left(\frac{r-g}{r+c}\right) \quad (14)$$

Finally, we check that playing D in the first period is the best strategy for the Competitive type. The Competitive type's payoff from the equilibrium strategy is,

$$\begin{aligned} E[U_i^H(\sigma,\sigma)] &= (1-\alpha)(t + p(-c) + (1-p)t) - 2\alpha c && (15)\\ &= p(1-\alpha)(-c-t) + (1-\alpha)2t - 2\alpha c && (16)\end{aligned}$$

Let $\hat{\sigma}$ be the strategy where the Competitive type plays C in the first period instead of D, and then follows σ in the remaining information sets.

$$E[U_i^H(\hat{\sigma},\sigma)] = (1-\alpha)(r+t) + \alpha(-l-c)$$

Sequential rationality for Competitive type is satisfied if,

$$\begin{aligned} E[U_i^H(\sigma,\sigma)] &\geq E[U_i^H(\hat{\sigma},\sigma)]\\ p(1-\alpha)(-c-t) + (1-\alpha)2t - 2\alpha c &\geq (1-\alpha)(r+t) + \alpha(-l-c)\\ (1-\alpha)(t-r) + \alpha(l-c) &\geq p(1-\alpha)(t+c)\\ \left(\frac{t-r}{t+c}\right) + \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{l-c}{t+c}\right) &\geq p\end{aligned}$$

Lemma 4 provides a condition that determines which lower bound, $p^\dagger$ or p^*, is binding for the strategy to be a PBE.

Lemma 4 (Strategy 1 minimum monitoring)
If $\alpha > \frac{r+c}{r+l}$ *then*

(a) $p^\dagger > p^*$

(b) The upper bound on monitoring is not binding:

$$\left(\frac{t-r}{t+c}\right) + \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{l-c}{t+c}\right) > 1$$

Proof: Recall that,

$$\mu^* = \frac{r-g}{r-g+l-c} \quad \Rightarrow \quad \frac{l-c}{r-g} = \frac{1-\mu^*}{\mu^*}$$

Rearranging the expression for $p^\dagger$,

$$p^\dagger = \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{1-\mu^*}{\mu^*}\right)\left(\frac{r-g}{r+c}\right) - \left(\frac{r-g}{r+c}\right) = \left(\frac{r-g}{r+c}\right)\left[\left(\frac{\alpha}{1-\alpha}\right)\left(\frac{1-\mu^*}{\mu^*}\right) - 1\right]$$

Noting that,

$$\begin{aligned}\left(\frac{\alpha}{1-\alpha}\right)\left(\frac{1-\mu^*}{\mu^*}\right) - 1 &= \frac{\alpha-\mu^*}{\mu^*-\mu^*\alpha} \\ &= \left(\frac{\alpha-\mu^*}{\alpha-\alpha\mu^*}\right)\left(\frac{\alpha-\alpha\mu^*}{\mu^*-\mu^*\alpha}\right) \\ &= p^*\left(\frac{\alpha}{1-\alpha}\right)\left(\frac{1-\mu^*}{\mu^*}\right)\end{aligned}$$

So we can write,

$$p^\dagger = p^*\left(\frac{r-g}{r+c}\right)\left(\frac{\alpha}{1-\alpha}\right)\left(\frac{1-\mu^*}{\mu^*}\right)$$

Plugging in $\frac{1-\mu^*}{\mu^*} = \frac{l-c}{r-g}$ yields,

$$p^\dagger = p^*\left(\frac{l-c}{r+c}\right)\left(\frac{\alpha}{1-\alpha}\right)$$

Hence we get that, $p^\dagger > p^*$ if,

$$\left(\frac{l-c}{r+c}\right)\left(\frac{\alpha}{1-\alpha}\right) > 1 \quad \Rightarrow \quad \alpha > \frac{r+c}{r+l}$$

To show (b), we just plug in $\alpha = \frac{r+c}{r+l}$ to get,

$$\begin{aligned}\left(\frac{t-r}{t+c}\right)+\left(\frac{\alpha}{1-\alpha}\right)\left(\frac{l-c}{t+c}\right) &= \left(\frac{t-r}{t+c}\right)+\left(\frac{r+c}{r+l-r-c}\right)\left(\frac{l-c}{t+c}\right)\\ &= \left(\frac{t-r}{t+c}\right)+\left(\frac{r+c}{t+c}\right)\\ &= 1\end{aligned}$$

And then observe that the upper bound is increasing in α.

3.2.3 Strategy 2

Under the **Temporary Compliance** strategy, $\sigma_i = (C, C, C, D)$, the Competitive type complies in the first round, the Cooperative type complies, and in the second round the Cooperative type complies if a comply signal was observed and violates if a defect signal was observed.

Lemma 5 (Strategy 2: Temporary Compliance)
The symmetric strategy profile $\sigma_i = (C, C, C, D)$ is a PBE under the belief system $\mu_i^{L,H}(\emptyset) = \alpha$, $\mu_i^{L,H}(\bar{s}) = \alpha$, $\mu_i^{L,H}(\underline{s}) = 1$ if the monitoring level p and initial belief α satisfy,

$$\alpha \leq \min\left\{\frac{r+c}{t+c}, \mu^*\right\}, \quad \textit{and} \quad p \geq \frac{t-r}{(1-\alpha)(t+c)} \tag{17}$$

Proof: Begin by verifying that beliefs are consistent with Bayes Rule when possible. As both types play C in the first period, no updating is possible as types pool, and hence beliefs don't change and stay at $P(\theta_{-i}|\bar{s}, a_i, a_{-i}) = \alpha$. There are no false positives in the monitoring structure, and since on the equilibrium path the $\underline{s}$ signal is not observed, we are free to set beliefs $\mu_i^{L,H}(\underline{s}) = 1$ so that D is sequentially rational in period 2 after a $\underline{s}$ signal. Sequential rationality in period 2 after signal $\bar{s}$ is satisfied for the Cooperative type since $\alpha \leq \mu^*$, and trivially for the Competitive type playing D.

Next, we verify sequential rationality in the first period. Since the equilibrium strategy specifies both opponent types playing C in the first period, then it is clearly sequentially rational for the Cooperative type to play C in the first period. We focus on the Competitive type. The payoff from the equilibrium strategy σ for the Competitive type is,

$$E[U_i^H(\sigma, \sigma)] = (1-\alpha)(r+t) + \alpha(r-c)$$

Let $\hat{\sigma}$ be the strategy where the Competitive type plays D in the first period instead of C, and then follows σ in the remaining information sets. The expected full game payoff from $\hat{\sigma}$ is,

$$\begin{aligned}E[U_i^H(\hat{\sigma}, \sigma)] &= (1-\alpha)(t + p(-c) + (1-p)t) + \alpha(t-c)\\ &= p(1-\alpha)(-t-c) + (1-\alpha)2t + \alpha(t-c)\end{aligned}$$

Sequential rationality for Competitive type is satisfied if,

$$\begin{aligned} E[U_i^H(\sigma,\sigma)] &\geq E[U_i^H(\hat{\sigma},\sigma)] \\ (1-\alpha)(r+t)+\alpha(r-c) &\geq p(1-\alpha)(-t-c)+(1-\alpha)2t+\alpha(t-c) \\ p(1-\alpha)(t+c) &\geq t-r \\ p &\geq \frac{t-r}{(1-\alpha)(t+c)} \end{aligned}$$

A $p \in [0,1]$ exists satisfying this relation if,

$$\frac{t-r}{(1-\alpha)(t+c)} \leq 1 \quad \Rightarrow \quad \alpha \leq \frac{r+c}{t+c} \tag{18}$$

3.2.4 Equilibrium Conditions

Lemma 6 defines the criteria for Strategy 1 or Strategy 2 to be an equilibrium below the No Visibility Threshold. Lemma 7 briefly addresses uninformative strategies which can also be an equilibrium below the No Visibility Threshold.

Lemma 6 (Criteria for Strategies to be EQM)
Suppose $\alpha \leq \mu^*$ and $l-c \geq t-r$. Then for any p, Temporary Compliance, Repeated Violation, or both must be equilibrium.

Proof: Since $\alpha \leq \mu^*$, then

$$\frac{\alpha}{1-\alpha} \leq \frac{\mu^*}{1-\mu^*} = \frac{r-g}{l-c}$$

and,

$$\left(\frac{\alpha}{1-\alpha}\right)\left(\frac{l-c}{r+c}\right) - \left(\frac{r-g}{r+c}\right) \leq \left(\frac{r-g}{l-c}\right)\left(\frac{l-c}{r+c}\right) - \left(\frac{r-g}{r+c}\right) = 0$$

Since $p \in [0,1]$ then $p \geq \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{l-c}{r+c}\right) - \left(\frac{r-g}{r+c}\right)$. Further, $p \geq 0 \geq \frac{\alpha-\mu^*}{\alpha-\alpha\mu^*}$.

Therefore when $\alpha \leq \mu^*$, Strategy 1 is an equilibrium if,

$$\left(\frac{\alpha}{1-\alpha}\right)\left(\frac{l-c}{t+c}\right) + \left(\frac{t-r}{t+c}\right) \geq p$$

Strategy 2 is an equilibrium if,

$$p \geq \frac{t-r}{(1-\alpha)(t+c)}$$

There will exist a p that does not satisfy either of these iff,

$$\frac{t-r}{(1-\alpha)(t+c)} > \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{l-c}{t+c}\right) + \left(\frac{t-r}{t+c}\right) \tag{19}$$

$$t-r > \alpha(l-c) + (1-\alpha)(t-r) \tag{20}$$

$$t-r > l-c \tag{21}$$

So we can conclude that as long as $l - c \geq t - r$, then if $\alpha \leq \mu^*$, then for any monitoring level, either Strategy 2 or Strategy 1 is an equilibrium.

Note that $l - c$ is measuring the downside risk of being taken advantage of compared to the status quo, while $t - r$ is measuring the benefit of comparative advantage to the Competitive types relative to mutual cooperation.

Lemma 7 (Uninformative Strategies)
If $\alpha \leq \mu^*$, then any symmetric strategy profile in which both types play D in period 1 and Competitive types play D in period 2 is a PBE with the belief system $\mu_i^{L,H}(\emptyset) = \mu_i^{L,H}(\bar{s}) = \mu_i^{L,H}(\underline{s}) = \alpha$.

Proof: Because both player types play D in the first period, the first period is irrelevant and the two-period game simply becomes a one-period stage game in the final period. Since $\alpha \leq \mu^*$, then Cooperative types can coordinate on C in the final period for all signals or coordinate on a specific signal. As both player types play D in period 1, the distribution of observed signals cannot be changed by a single player changing their period 1 action to C, hence no deviation to C in period 1 would be rewarded for the Competitive type.

3.2.5 Full Visibility Threshold

Theorem 8 gives an upper bound on the belief α, above which only the strategy profile in which all players defect in both periods is an equilibrium.

Theorem 8 (Full Visibility Threshold)
Define the Full Visibility Threshold as $\gamma^* = \frac{2r+c-g}{2r+l-g}$. Then,

(a) $\gamma^* > \mu^*$

(b) Repeated Violation is an equilibrium for all $\alpha \leq \gamma^*$

(c) If $\alpha > \gamma^*$ then at any monitoring level the only equilibrium is the Defect Strategy.

Proof: (a) Using that $r + c > 0$ and $l - c > 0$, then

$$\gamma^* = \frac{(r-g)+(r+c)}{(r+l-g-c)+(r+c)} > \frac{r-g}{r+l-g-c} = \mu^*$$

(b) We verify the bounds on p. Show that $p^\dagger \leq 1$ for $\alpha \leq \gamma^*$:

$$\begin{aligned}
\alpha &\leq \gamma^* \\
\frac{\alpha}{1-\alpha} &\leq \frac{\gamma^*}{1-\gamma^*} \\
\frac{\alpha}{1-\alpha} &\leq \frac{2r+c-g}{l-c} \\
\frac{\alpha}{1-\alpha} &\leq \frac{r+c}{l-c} + \frac{r-g}{l-c} \\
\left(\frac{\alpha}{1-\alpha}\right)\left(\frac{l-c}{r+c}\right) &\leq 1 + \frac{r-g}{r+c} \\
\left(\frac{\alpha}{1-\alpha}\right)\left(\frac{l-c}{r+c}\right) - \frac{r-g}{r+c} &\leq 1 \\
p^\dagger &\leq 1
\end{aligned}$$

For $p^* = \frac{\alpha-\mu^*}{\alpha-\alpha\mu^*}$, note that p^* is a strictly increasing function in α, and $p^* = 1$ when $\alpha = 1$. Then $1 > \gamma^* \geq \alpha$ implies $p^* < 1$.

(c) Assume $\alpha > \gamma^*$, and hence $\alpha > \mu^\star$ as well by (a). Therefore Temporary Compliance cannot be an equilibrium since it requires $\alpha \leq \mu^\star$. Using the result in (b) Repeated Violation cannot be an equilibrium since the required monitoring level must satisfy $p \geq p^\dagger > 1$ for $\alpha > \gamma^*$. Only the Defect Strategy will be an equilibrium.

3.3 Proposition 3

Otherwise, an agreement will occur so long as the minimum level of information exchange required to reassure each side about catching the other cheating is not too costly. The minimum level of information exchange increases in the belief that either side is a Competitor

When states lack sufficiently confident beliefs about one another's incentives for cooperation, agreements will only be viable if they provide enough monitoring for states to be able to detect whether cheating has occurred and change their behavior to avoid further disadvantages. The more suspicious states are that the other is a Competitor, the greater their need to accurately monitor behavior. Better monitoring (a higher p in the model) is costly, and to some extent, states will see it as worthwhile to pay these costs in order to risk cooperation and reap its benefits. But at very high costs, the agreement will not be worth it overall.

That cooperation on implementing an agreement with some level of monitoring can occur in equilibrium is shown by Lemma 3, Lemma 5.

We also consider what happens to the minimum level of monitoring (p) as the probability that the other side is a Competitor (α) increases.

Lemma 9 (Monitoring bounds increasing in α)
The minimum monitoring bound for the Repeated Violation equilibrium is strictly increasing in α for $\alpha > \mu^*$.

Proof: The minimum monitoring bound is $\max\{p^*, p^\dagger, 0\}$. When $\alpha > \mu^*$, then $p^*, p^\dagger > 0$, and hence one of these two levels is binding. We therefore show that both of these terms are strictly increasing in α. For $p^* = \frac{\alpha-\mu^*}{\alpha-\alpha\mu^*}$ we evaluate the derivative with respect to α:

$$\frac{d}{d\alpha}\left(\frac{\alpha-\mu^*}{\alpha-\alpha\mu^*}\right) = \frac{d}{d\alpha}\left(\frac{1-\mu^*\alpha^{-1}}{1-\mu^*}\right) = \frac{\mu^*\alpha^{-2}}{1-\mu^*} > 0$$

For $p^\dagger$, since $\frac{\alpha}{1-\alpha}$ is strictly increasing, and $l > c$ implies $\frac{l-c}{r+c} > 0$, then

$$\alpha \uparrow \quad \Rightarrow \quad p^\dagger = \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{l-c}{r+c}\right) - \left(\frac{r-g}{r+c}\right) \uparrow$$

The lower bound on the monitoring level is the maximum of two functions that are both strictly increasing in α, which implies that the maximum of the two functions is strictly increasing in α.

To get to the point of states selecting a monitoring level that can both be implemented in a cooperative equilibrium and is not too costly, we next consider the bargaining game.

3.3.1 Bargaining Game

In the bargaining stage, the players must decide on either a treaty with monitoring level p or no treaty at all. If a monitoring level is agreed upon, then that level is implemented and play moves to the second stage. If there is no agreement, then the game ends with both players receiving status quo per period payoffs $(-c_1, -c_2)$.

The cost of monitoring level p is given by the function $c(p)$, which is increasing in p. Low monitoring is less costly than high monitoring. This might be a "cost" in terms of the resources involved to implement an intrusive monitoring measure or other kinds of costs associated with allowing observation of the state's territory beyond the scope of the adversary's unilateral ability to collect information.

Lemma 10 (Optimal monitoring)
The minimum monitoring will be the optimal monitoring for Repeated Violation equilibrium if the cost of monitoring increases sufficiently quickly.

Proof: For the Repeated Violations equilibrium, the payoff for the Cooperative type is,

$$2r(1-\alpha) - 2l\alpha + p\alpha(l-c) - cost(p)$$

A sufficient condition to show that the minimum level of monitoring is the optimal one is that the derivative of the payoff with respect to p is negative:

$$\frac{d}{dp}\left(2r(1-\alpha) - 2l\alpha + p\alpha(l-c) - cost(p)\right) = \alpha(l-c) - cost'(p) < 0$$

For any α, the lowest possible monitoring level is optimal if the cost function increases faster than $\alpha(l-c)$. Note that this constraint is increasing in α. Since Repeated Violations is an

equilibrium for $\alpha \leq \gamma^*$, then it is sufficient to have $cost'(p) > \gamma^*(l-c)$. Note that the cost of monitoring always increasing faster than $\gamma^*(l-c)$ is just a sufficient condition.

Lemma 11 shows that Strategy 1 Repeated Violation will be preferable over Strategy 2 Temporary Compliance under conditions.

Lemma 11
Under key conditions for the cost of monitoring, the Temporary Compliance equilibrium will yield a lower payoff to the Cooperative type than the Repeated Violation equilibrium.

Proof: This is only relevant for $\alpha \leq \mu^*$ since this is the only region where Temporary Compliance can be an equilibrium.

The minimum monitoring level for Temporary Compliance is,

$$p^{min}(\alpha) = \frac{t-r}{(1-\alpha)(t+c)}$$

And the Cooperative type's payoff at this minimum monitoring level is,

$$(1-\alpha)2r + \alpha(r-l) - cost(p^{min}(\alpha))$$

This payoff is decreasing in α, since its derivative is,

$$\begin{aligned}\frac{d}{d\alpha}\Big((1-\alpha)2r + \alpha(r-l) - cost(p^{min}(\alpha))\Big) &= -2r + r - l - cost'(p^{min}(\alpha))\frac{d}{d\alpha}p^{min}(\alpha) \\ &= -r - l - cost'(p^{min}(\alpha))\left(\frac{t-r}{(1-\alpha)^2(t+c)}\right) \\ &< 0\end{aligned}$$

Here we have assumed that cost is increasing in monitoring level. Therefore, the maximum payoff for the Temporary Compliance strategy is achieved at $\alpha = 0$, and monitoring level $p^{min} = (t-r)/(t+c)$, yielding

$$2r - cost\left(\frac{t-r}{t+c}\right)$$

In contrast, the payoff from choosing $p = 0$ and the Repeated Violation strategy is (assuming no monitoring is costless), $(1-\alpha)2r - \alpha 2l$, and achieves its minimum at $\alpha = \mu^*$. Therefore, a sufficient condition that the Temporary Compliance strategy yields a lower payoff to the Cooperative type than the Repeated Violation strategy is:

$$\begin{aligned}2r - cost\left(\frac{t-r}{t+c}\right) &< (1-\mu^*)2r - \mu^* 2l \\ 2\mu^*(r+l) &< cost\left(\frac{t-r}{t+c}\right)\end{aligned}$$

Lemma 12 (Pooling on monitoring level)
Both Competitive types and Cooperative types will pool on offering and accepting the same level of monitoring. There is no separating equilibrium where Competitive types will accept or offer a different level of monitoring.

Proof: At any given level of monitoring, the implementation game payoff for the Competitive type will be either the same or higher than that of the Cooperative type. This is the case because the Competitive type will always get the same payoff as the other player if both states violate in the two rounds, at least one higher temptation payoff if the other player complies in the first round, and two higher temptation payoffs if the other player complies in both rounds. Due to its lower temptation payoff, the Cooperative type cannot get a payoff that is higher than the Competitive type.

At any level of monitoring, there is therefore a strategy for the Competitive type that will get it at least the same payoff as the Cooperative type. That strategy is to adopt the same actions that the Cooperative type would in that circumstance.

Suppose there is a monitoring level that only Competitive types would accept or reject, which in effect reveals their identity to the other player. If a Competitive type is revealed, the best response for either type of opponent is to reject the offer and get $-2c$. This is true because if the offer is accepted, then in the implementation, both sides will violate in the second round, the Competitive type will have no incentive to comply in the first round since its identity has already been revealed, and so both players will also violate in the first round. The payoff for both players will be: $-2c + c(p)$.

If a Cooperative type offers a level of monitoring and it is accepted, then $E[U_i^L] > -2c + c(p)$, because otherwise, the player would have preferred the no-agreement status quo without paying the cost for monitoring. The Competitive type can get at least the same payoff, $E(u) > -2c + c(p)$, for the same level of monitoring. This payoff is greater than $-2c$, which is what the Competitive type would have gotten if it revealed its type: $E[U_i^L] > -2c + c(p) > -2c$.

The Competitive type would accept a lower level of monitoring, but the Cooperative type would not have offered it in the first place, preferring the no-agreement status quo $-2c$ instead. If, on the other hand, the Competitive type did the offering, and offered its preferred lower level of monitoring, then in doing so it would reveal its type. The best response strategy of the other player then changes to rejecting the offer and both sides get $-2c$. The Competitive type thus gets no benefit from offering a level that the Cooperative type would not accept, and therefore will also choose not to offer it, so the result is pooling.[15]

Both types of states therefore behave the same way at the treaty design stage, offering and accepting the same level of monitoring. The Competitor can always pretend to be a Cooperator and then gain an advantage by violating when an agreement is implemented. To do so, she accepts the same monitoring level the Cooperator would, which is sufficiently informative to detect violations but not prohibitively costly. Anything else, such as a preference for lower monitoring, would reveal her true type, and lead to no agreement and no

[15] There are trivial separating equilibria where states are indifferent about what they offer, and all offers are always rejected, so types can offer different monitoring levels.

opportunity to gain an advantage from violation.

4 Comparative Statics and Agreement Pathways

To evaluate which level of monitoring is chosen under different beliefs (α), and which strategies are played in equilibrium, we need to make some additional assumptions about the other parameters. We set r, g, t, l, and c to reasonable values according to their relationship as outlined in the game payoffs. Then, keeping these constant we evaluate the optimal level of monitoring p which would be chosen in equilibrium.[16] This is the comparative static of interest; as beliefs about the adversary change, how does this affect whether states are able to sign a treaty, and at what level of monitoring? Numerically, we set $r = 3, l = 2, t = 5, g = 2, c = 0$.[17]

We set the cost function according to the conditions in Lemma 11. Monitoring cost is high enough that a state would not want to pay for monitoring if it was fairly sure that the other side would comply anyway. The cost rises quickly enough by Lemma 10, meaning that states would not want to pay for more than they need, given beliefs about the opponent. In this case, the cost function is a simple linear form, $c(p) = 3p$.

The results of the analysis are presented in Figure 4. Along the x-axis is α, capturing the belief about the adversary. At low levels of α on the left side of the plot, the states believe one another to be likely Cooperator types, and at high levels of α on the right, states believe one another to be likely Competitor types.

Figure 5 recreates the numerical analysis, as well as other experiments with alternate numerical values for the set parameters, in a more graphical format for illustration. This figure labels the parameter spaces where agreements are possible and where they are not.

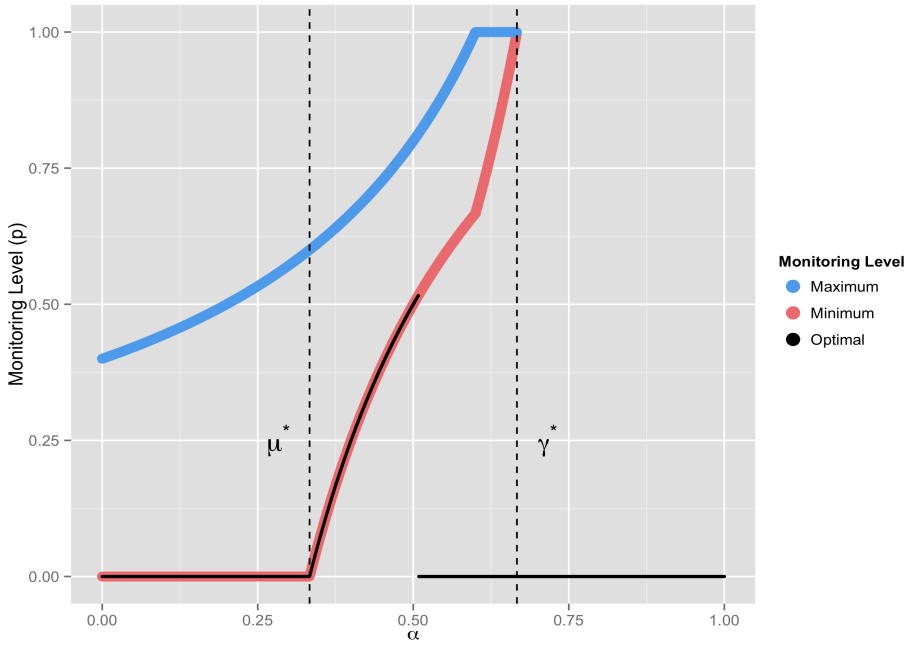
A change in beliefs could mean either new doubts about prior assessment or a more confident estimate about the opponent's incentives. An increase in uncertainty is a case where the direction of change is important. If beliefs change towards either of the two extremes, they become more certain, and if they change towards the middle, towards $\alpha = .5$, they become less certain. Importantly, a shift from an area of relative certainty *towards* the middle means that the belief is still that, on average, the opponent will be the previously expected type. We can differentiate this kind of belief change as an "increase in uncertainty" as opposed to a fundamental change in expectations, which would entail a shift to the other side of the beliefs spectrum.

Corollary 1. An increase in uncertainty makes an agreement more likely. Shaken confidence that the other side is a Cooperator leads to low but non-zero information exchange; shaken confidence that the other side is a Competitor leads to high information exchange.

[16] While analytic solutions are often preferred, in this case demonstrating the implications of equilibrium conditions using numeric simulations provides an effective way to quickly understand the relationships between the main variables of interest. Alastair Smith, "Political Groups, Leader Change, and the Pattern of International Cooperation," *Journal of Conflict Resolution* 53, no. 6 (2009): 853–877.

[17] Note that in the payoff matrix, the sucker payoff is $-l$.

Figure 4: Comparative Statics Plot – optimal monitoring selected



Increases in uncertainty imply predictions about agreement outcomes that are both distinct from and more precise than general belief changes. If beliefs change from either extreme and become less certain, the likelihood of crossing a cooperation threshold increases. The starting point—initial belief—matters for treaty form as it determines which threshold, and consequently which range of monitoring, will be closest.

Figure 6 illustrates the intuition of the hedging and risking "pathways" for agreement. As beliefs change from relative certainty towards uncertainty (the middle), the likelihood of crossing either of the two thresholds increases. Outcomes for monitoring types depend on where beliefs started before the shift.

Figure 5: Illustration of Parameter Space and Cooperation Outcomes

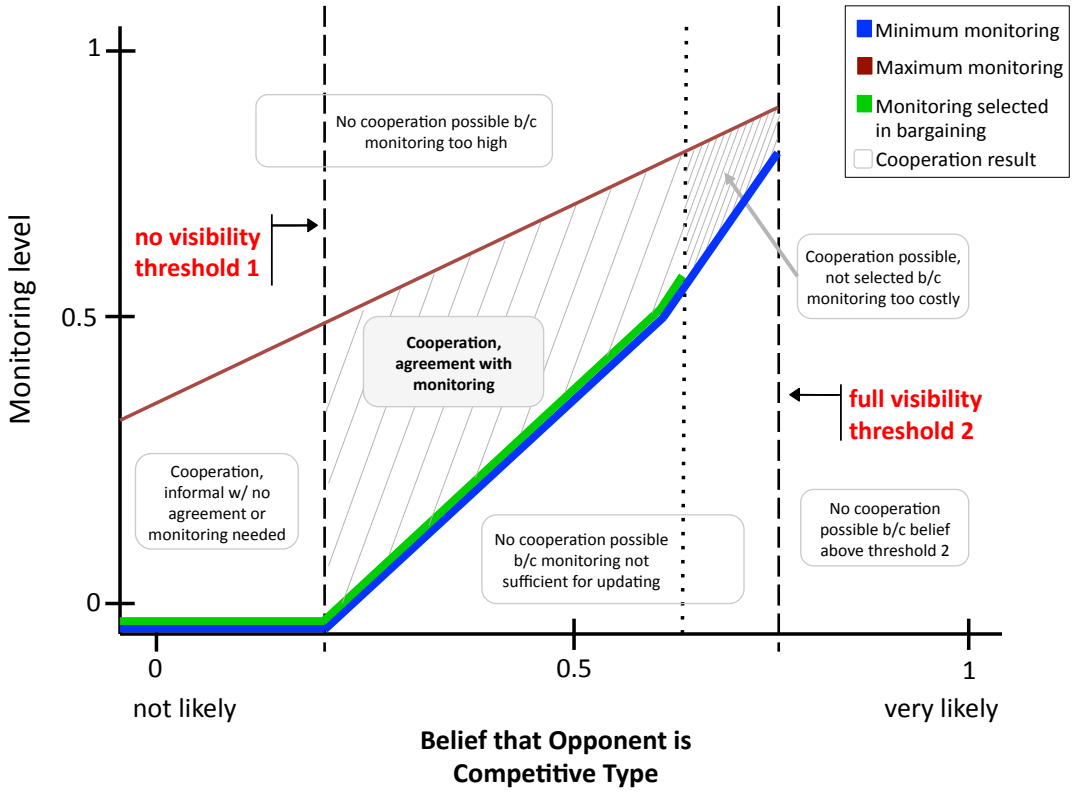
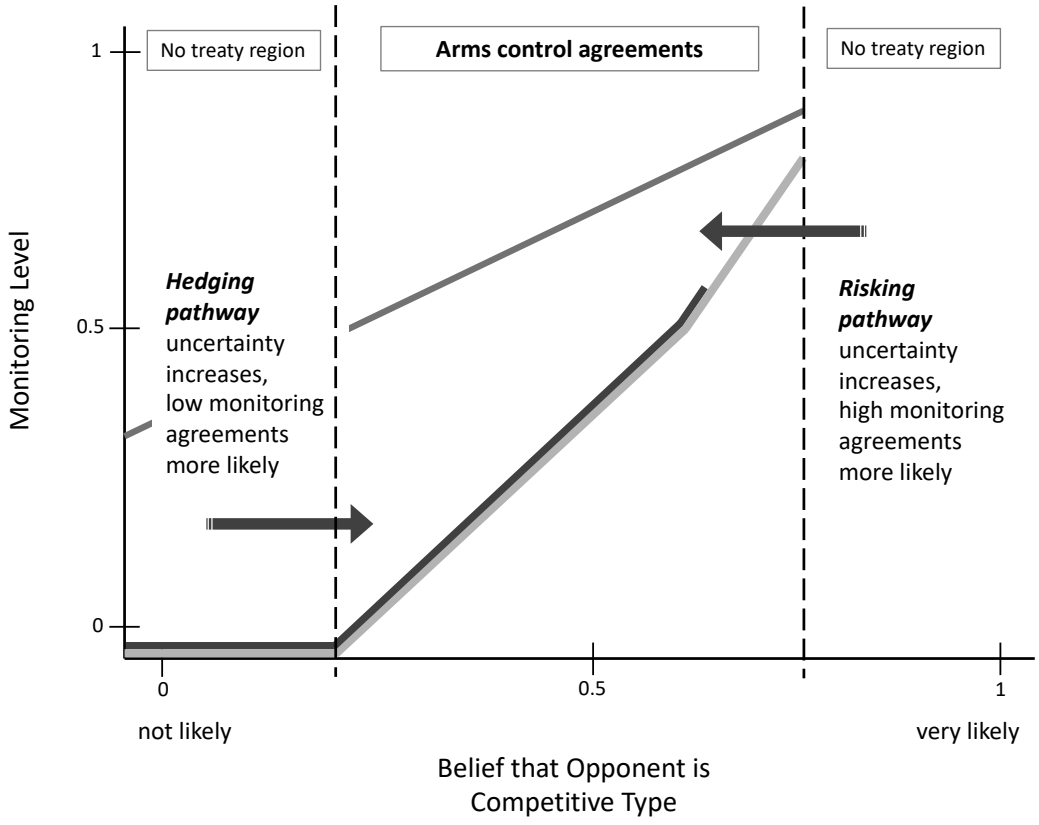


Figure 6: Illustration of Agreement Pathways



Online Appendix Part II
Statistical Analysis for *Enemies in Agreement: Political Volatility and the Design of Arms Control*

Jane Vaynman

Contents

1 Dataset Details **B3**

2 Explanatory Variables **B5**

3 Control Variables **B9**

4 Descriptive Statistics **B14**

5 Statistical Approach **B18**

6 Full Results: Hypothesis 3 **B19**

7 Full Results: Hypotheses 1 and 2 **B19**

8 Alternate Measures of Baseline Beliefs **B26**

9 Alternate Measures of Domestic Political Shifts **B31**

10 Socialization and Learning Results **B34**

11 Other Robustness Checks: Rare Events and Time Period Controls **B34**

List of Figures

1 Ideal point distance over time B9
2 Arms Control Treaties over Time B15
3 Dependent Variable Treaty Coding, Dyadic View B16
4 Effects of Cabinet or Executive Change on Treaty Types (generalized ordered logit) B24
5 Effects of Demonstrations, Riots, and Strikes on Treaty Types (generalized ordered logit) B25
6 Effects of Cabinet or Executive Change on Treaty Types (multinomial logit) B27
7 Effects of Demonstrations, Riots, and Strikes on Treaty Types (multinomial logit) B28

List of Tables

1 Agreement information provision types and members B15
2 CE Effects on Probability of a Treaty (logit) B20
3 DRS Effects on Probability of a Treaty (logit) B21
4 Cabinet/Executive Volatility Effects on Agreement types (generalized ordered logit) B22
5 Demonstrations, Riots, and Strikes Volatility Effects on Agreement Types (generalized ordered logit) B23
6 Cabinet/Executive Volatility Effects on Agreement Types (multinomial logit) B29
7 Demonstrations, Riots, and Strikes Volatility (DRS events) Effects on Agreement Types (multinomial logit) B30
8 Alternate Beliefs Measures, CE Change (generalized ordered logit) B32
9 Alternate beliefs measures, DRS Change (generalized ordered logit) B33
10 Socialization and Learning, CE Change Effects B35
11 Socialization and Learning, DRS Change Effects B36
12 Time Periods (logit) B38
13 Time Periods (generalized ordered logit) B39

This portion of the Appendix provides additional information, full empirical results, and robustness checks for Chapter 3 of *Enemies in Agreement: Political Volatility and the Design of Arms Control.* Replication data and code are available online at Harvard Dataverse https://doi.org/10.7910/DVN/EUFGGL or upon request from the author.

1 Dataset Details

The Adversarial Agreements Dataset takes the form of dyad-year observations, where each pair of states in any given year can be observed as signing no agreement, a low monitoring agreement, or a high monitoring agreement. This section provides additional details on the case universe of agreements that are included in the dataset. Agreements are coded as having either a "Low" degree of information exchange or a "High" degree of information exchange. This variable was coded based on the text of the agreement. In this Appendix, as in the book itself, the terms "agreement" and "treaty" are used interchangeably. The legalization level of each agreement was coded for this data, but it is not used in this analysis.

1.1 Criteria for Inclusion

Arms control agreements are defined as any formal agreement that limits a security-related capability or behavior of states. An agreement is considered formal as long as it is recorded in written form. Agreements can take a variety of legal forms, including a ratified treaty or a military-level agreement. The collection includes agreements that: 1) establish a limitation (in quantity or quality) of armaments or military personnel; 2) restrict movement, basing, or fortifications on a select territory; 3) restrict or ban either the use of certain weapons or certain military activities during a future conflict; 4) call for disarmament or demilitarization of troops or territory; 5) limit states' abilities to launch a surprise attack or build new capabilities in secret. These agreements all address activities or capabilities that have the potential to give one adversary an advantage in a conflict.

Some security-related agreements define borders between two states without including additional terms of what may or may not happen along those borders. These are not included in the dataset. War termination and ceasefire agreements that contain arms control provisions are included, as are confidence-building measures (CBM), which often make fewer demands on participants but still limit first-strike incentives and secrecy in military development. These two types of agreements are less commonly referred to as "arms control" by today's policy practitioners, but including them as adversarial cooperation agreements is not controversial—existing research often notes that the Versailles Treaty is a form of imposed arms control and that CBMs between the United States and USSR or between India and Pakistan are smaller-scale controls. The 225 included agreements represent the full universe of arms control treaties rather than a sample.

A wide range of sources was used to collect both the list of all agreements and the full text of each. These include: an essential arms control compilation volume by Jozef Goldblat

titled *Arms Control: The New Guide to Negotiations and Agreements,*[1] the UN Treaty series database, NGO sources, government foreign ministry records, UK parliamentary records, and printed histories for individual treaties or sets of treaties. The Russian Ministry of Foreign Affairs treaty archive, as well as a wide range of secondary literature, were used to identify agreements that are less likely to be recorded in Western sources. Several agreements available in Russian only are included in the collection. The date range surveyed was from 1700-2010, though only a subset of these agreements, from 1945-2010 (H1 and H2) and 1816-2010 (H3), is used in the statistical analysis due to data limitations on explanatory variables.

1.2 Dependent Variable Coding

The full text of each treaty was coded for information provisions. The codes are: (1) *None:* No reference to information exchange or only voluntary measures included; (2) *Reporting and Notification:* Information is given by one state to the other, but no mandated capacity for the adversary to have independent access to information; (3) *Observation/Monitoring:* One side is allowed to directly observe the capabilities or activities of the other state; (4) *Verification:* Information is collected, observed, and then checked by the opponent, primarily through the use of on-site inspections. Coding was conducted using a close reading of each treaty text, following a codebook to identify information provisions. The coding was initially done at a very detailed level, and then groups of codes were aggregated into more general categories. For example, textual codes include: reporting to state parties, reporting to an independent agency, reporting in response to an inquiry, and reporting with details specified in the agreement. All the reporting-related codes combine into the broader "reporting and notification" code. As a secondary check, the codes were compared with descriptions of treaties and their provisions in other studies. While not all treaties are discussed in existing historical work, several sources include assessments of information in selected treaties.[2] This comparison shows that the coding in this project is consistent with other expert assessments.

1.3 Agreement Members

Multilateral and bilateral agreements are included in the data. Bilateral agreement members are self-explanatory, but challenges emerge for coding members of multilateral agreements. The issue of multi-party events comes up for studies of conflict, as of course numerous wars involve many state parties. In that data, multilateral events are simply turned into a collection of bilateral events between all members, and as a consequence conflicts such as the two World Wars are over-represented as numerous bilateral cases.

---

[1] Jozef Goldblat, *Arms Control: The New Guide to Negotiations and Agreements* (London, UK: Sage Publications, 2002).

[2] These include: Trevor N. Dupuy and Gay M. Hammerman, *A Documentary History of Arms Control and Disarmament* (New York, NY: T.N. Dupuy Associates, 1973); Serge Sur, ed., *Verification of Current Disarmament and Arms Limitation Agreements* (Aldershot, England: Dartmouth Publishing Company, 1991).

In the case of adversarial treaties, numerous agreements have over 50 and even over 100 members. Treating all parties as members would create massive over-representation for some treaties. To better fit a theory based on states' decisionmaking and evaluation of tradeoffs the data only count some states as treaty "designers," focusing on those which are most likely to have something to gain or lose from cooperation, and therefore are likely to have an impact on treaty provision choices. In most cases, not all 100 participants had any influence on the form of the treaty; most are joiners into a negotiation that is largely controlled by more powerful actors. I apply a systematic cut-off rule for multilateral treaty participants. For treaties with more than ten participants, I only count the ten most powerful states, as ranked by their CINC scores. This method narrows down the participants to states which are most likely to both have the negotiating leverage to influence the shape of the treaty, and the military capabilities that the treaty seeks to limit. Historical evidence supports the assumption that in large multilateral treaties, both the intended targets of limitations – the dyad pairs seeking to restrain one another – and the treaty designers are usually only a few powerful states. For example, the United States and Russia were often the driving forces in the design of major multilateral treaties such as the Nuclear Nonproliferation Treaty and the Biological Weapons Convention.

The date of each agreement is recorded as the date the agreement was signed. I do not use ratification dates for several reasons. First, not all of the agreements are treaties that would require ratification. Second, the decision to pursue a security agreement, the negotiation of its design, and the actions surrounding compliance are the purview of the executive. The executive can choose to comply with an agreement or violate it regardless of whether ratification has occurred. While ratification may carry additional signaling or commitment benefits, it is not an essential element of security agreements. Finally, the timing between signature and ratification can be determined by a number of domestic political factors that have little to do with the security tradeoffs that drive the decisions on whether limitations are pursued in the first place and the manner in which states decide to monitor compliance.

2 Explanatory Variables

The key explanatory variables in the analysis are domestic political volatility and prior beliefs about the adversary.

The political volatility variables are created using the Cross-National Time-Series (CNTS) dataset.[3] A number of studies have used the CNTS data to capture political volatility or instability, but there is no consensus on the best approach or the best measures to use. To list a few: Sasikumar and Way (2009) codes a measure of political instability using a weighted index of CNTS variables, combining general strikes, demonstrations, and government crises; Ghosn (2011) uses demonstrations and government crises as separate variables; Brunetti and Weder (1998) use number of changes in the national constitution as a measure of policy uncertainty; Roe and Siegel (2011) use government crisis with a 1 percent decay rate assess

[3] Arthur S. Banks, *Cross-National Time-Series Data Archive*, Databanks International, Jerusalem, Israel, 2011.

the impact of crises over a 30-year period.[4] The variables used in this analysis build on ideas developed in such prior studies and adapt the measures to better fit the concept of foreign policy relevant domestic political volatility.

The Cross-National Time-Series data is monadic, recording events that happen in each state each year. I convert the data to dyad-year observations. Both political volatility variables are coded 1 if volatility occurred in a given year in one or both of the dyad partners, and 0 otherwise. A political shift in one state would be sufficient to cause an increase in uncertainty in the dyad. From the point of view of the observing state, the shifting one may be changing in its incentives. Conversely, from the point of view of the shifting state, it may expect that the observer sees different incentives in response to an altered adversary. For example, food riots in Egypt cause Israel to be less certain about Egypt's incentives for arms racing and aggression (perhaps to distract from domestic unrest) vs. cooperating and saving resources for domestic welfare programs. At the same time, Egyptian leaders would be uncertain as to whether Israel would see the unrest as weakness and try to take advantage or see the unrest as a new reason to cooperate on arms control, since perhaps broader relations with Egypt would be affected by pressures on the government.

Demonstrations, riots, and strikes: Demonstrations, riots, and political strikes are indicators of public pressures that could cause a government to make changes in its policies, particularly in allocating resources to "guns vs. butter." The issue here is not necessarily how the government responds, but the signals that these events send to an adversary about pressures influencing the government's future policies. Observing such events, an adversary could become less certain in its preexisting assessment of the state's incentives. The binary variable "DRS events" is coded 1 if more than one demonstration, riot, or strike occurred in either dyad state in the last year, and 0 otherwise.

The measure seeks to capture events that would be substantial enough as to be observable internationally as domestic pressures. Since the data does not record the sizes or issue areas associated with these social unrest events, we can rely on their frequency. Coding more than one event per year reduces noise and counts only cases that might in fact be interpreted as public pressure on government policy, rather than an occasional or sporadic public outcry. In the data, many states experience one demonstration in any given year, while fewer experience more than one event. Approximately the same frequency levels apply for general strikes and riots. The distribution suggests that some level of unrest is normal, but far fewer cases experience multiple events, so this heightened level is more likely to be observed as a departure from normal domestic political conditions.

Cabinet and executive changes: A signal of policy change can come from upper-level

[4] Karthika Sasikumar and Christopher R. Way, "Testing Theories of Proliferation in South Asia," in *Inside Nuclear South Asia*, ed. Scott D. Sagan (Stanford Security Studies, 2009); Faten Ghosn, "Domestic Unrest and the Initiation of Negotiations," *International Negotiation* 16, no. 1 (2011): 69–85; Aymo Brunetti and Beatrice Weder, "Investment and Institutional Uncertainty: A Comparative Study of Different Uncertainty Measures," *Weltwirtschaftliches Archive* 134, no. 3 (1998): 513–533; Mark J. Roe and Jordan I. Siegel, "Political Instability: Its Effects on Financial Development, Its Roots in the Severity of Economic Inequality," *Journal of Comparative Economics*, no. 39 (February 2011): 279–309.

shifts within the government. The executive may make changes to government structure in response to public pressures, internal pressures such as competition among elite groups, or to reflect changes in the beliefs of new leaders. These leadership shifts raise uncertainty about foreign policy because new leaders can either choose to continue the policies of their predecessors or pursue new directions. This top-level policy can come from leaders themselves or from the influence of other key actors in senior positions.

To capture top-down volatility, a new measure is constructed using CNTS data on cabinet and executive changes. The variable is coded 1 if there was at least one executive change or more than one major cabinet change in a given year in one or both of the dyad partners, and 0 otherwise. Again, counting only more than one cabinet change is an effort to remove some noise from this variable and capture only moments that suggest a serious possibility of changes in foreign policy that would be observed by another state. A cabinet change in CNTS is defined as over 50 percent of the cabinet replaced. About 40% of observations experience one cabinet change, which suggests that a relatively high degree of cabinet turnover is part of normal political affairs, and likely not a sign of political volatility. A far lower percent of the observations experience more than one cabinet change, which suggests that multiple changes would in fact be interpreted as associated with a possibility of new policy direction. Changes in the executive happen less frequently, are more likely to always be observed by other states, and are more likely to each be indicators of an increase in uncertainty. The variable is also coded 0 if there were over four of either cabinet or executive changes in one year. Such a high frequency of leadership turnover suggests that the government might not have the capacity to actually conduct international diplomacy. In practice, very few observations have such a high level of leadership turnover.

Other variables available in the CNTS data, such as revolutions, guerrilla warfare, and even government crises, are indicators of instability more serious than "volatility" and so are not appropriate measures for the concept in this analysis.[5] When these events occur, a government may be likely to turn its attention on responding to an immediate threat at home, rather than looking to international cooperation as a possible solution.[6] From the adversary's perspective, during such changes it may not even be clear who in the state government has the authority to sign promises of arms limitation, so adversaries are likely to wait until the right interlocutor can be identified. At lower levels of volatility (or what other studies call "policy uncertainty" or policy volatility), the government is still in control of determining policy and there are incentives to cooperate internationally rather than wait for the unrest to abate and incur the costs of noncooperation. At the end of instability periods such as a revolution or a coup, when the government or challenger group gain control, may also be moments of new foreign policies emerging, but here we again want to measure the new government coming in rather than the revolution starting.

---

[5] For example, "government crisis" is defined as "any rapidly developing situation that threatens to bring the downfall of the present regime excluding situations of revolt aimed at such overthrow." These include situations where "military law was declared, a state of siege, the suspension of the constitution, a vote of no confidence by a parliamentary majority, or a call for impeachment of top officials."Banks, *Cross-National Time-Series Data Archive*.

[6] Ghosn, "Domestic Unrest and the Initiation of Negotiations."

Beliefs about the adversary: To capture the prior belief about the other state, I look at states' general foreign policy positions. The measure of "policy ideal point" estimates, developed by Bailey, Strezhnev, and Voeten calculates each state's policy position based on its voting record in the UN General Assembly.[7] States that have closely aligned policy positions (low ideal point distance) are likely to see their incentives for complying with cooperation as mutually aligned, and so are likely to expect one another to be Cooperative types. On the other hand, states with distant policy positions are likely to see their security interests as counter to one another, and so are more likely to have high incentives to take advantage by cheating on cooperation.

The measure is an improvement on prior efforts that similarly used UN voting patterns to capture the nature of states' positions relative to one another. The Affinity of Nations measure of policy similarity used UN voting patterns to calculate the maximum metric distance between two states.[8]

The degree of similarity or difference between the policy positions of each pair of states can be determined by calculating the distance between the ideal points. While the distance in UN ideal point estimates is the best available proxy for capturing states' beliefs about one another's incentives, it only measures states that are members of the UN, which narrows the statistical analysis to the post-1945 period. I also test several alternative measures of beliefs, which is detailed in the last section of this Appendix.

Ideal points distance has both benefits and drawbacks as a measure for baseline beliefs about an opponent's incentives for violating a deal. The measure has important advantages in capturing policy positions in a more fine-grained way than alternative measures and is consistent over time. The most significant drawback is that the measure is better at capturing a state's policy position relative to the US-led liberal order rather than a regional rival. Some rival states measure as closely aligned on ideal points because they both adopt positions that are either close to or in opposition of the US-led order. However, the problem in how rivals are characterized by this measure may also be a limited one when it comes to expectations about cheating on cooperation. Rivals aligned in their mutual opposition to the liberal order, for example, may also have strong incentives to comply with an agreement, possibly to avoid interference from the United States.

The measure does reasonably well on "face-validity" checks with a few dyads that are intuitively important for thinking about arms control outcomes. Ideal point distance changes line up with detentes and spikes in suspicion between the United States and the USSR, the end of the Cold War, and a resumption of greater tension two decades later. The range of variation for the India-Pakistan dyad is much narrower over time, as they do measure as relatively aligned, but particularly low distances do line up with relative detentes. Some

[7] Michael A Bailey, Anton Strezhnev, and Erik Voeten, "Estimating Dynamic State Preferences from United Nations Voting Data," *Journal of Conflict Resolution* 6, no. 2 (2017).

[8] Anton Strezhnev and Erik Voeten, *United Nations General Assembly Voting Data*, v. 4, August 2012. There, the metric distance calculation uses a method developed by: Curtis S. Signorino and Jeffrey M. Ritter, "Tau-b or Not Tau-b: Measuring the Similarity of Foreign Policy Positions," *International Studies Quarterly* 43, no. 1 (March 1999): 115–144.

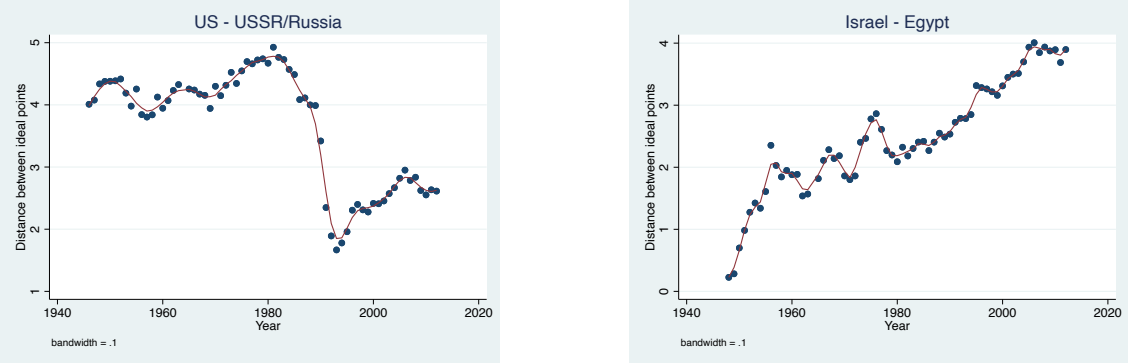


Figure 1: Ideal point distance over time

military rival dyads are more problematic – Ethiopia and Eritrea measure very close on distance despite active conflict periods.

Time lags: The analysis was run with both prior beliefs lagged three years with domestic volatility occurring in the last two years as well as with prior beliefs lagged two years with volatility measures in the last year. The results are not meaningfully different. The lags are intended to ensure that the analysis captures the beliefs that existed prior to the occurrence of domestic political volatility and not any shifts in policy positions that might be coming about as a result of the volatility.

3 Control Variables

The primary datasets used for control variables are the Correlates of War (COW) Militarized Interstate Dispute dataset and National Material Capabilities dataset. I use the EUGene (v. 3.204) program to generate dyad-year observations.[9] A number of the variables are created using the procedures used by Bennett and Stam.[10]. These control variables are selected because there is reason to think that they co-vary with the key independent variables (prior relationship between the states and domestic political volatility) and may have an effect on the dependent variable (agreement signature and type of information provisions).

Balance of power: The balance of power (in terms of material capabilities) likely has effects on agreements because it could be a factor in both the benefits of cooperation and the costs of maintaining the status quo. Equally balanced states may be more willing to sign treaties because in those cases limitations are perceived to be even and less likely to change the overall balance of power. Many arms control scholars have argued that arms control agreements are more likely to emerge when military capabilities are evenly balanced.[11]. However, some

[9] D. Scott Bennett and Allan C. Stam, "EUGene: A Conceptual Manual," *International Interactions* 26 (2000): 179–204.

[10] D. Scott Bennett and Allan C. Stam, *The Behavioral Origins of War* (Ann Arbor, MI: University of Michigan Press, 2004).

[11] Thomas Risse-Kappen, "Did Peace Through Strength End the Cold War? Lessons from INF," *Interna-*

treaties are imposed rather than signed completely voluntarily, so highly uneven dyads might be those where a powerful state can impose its preferred negotiated solution on a weaker adversary. Hegemony might lead to more intrusive information demands in treaties. Here, dyads facing asymmetry may face more cost to maintaining the status quo, such as a conflict continuing without surrender, and are more likely to accept costly institutional tools. The question of whether parity or inequality leads to more or less cooperation or a certain type of cooperation is not dissimilar from the debate in security studies on which of these types of dyads are more war-prone, where arguments have been made both ways. Finally, relative power balance may also have an effect on whether states see one another as having aligned or diverging security interests.

I generate the balance of power variable using CINC scores as the power measure. In non-directed dyads, it does not matter which state is A or B, and I set the greater power state as A. *Balance of power* $= \frac{CountryApower}{sum(A+Bpower)}$ The variable ranges from .5=even balance of power to 1 = asymmetric balance of power. The CINC scores for each state in the dyad come from the COW dataset.

Recent conflict: States may be more likely to engage in arms control if they recently engaged in conflict. Avoiding competition, future conflict, or a resumption of the recent conflict would give states more to gain from cooperation. Numerous wars have ended with ceasefire agreements, some of which have arms control provisions and some which do not. Case study evidence also suggests that arms control agreements are sometimes signed following crises where escalation to war was avoided. Recent conflict may also inform perceptions of the adversary's security interests. Finally, it may affect political volatility, for example if leadership was altered in the course of conflict, or if conflict led to public grievances over the domestic provision of goods or government services.

The binary recent conflict variable that is coded 1 if a MID level 2 or higher occurred in the last five years, and zero otherwise. This eliminates minor conflicts, like fishing disputes and non-reciprocated incidents, but gives a relatively permissive time frame for how long states may take to respond to security tensions.

Arms race: States engaged in an arms race may have different incentives for arms control than those not spending resources on military competition. This variable is likely to capture the costs of maintaining the status quo in some cases. I construct the arms race variable using Diehl's (1983) definition of an 8 percent or more growth in military expenditures over 5 years in both states.[12]. This measure was used by Bennett and Stam in their analysis on the causes of war,[13] and is considered the best available in other studies as well, including Rider 2011.[14]. Military expenditures data come from the COW National Material Capabilities

tional Security, 1991, 162–188.

[12] Paul F. Diehl, "Arms Races and Escalation: A Closer Look," *Journal of Peace Research* 20, no. 3 (1983): 205–212.

[13] Bennett and Stam, *The Behavioral Origins of War*.

[14] Toby J. Rider, Michael G. Findley, and Paul F. Diehl, "Just Part of the Game? Arms Races, Rivalry, and War," *Journal of Peace Research* 48, no. 85 (2011): 85–100.

data.

Some scholars have been concerned that this measure does not account for whether the arms build-up is actually directed at the other state in the dyad.[15] However, "inadvertent" arms races, or in other words, mutual state build-ups not directed at one another, can lead to suspicion and further arms build-up, and so likewise drive incentives for arms control. States could have higher incentives to cooperate during any type of military build-ups as a way to mitigate burgeoning rivalry, mitigate the risk of direct competition in the future, or constrain the costs of the arms build-up. The simple Diehl measure, which does not include information about the relationship between the states, is still appropriate for the study of arms control design. Ideally, an arms race variable would also incorporate expected expenses in military planning for future years, as arms control may also be a way to avoid an impending arms race rather than mitigate an ongoing one. However, data limitations do not allow for this kind of analysis.

Wealth: Countries have different baseline capabilities to observe their adversaries and different levels of resources to finance monitoring measures. These factors are likely to affect the cost and relative benefit of agreement-mandated monitoring, and thus affect states' likelihood of pursuing specific treaty forms. For wealthier states, the cost (and also the benefit) of monitoring may be lower because they likely already have high investment in intelligence or technology to observe other states. It may also be the case that wealthier states can afford to maintain arms races for longer than poorer states and might therefore have different incentives for pursuing arms control treaties.

I use a wealth measure to capture differences in monitoring technology or resources that might be available to states. In order to get a measure with maximum coverage over the time frame and number of countries in my data, I use the energy component of COW National Material Capabilities as a proxy for wealth. I include the measure for both state A and state B because both states are agreeing to a monitoring level in bargaining and the burden of costs of monitoring may be negotiated as well (for example, with the wealthier state contributing more to the costs of monitoring.) This captures the possible difference between relatively wealthy vs. relatively poor dyads. I create the measure as: $ln(energyA + energyB)$. As a robustness check, I also create the same measure using GDP, where data is only available after 1945. For this alternative measure, I use Kristian Skrede Gleditsch's Expanded Trade and GDP Data (v.4.1).[16]

Contiguity: States that share a land border are more likely to have militarized disputes and contested territory. This suggests that they are also more likely to have agreements that resolve those disputes or manage territorial tensions. Again, these states may have more to gain from cooperation than non-contiguous dyads, a greater gain from violating an

[15] Douglas M. Gibler, Toby J. Rider, and Marc L. Hutchinson, "Taking Arms Against a Sea of Troubles: Conventional Arms Races During Periods of Rivalry," *Journal of Peace Research* 42, no. 2 (March 2005): 131–147.

[16] Kristian Skrede Gleditsch, "Expanded Trade and GDP Data," *Journal of Conflict Resolution* 46, no. 5 (October 2002): 712–724.

agreement, or more to lose when an opponent chooses to violate. I use a binary variable of whether the dyad is contiguous on land, based on COW data.

Nuclear weapons: Nuclear-armed states face more acute existential threats from conflict escalation. These states may therefore have stronger incentives for cooperation, both in the nuclear arms control arena and also in conventional arms, because the costs associated with maintaining a risky or unstable status quo are very high. It may also be the case that nuclear weapons create a particularly active new arena for cooperation, and nuclear states are likely to have more agreements than non-nuclear states because non-nuclear states simply do not have that technology to limit. Finally, because of the delivery systems used for nuclear weapons, states that possess them are also likely to have other advanced technology applicable for observing adversaries, such as satellites and advanced aircraft. I code nuclear weapons 1 if both states in the dyad possess nuclear weapons and 0 otherwise.

Ongoing treaty: Having ongoing cooperation could make new cooperation either more or less likely. New cooperation could be more likely if the prior treaty or treaty negotiations set up a certain agenda for cooperation, identifying other weapons or activities that should be limited as separate treaties. On the other hand, having recently signed a treaty could mean that the states are essentially already doing all the cooperation that was possible. The situation here is not unlike that of studies of war initiation, where scholars usually control for "ongoing conflict."

The ongoing treaty variable is coded as 1 if a treaty was signed by the dyad in the last 5 years, and zero otherwise. The data on treaty signature is from the adversarial agreements dataset. Defining ongoing treaties by stated duration or "in force" status is not useful for capturing the possible effects of recently concluded cooperation. Few arms control treaties have a limited duration and others remain technically in force even though they are no longer explicitly followed by their participants.

Temporal and dyadic dependence: To correct for possible time dependence, I use the Carter and Signorino (2009) method, including time, time squared, and time cubed variables.[17] Time in this context is the number of years after the last agreement. I address possible dyadic dependence by using dyad-clustered standard errors.

3.1 Variables for Robustness Checks

Beyond the main analysis, a number of additional controls are included in additional model specifications to test robustness of the results.

Satellites: An argument is sometimes made by both policymakers and scholars that states without these satellite capabilities do not pursue arms control because they lack the tech-

[17] David B. Carter and Curtis S. Signorino, "Back to the Future: Modeling Time Dependence in Binary Data," *Political Analysis* 18, no. 3 (2010): 271–292.

nological capacity to observe compliance.[18] The question of satellite capabilities is a more specific version of a parameter already included in the models: a state's baseline ability to observe the other side, which is measured by wealth. I use data on when states acquire government-controlled satellites to test whether this capability is associated with treaties.[19] Early's data goes through 2003, so I extend it through 2007 by applying capabilities that existed in 2003 to the next four years as well. While this basic technique may miss a few states that have obtained satellite capability in the last few years, we would not expect states to lose any capability they had already acquired.

Unit-specific effects: We might have some concerns that certain key characteristics of states' relationships are omitted from the analysis. In other words, there might be certain characteristics unique to the dyad which affect those states' beliefs about one another as well as their likelihood to sign agreements. One way to address this problem would be to use a unit-level specific control, namely a dyad fixed effects model. However, in an FE model all units which do not have variation in the DV over the full panel would be dropped. In the treaty cooperation case, many dyads never sign an agreement for reasons that are being directly modeled by the other variables. An FE model would drop a large amount of data that is informative for understanding which conditions are associated with cooperation. The main regressions do not test a fixed effects specification for this reason. In international relations applications, there is no consensus on how to deal with possible dyad unit effects, and some scholars strongly advise against fixed effects models.[20] The logit model used to test Hypothesis 3 does implement an FE specification as a robustness check. (A random effects model (not reported) also gave the same results (direction, approximate magnitude, and significance) for tests of H3. However, the substantive meaning of unit effects in this case – unobserved characteristics or propensities for cooperation in particular dyads – suggests that, if anything, an FE model is more appropriate. A Hausman test comparing the RE and FE specifications likewise suggests the FE model is better.) There are limited options for fixed effects for multinomial logit. A user-written Stata command introduced in 2014, femlogit, is supposed to calculate a fixed effects multinomial logit, but it failed to run on this data, possibly due to data size and number of variables.

Instead of accounting for all individual dyads through fixed effects, I check for possible effects from a couple of dyads which we might expect to be both particularly unique in the relationship between the two states and have an additional propensity for cooperation or competition that is not captured by other variables. I code a binary variable for whether the dyad is US-USSR (or US-Russia) and a second binary variable for whether the dyad is India-Pakistan. Both pairs of states have been involved in rivalries with repeated periods of detente and high tension. They have a long history of adversarial interaction, and both

[18] See for example Glaser 1994 on the role of satellite observation in making arms control possible.Charles L. Glaser, "Realists as Optimists: Cooperation as Self-Help," *International Security* 19, no. 3 (1994): 50–90

[19] Bryan Early, "Exploring the Final Frontier: An Empirical Analysis of Global Civil Space Proliferation," *International Studies Quarterly*, Forthcoming,

[20] Nathaniel Beck and Jonathan N. Katz, "Throwing out the Baby with the Bath Water: A Comment on Green, Kim, and Yoon," *International Organization* 55, no. 2 (Spring 2001): 487–495.

have also signed numerous agreements. Given their history and nature of competition, these dyads would be the most likely candidates for unit-level effects as an omitted variable in accounting for their cooperation patterns.

Joint democracy: The main regressions do not control for joint democracy because the key characteristics of democracies that may contribute to beliefs or higher uncertainty and treaty outcomes are proxied by other control variables. Joint democracy may be a noisier variable to capture the same concepts. For example, democracy could mean lower costs of monitoring. Joint democratic states could have more access to information about one another because their political systems are more transparent, and so we might expect that they have reduced costs of observing one another's compliance in arms control. However, it is not clear that democracies will always be any less secretive or any less able to protect information on military capabilities or technologies. Lower costs of observing compliance could be better captured by identifying dyads with high resource capacity that can invest more in intelligence or technology rather than generalizing about democratic ones. Joint democracies are on average wealthier dyads in my data, and all the models already control for dyad wealth.

We might also hypothesize that democracies are, because of the nature of their political systems, more interested in formal legal institutions in general.[21] This would suggest that jointly democratic dyads are more likely to have more treaties. However, the analysis already controls for treaties in the last five years (last ten years tested as a robustness check) which captures treaty propensity for a number of reasons that might include either interest in international institutions or recurring security challenges. Democratic dyads on average have more treaties in the past five or ten years than mixed or autocratic dyads.

In effect, the other control variables are already capturing several factors we might expect to be different among democratic states, but in a way that allows us to assess nondemocratic dyads along the same more detailed continuum as well. However, even if this logic for choosing controls with better precision is not convincing, or if there is still concern about democracy having separate effects, robustness checks were performed using a democratic similarity measure. I create the variable joint democracy, which is coded 1 if both states in the dyad have a polity score greater than 10, and 0 otherwise.

4 Descriptive Statistics

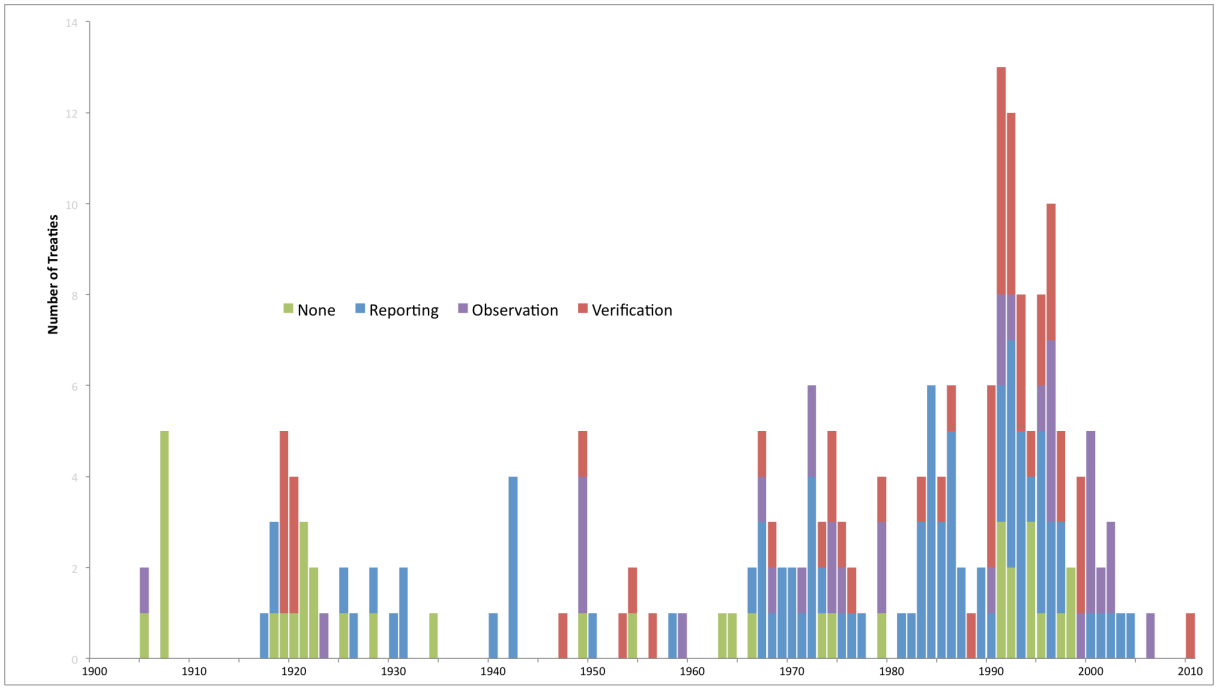
The tables below provide some additional information about the data. Table 1 is included in Chapter 3 of the book.

[21] Some research in international relations suggests that democracies are more willing to comply with international commitments because of their existing commitments to domestic constitutional obligations, so perhaps a similar logic would suggest that they are more likely to form them as well. Kurt Taylor Gaubatz, "Democratic States and Commitment in International Relations," *International Organization* 50, no. 1 (Winter 1996): 109–139.

Table 1: Agreement information provision types and members

Type of agreement	**Bilateral**	**Small multilat.** (3–9)	**Large multilat.** (10+)	***Total***
Low monitoring	67	25	45	137
High monitoring	32	11	45	88
Total	99	36	90	225

Figure 2: Arms Control Treaties over Time



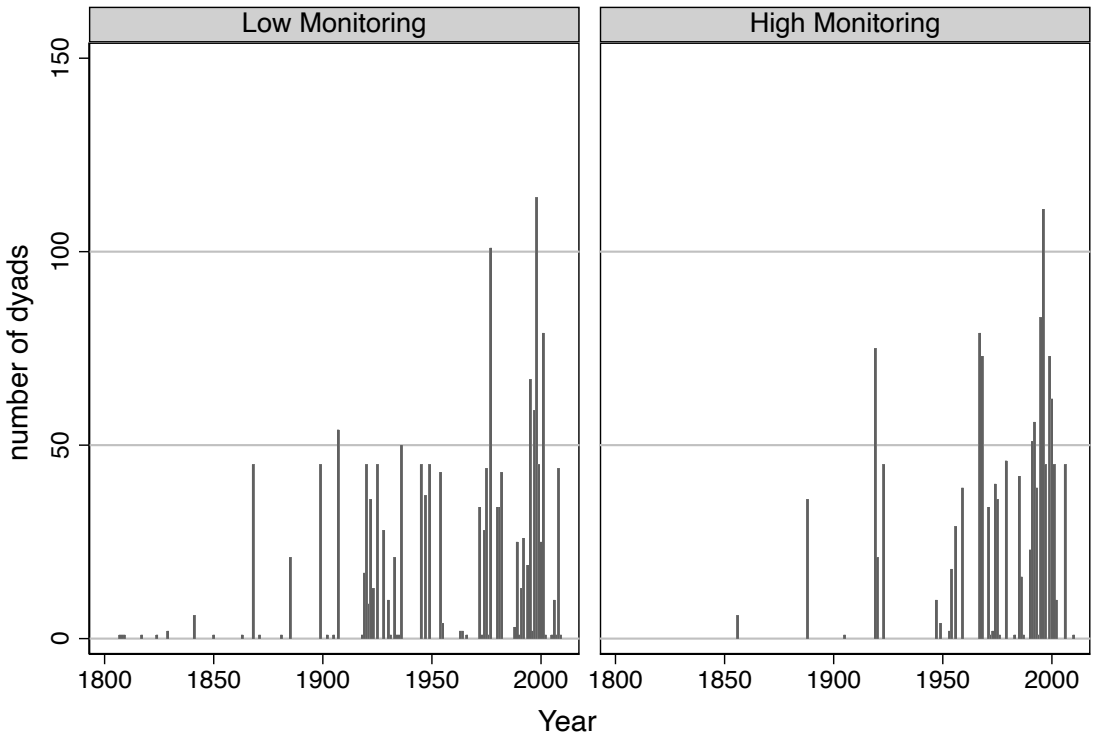
4.1 Dyadic Observations

For the statistical analysis, the treaty data are converted into dyadic form, where each observation is a pair of states signing no agreement, a low monitoring agreement, or a high monitoring agreement. Dyadic interactions can better reflect the strategic nature of signing an agreement and closely fit the theoretical view of treaties as outcomes of attributes and beliefs of two actors.

The process of creating dyads from agreement observations incorporates the fact that some treaties are multilateral in terms of their designers, but the limitations are structured as being between two sides. For example, states A, B, and C may be members of a treaty, but the treaty is an agreement between states A and B together against C. In this case, a dyad between A and B would not accurately reflect the nature of the agreement. In many cases, the treaty text explicitly identifies parties on one side and parties on the other. For these treaties, dyads are created only between states on opposite sides of the agreement, so A-C and B-C, but not A-B. The coding is based only on the text of the treaty itself, not on any additional assumptions about the relationships between members.

Figure 3 shows the dyadic view of the agreements, identifying the number of pairs of states that sign an agreement in a given year. As in all the statistical tests that follow, the provision categories are collapsed into two: low and high degree of information exchange.

Figure 3: Dependent Variable Treaty Coding, Dyadic View



The analysis starts with the assumption that any pair of states could conclude an arms control agreement in principle, regardless of whether they tried or not. In practice, this means all dyad-years are used as observations.[22]

Unlike some past studies of conflict that see only some dyads as being plausibly able to engage in war or militarized dispute, the analysis here does not restrict the data to "politically relevant dyads" for theoretical reasons specific to the nature of arms control treaties. Narrowing to politically relevant dyads seeks to limit the analysis to those pairs of states that researchers believe to have a sufficiently high probability for conflict. However, this approach would be problematic for arms control. States with little prior connection may still be involved in an arms control agreement, particularly on issues that present a wide-ranging security threat. Indeed, about 20 percent of treaty dyads are observed in non-politically relevant dyads. Second, arms control can happen in anticipation of future threats and capabilities states do not yet possess. This anticipation makes states with little current interaction possibly threats in the future, and therefore possible targets for arms limitation efforts. While it is the case that some dyads have a very low probability of ever signing an agreement, that probability is explicitly modeled through the explanatory and control variables.

Rather than introducing strong assumptions about which states would have a zero probability for arms control, I apply only one modest method of limiting the observations. Members of NATO and the Warsaw Pact share capabilities and had a very high degree of coordination in their war planning. These alliances are highly institutionalized, durable, and constantly maintained. Dyads between members are not able to create arms control agreements between themselves because their military capabilities, behaviors, and security goals are in many ways shared; a limitation or restraint would be more akin to a unilateral decision rather than a negotiated bargain between sides with opposing security interests.

Some dyads between NATO countries are still created from the treaty data because, during the Cold War, the "sides" were often not stated explicitly in treaty preambles but still understood by all participants. But an observation of a treaty between two NATO or two Warsaw states is not an attempt to limit one another as adversaries, but rather the agreement was negotiated alongside the alliance partner. The dataset thus drops all dyads within members of NATO and within the Warsaw Pact during the time frames that states were members of each alliance, including observed treaty dyads. The majority of dropped treaty dyads are US-UK.

Finally, if a pair of states signed more than one treaty in a given year, the treaty with the most intrusive monitoring is counted as an instance of cooperation. This accounts for the difference in the number of unique treaties listed in treaty data and the unique treaties that show up in the dyadic analysis.

[22] Following past studies, this analysis uses nondirected dyads generated using the EUGene software, which combines data from the COW project, Polity IV project, and several other common sources. Several of the other control variables are either directly output by the program (contiguity, energy expenditure) or derived through manipulations of provided variables (moving average of arms buildup, joint democracy, etc.) D. Scott Bennett and Allan C. Stam, *EUGene*, 3.1 (December 2005).

5 Statistical Approach

The H3 models are standard logistic regression with dyad-clustered standard errors. For the conditional Hypotheses 1 and 2, the generalized ordered logit (gologit) allows me to use the ordering information of the dependent variable, while making few assumptions about the effects of independent variables and controls. The model and its application in Stata are detailed in several articles by Richard Williams.[23] Although generalized ordered logits are not very widely used in political science, they appear more commonly in sociology, health, and education research.[24]

The main results use a fully unconstrained version of the generalized ordered logit. Proportional odds are not imposed on any of the variables. On the one hand, there are no clear theoretical reasons to argue that some variables should be constrained and not others. On the other hand, using an unconstrained model when at least some variables should have been constrained to proportional odds likely weakens the results. It also increases the possibility that some observations have an outcome with a predicted probability less than zero, and indeed, nearly all the model specifications have some in-sample cases with predicted probabilities below zero. This is because, over a large parameter space, nonparallel lines must cross. As indicated by the gologit2 documentation, negative predicted probabilities are not necessarily a problem. The recommendations are to leave as is if there are relatively few cases of negatively predicted probabilities, or simplify the model or data. Although there is no clear theoretical basis for constraining some variables, there is an empirical test to help assess which variables do or do not meet the proportional odds assumption. I ran this test on the main models, and in accordance with the results then ran a partially constrained generalized ordered logit, with the following variables constrained: wealth, balance of power, contiguity, recent dispute, and arms race. The results were almost exactly the same as for the unconstrained model.

The results tables are presented below, with results reported as odds ratios and relative risk ratios for better interpretability. However, it is important to note that the generalized ordered logit model is more difficult to interpret simply from coefficients or relative risk ratios. The marginal effects plots in the main text of the article provide a clearer picture of the results. As a robustness check, and to address concerns that might arise based on the gologit model itself, results are reported for both the multinomial logit and generalized ordered logit for most models. Importantly, this demonstrates that the results are not model specific and are consistent in direction, significance level, and rough magnitude of substantive effects.

The multinomial logit is also well-suited for studying categorical variables, but it is

[23] Richard Williams, "Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables," *Stata Journal* 6, no. 1 (2006): 58–82; Richard Williams, "Understanding and interpreting generalized ordered logit models," *The Journal of Mathematical Sociology* 40, no. 1 (2016): 7–20. For a further technical discussion, see also James W. Hardin and Joseph M. Hilbe, *Generalized Linear Models and Extensions*, Second (College Station, TX: Stata Press, 2007).

[24] For a political science example, see Regina P Branton and Bradford S Jones, "Reexamining Racial Attitudes: The Conditional Relationship Between Diversity and Socioeconomic Environment," *American Journal of Political Science* 49, no. 2 (2005): 359–372.

somewhat less appealing in this case because it cannot incorporate the ordering information for the dependent variable – the information that a high monitoring treaty is more intrusive than a low monitoring one is not used. The model allows all coefficients for all variables to vary across categories. Therefore, a multinomial logit is the least restrictive approach, and results which are significant from the mlogit are usually even stronger when the analysis is redone using models that impose restrictions.

6 Full Results: Hypothesis 3

Tables 2 and 3 show the full results for the logit models of treaty outcomes. Note that time, time squared, and time cubed controls were included in this model but coefficients are not reported. All were significant but very close to 1 for the odds ratio, indicating a very small magnitude effect. Model 1 is the main model, for which marginal effects were calculated and reported both here and in Chapter 3. The following models introduce several additional variables to test alternative explanations.

7 Full Results: Hypotheses 1 and 2

The ordered logit models are presented in the main text of Chapter 3 as well as here. The multinomial logit results reported here are an additional check to show that the results are robust to different modeling alternatives.

All tables report either odds ratios or relative risk ratios (for multinomial logit results). All models include dyad-clustered standard errors and variables for $time$, $time^2$, and $time^3$, which are not reported.

The marginal effects plots show a comparison of treaty outcomes for stable vs. volatile conditions across the range of beliefs. Control variables are held constant in this analysis: contiguity, ongoing treaty, nuclear weapons, recent militarized dispute, and arms race are held at median values, and balance of power, resources, and time variables at their means.

7.1 Generalized Ordered Logit Results

Tables 4 and 5 present the results for generalized ordered logit models for the cabinet or executive change volatility measure. Model 1 is the base model used in Chapter 3, and the following models include additional control variables. Controls for time, time squared, and time cubed were included in this model but coefficients are not reported. All were significant but very close to 1 for the odds ratio.

7.2 Multinomial Results

Tables 6 and 7 present the results for multinomial models. Controls for time, time squared and time cubed were included in this model but coefficients are not reported. All were significant but very close to 1 for the odds ratio.

Figures 6 and 7 show the results of margins analysis which isolates the effect of domestic political volatility at different levels of beliefs about the adversary. These figures can be

Table 2: CE Effects on Probability of a Treaty (logit)

	Model 1	Model 2	Model 3	Model 4	Model 5 (FE)
cabinet or executive change	**1.454*** ** (0.075)	**1.224*** ** (0.072)	**1.429*** ** (0.074)	**1.430*** ** (0.074)	**1.237*** ** (0.056)
balance of power	0.0104*** (0.002)	0.00629*** (0.001)	0.0164*** (0.003)	0.0166*** (0.003)	0.0111*** (0.004)
contiguity	3.726*** (0.412)	4.482*** (0.603)	3.892*** (0.475)	3.909*** (0.485)	1.566* (0.285)
recent treaty	5.715*** (0.515)	6.079*** (0.689)	4.714*** (0.437)	4.713*** (0.438)	1.137 (0.087)
nuclear weapons	2.098** (0.565)	2.075* (0.621)	1.476 (0.406)	1.492 (0.468)	3.225*** (0.501)
arms race	1.549*** (0.076)	1.748*** (0.096)	1.571*** (0.078)	1.571*** (0.078)	1.596*** (0.077)
recent MID	1.876*** (0.193)	1.235 (0.150)	1.949*** (0.223)	1.953*** (0.228)	1.249** (0.101)
wealth	1.728*** (0.035)		1.597*** (0.034)	1.596*** (0.034)	1.613*** (0.032)
wealth (gdp)		2.075*** (0.063)			
satellites			2.608*** (0.231)	2.619*** (0.233)	
treaty experience			0.821* (0.065)	0.821* (0.065)	
rivalry			1.188 (0.252)	1.210 (0.275)	
US–Russia				1.065 (0.378)	
India–Pakistan				0.571 (0.192)	
N	688,027	595,276	688,027	688,027	62,664

$*p < 0.05, **p < 0.01, ***p < 0.001$

Table 3: DRS Effects on Probability of a Treaty (logit)

	Model 1	Model 2	Model 3	Model 4	Model 5 (FE)
demonstrations, riots or strikes	**1.717*** (0.083)	**1.479*** (0.078)	**1.756*** (0.084)	**1.757*** (0.085)	**1.258*** (0.066)
balance of power	0.0111*** (0.002)	0.00659*** (0.001)	0.0177*** (0.004)	0.0179*** (0.004)	0.0135*** (0.006)
contiguity	3.432*** (0.401)	4.174*** (0.571)	3.934*** (0.492)	3.967*** (0.503)	1.390 (0.285)
recent treaty	7.047*** (0.689)	6.520*** (0.781)	5.992*** (0.601)	6.000*** (0.602)	1.793*** (0.159)
nuclear weapons	1.643 (0.446)	1.778 (0.539)	1.228 (0.324)	1.214 (0.364)	2.551*** (0.447)
arms race	1.730*** (0.085)	1.790*** (0.098)	1.770*** (0.087)	1.770*** (0.087)	1.579*** (0.079)
recent MID	1.794*** (0.185)	1.333* (0.160)	2.076*** (0.245)	2.068*** (0.249)	1.376*** (0.129)
wealth	1.696*** (0.037)		1.542*** (0.034)	1.541*** (0.034)	1.435*** (0.038)
wealth (gdp)		1.987*** (0.063)			
satellites			2.592*** (0.218)	2.599*** (0.220)	
treaty experience			0.949 (0.076)	0.949 (0.076)	
rivalry			0.758 (0.183)	0.757 (0.203)	
US–Russia				1.181 (0.437)	
India–Pakistan				0.801 (0.261)	
N	631323	578911	631323	631323	43857

$*p < 0.05, **p < 0.01, ***p < 0.001$

Table 4: Cabinet/Executive Volatility Effects on Agreement types (generalized ordered logit)

	M1		M2		M3		M4	
1: Low Monitoring								
CE changes	0.988	(0.101)	0.953	(0.099)	0.990	(0.102)	0.986	(0.101)
policy distance	0.878*	(0.046)	0.878*	(0.047)	0.877*	(0.046)	0.874*	(0.047)
CE change X policy	1.241***	(0.075)	1.246***	(0.078)	1.242***	(0.075)	1.245***	(0.075)
balance of power	0.009***	(0.002)	0.016***	(0.004)	0.009***	(0.002)	0.009***	(0.002)
contiguity	4.052***	(0.536)	4.586***	(0.632)	4.050***	(0.539)	4.101***	(0.548)
recent treaty	8.722***	(1.031)	7.270***	(0.851)	8.359***	(0.976)	8.757***	(1.032)
nuclear weapons	2.036*	(0.583)	1.558	(0.438)	2.029*	(0.589)	1.974*	(0.649)
recent MID	1.214	(0.154)	1.241	(0.154)	1.207	(0.155)	1.192	(0.180)
arms race	1.835***	(0.110)	1.857***	(0.112)	1.823***	(0.109)	1.836***	(0.110)
wealth (resources)	1.837***	(0.048)	1.650***	(0.044)	1.857***	(0.050)	1.836***	(0.048)
satellites			2.673***	(0.255)				
treaty experience					0.745**	(0.074)		
US–Russia							1.425	(0.548)
India–Pakistan							0.824	(0.243)
2: High Monitoring								
CE change	0.780*	(0.092)	0.763*	(0.090)	0.783*	(0.094)	0.783*	(0.092)
policy distance	0.844**	(0.053)	0.845**	(0.053)	0.844**	(0.054)	0.838**	(0.054)
CE change X policy	1.380***	(0.088)	1.386***	(0.089)	1.379***	(0.088)	1.381***	(0.087)
balance of power	0.011***	(0.003)	0.019***	(0.005)	0.011***	(0.003)	0.011***	(0.003)
contiguity	3.366***	(0.490)	3.718***	(0.565)	3.371***	(0.494)	3.409***	(0.504)
recent treaty	5.465***	(0.682)	4.764***	(0.576)	5.178***	(0.653)	5.434***	(0.684)
nuclear weapons	2.070*	(0.612)	1.729	(0.509)	2.051*	(0.614)	2.129*	(0.707)
recent MID	1.530**	(0.224)	1.569**	(0.227)	1.517**	(0.225)	1.556**	(0.250)
arms race	1.528***	(0.117)	1.563***	(0.124)	1.515***	(0.118)	1.525***	(0.117)
wealth (resources)	1.715***	(0.054)	1.572***	(0.053)	1.736***	(0.056)	1.715***	(0.054)
satellites			2.110***	(0.240)				
treaty experience					0.712*	(0.096)		
US–Russia							1.065	(0.400)
India–Pakistan							0.443**	(0.136)
N	513024.000		513024.000		513024.000		513024.000	

$*p < 0.05, **p < 0.01, ***p < 0.001$

Table 5: Demonstrations, Riots, and Strikes Volatility Effects on Agreement Types (generalized ordered logit)

	M1		M2		M3		M4	
1: Low Monitoring								
DRS events	1.269*	(0.130)	1.293*	(0.135)	1.285*	(0.132)	1.274*	(0.132)
policy distance	0.857*	(0.057)	0.863*	(0.056)	0.856*	(0.057)	0.856*	(0.057)
DRS events X policy	1.212**	(0.085)	1.193*	(0.083)	1.209**	(0.085)	1.210**	(0.085)
balance of power	0.009***	(0.002)	0.017***	(0.004)	0.009***	(0.002)	0.009***	(0.002)
contiguity	4.043***	(0.526)	4.562***	(0.619)	4.040***	(0.530)	4.063***	(0.538)
recent treaty	8.508***	(1.013)	7.084***	(0.835)	8.056***	(0.954)	8.525***	(1.010)
nuclear weapons	1.823*	(0.514)	1.382	(0.382)	1.811*	(0.519)	1.808	(0.585)
recent MID	1.238	(0.163)	1.272	(0.159)	1.229	(0.163)	1.241	(0.193)
arms race	1.842***	(0.110)	1.865***	(0.113)	1.829***	(0.110)	1.841***	(0.110)
wealth	1.782***	(0.048)	1.600***	(0.042)	1.807***	(0.049)	1.782***	(0.048)
satellites			2.706***	(0.256)				
treaty experience					0.682***	(0.068)		
US–Russia							1.143	(0.454)
India–Pakistan							0.808	(0.237)
2: High Monitoring								
DRS events	1.094	(0.139)	1.082	(0.136)	1.104	(0.141)	1.089	(0.140)
policy distance	0.846*	(0.062)	0.851*	(0.061)	0.843*	(0.063)	0.841*	(0.062)
DRS events X policy	1.277**	(0.106)	1.270**	(0.103)	1.276**	(0.106)	1.280**	(0.106)
balance of power	0.010***	(0.003)	0.018***	(0.005)	0.009***	(0.003)	0.010***	(0.003)
contiguity	3.518***	(0.512)	3.820***	(0.573)	3.509***	(0.514)	3.490***	(0.524)
recent treaty	5.000***	(0.669)	4.257***	(0.548)	4.742***	(0.631)	4.957***	(0.662)
nuclear weapons	1.763	(0.517)	1.492	(0.432)	1.755	(0.522)	1.886	(0.619)
recent MID	1.483*	(0.230)	1.545**	(0.234)	1.480*	(0.233)	1.562**	(0.264)
arms race	1.536***	(0.119)	1.557***	(0.123)	1.532***	(0.120)	1.532***	(0.119)
wealth	1.690***	(0.050)	1.550***	(0.048)	1.709***	(0.052)	1.691***	(0.050)
satellites			2.035***	(0.225)				
treaty experience					0.750	(0.112)		
US–Russia							0.797	(0.319)
India–Pakistan							0.428**	(0.132)
N	508666.000		508666.000		508666.000		508666.000	

$*p < 0.05, **p < 0.01, ***p < 0.001$

Figure 4: Effects of Cabinet or Executive Change on Treaty Types
(generalized ordered logit)

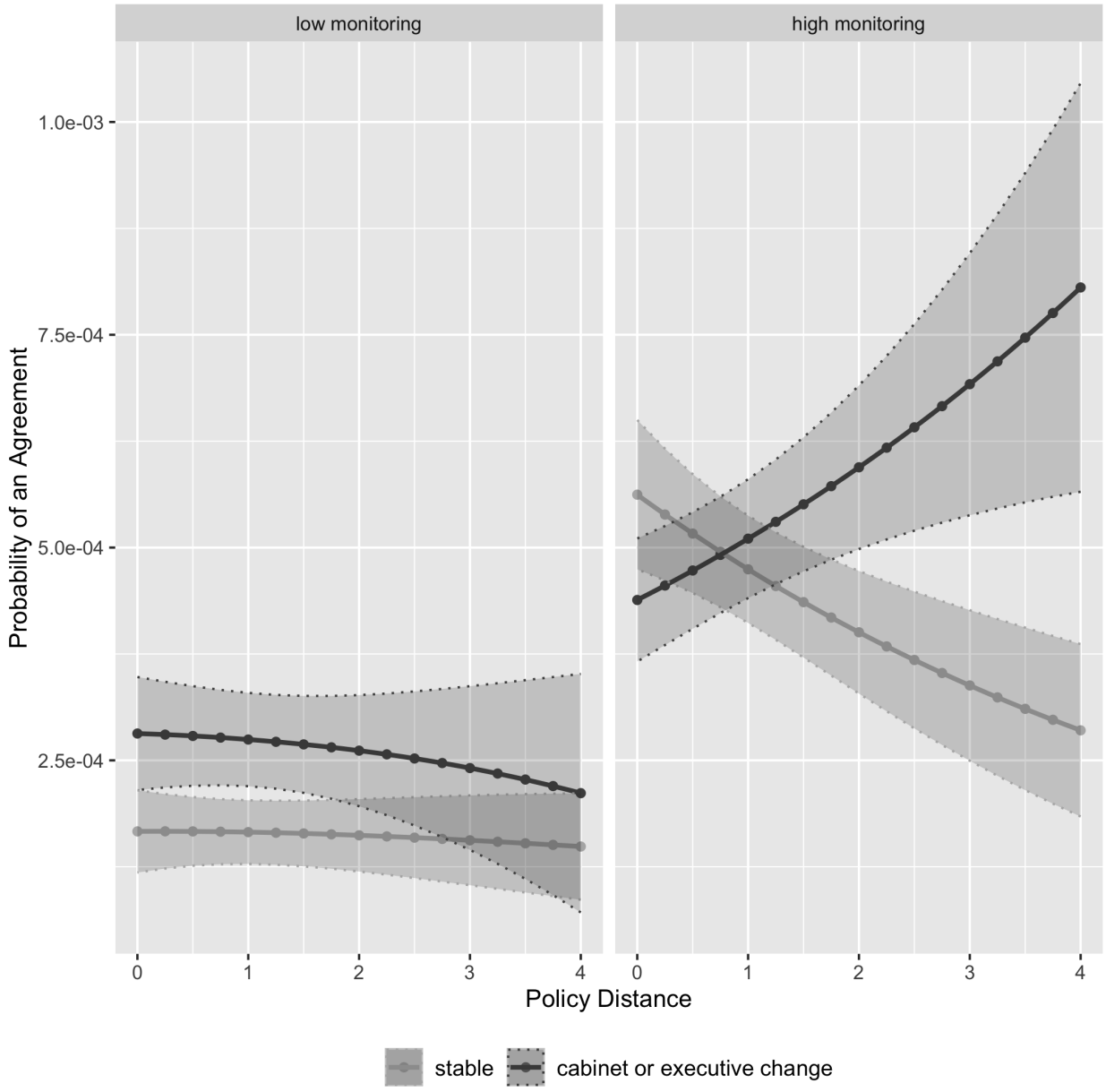
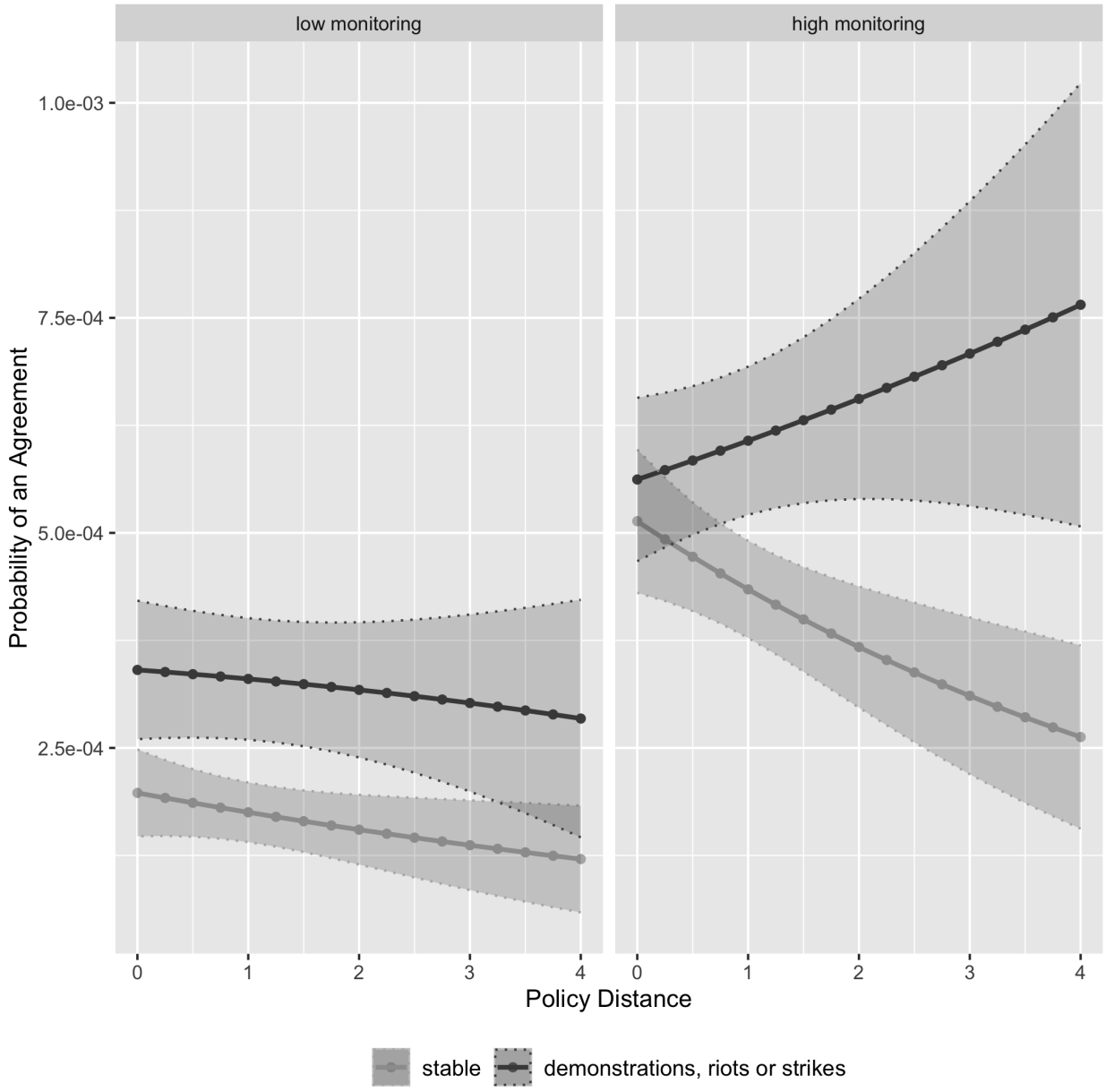


Figure 5: Effects of Demonstrations, Riots, and Strikes on Treaty Types (generalized ordered logit)



interpreted in the same way as the figures for the generalized ordered model which are provided and discussed in the main text of Chapter 3.

8 Alternate Measures of Baseline Beliefs

Several alternative proxies for the baseline beliefs variable were also tested. The United Nations General Assembly (UNGA) voting-based ideal point distance measure, though imperfect, is the best proxy available for the temporal and geographic span of the data. However, I also tested how other possible measures fare in its place. In each case, the new measure was also used to create the interaction term. The tables below provide the results for each of the volatility measures paired with the alternate beliefs measure, with all the base controls kept the same. Results for the controls are not reported.

Model 5 uses Yonatan Lupu's measure of "treaty commitment distance" between states.[25] It is calculated the same way as the UNGA votes based policy distance measure, but instead of votes in the UN determining a state's ideal point, it is rather the states' membership in international treaties.

By Lupu's own assessment, his measure and the UNGA measure may capture somewhat different preference spaces. He suggests that the UNGA-based measure more likely captures states' views on current issues on the international agenda while the treaty commitment distance measure is more likely to capture states' preferences for future commitments through international institutions. (Lupu finds a positive but weak correlation between his treaty commitment distance measure and the UNGA ideal point distance measure.) Engagement in international trade is also a strong predictor of treaty commitment distance, which may suggest a focus on economic preferences. In the context of the questions addressed in *Enemies in Agreement*, preferences over current policy issues are closer to the concept of states' estimates of their adversaries' security incentives. It is possible that Lupu's measure could capture states' attitudes towards the utility of formal institutions in general, and thus test a constructivist-based explanation that states with greater experience and acceptance of international institutions as tools are more likely to use them to manage security relationships as well. The results of the robustness test suggest some support for this alternative hypothesis, as higher treaty commitment distance is associated with a lower likelihood of both low and high monitoring arms control treaties.

Model 6 uses a measure of enduring rivalries developed by William Thompson.[26] The rivalries are determined based on historical assessments of cases, rather than a measure of military spending or changes in capabilities. For each dyad-year, a binary variable is coded for whether or not the two states were engaged in a rivalry. In the context of this analysis, we would expect that states in a rivalry would be more likely to expect that their possible cooperation partner has Competitive incentives, while those not in a rivalry are likely to hold beliefs that the opponent is Cooperative. Unfortunately, this measure is not

[25] Yonatan Lupu, "Why Do States Join Some Universal Treaties but not Others? An Analysis of Treaty Commitment Preferences," *Journal of Conflict Resolution*, Forthcoming,

[26] William R. Thompson and David R. Dreyer, *Handbook of International Rivalries: 1494-2010* (2012).

Figure 6: Effects of Cabinet or Executive Change on Treaty Types (multinomial logit)

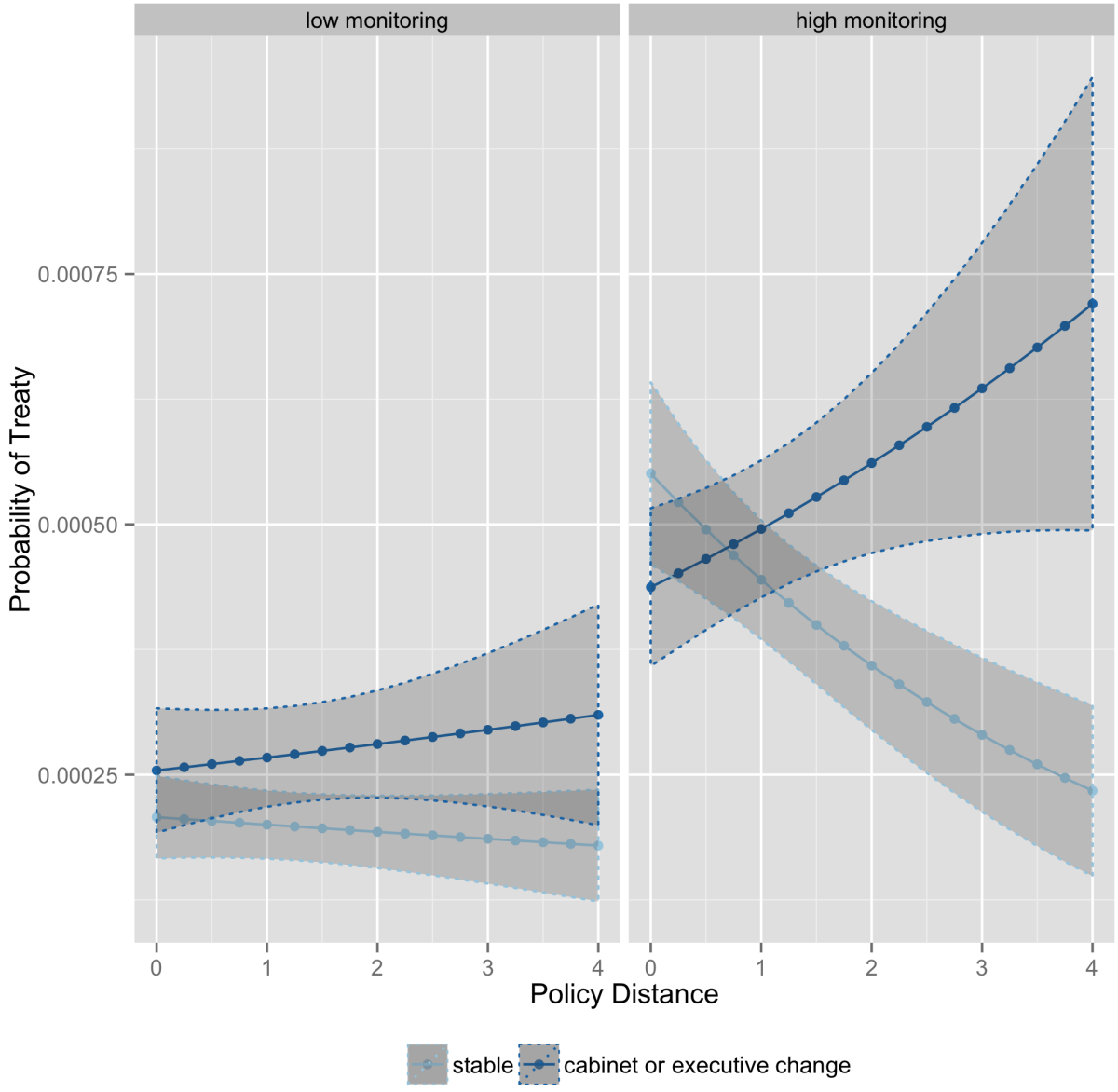


Figure 7: Effects of Demonstrations, Riots, and Strikes on Treaty Types (multinomial logit)

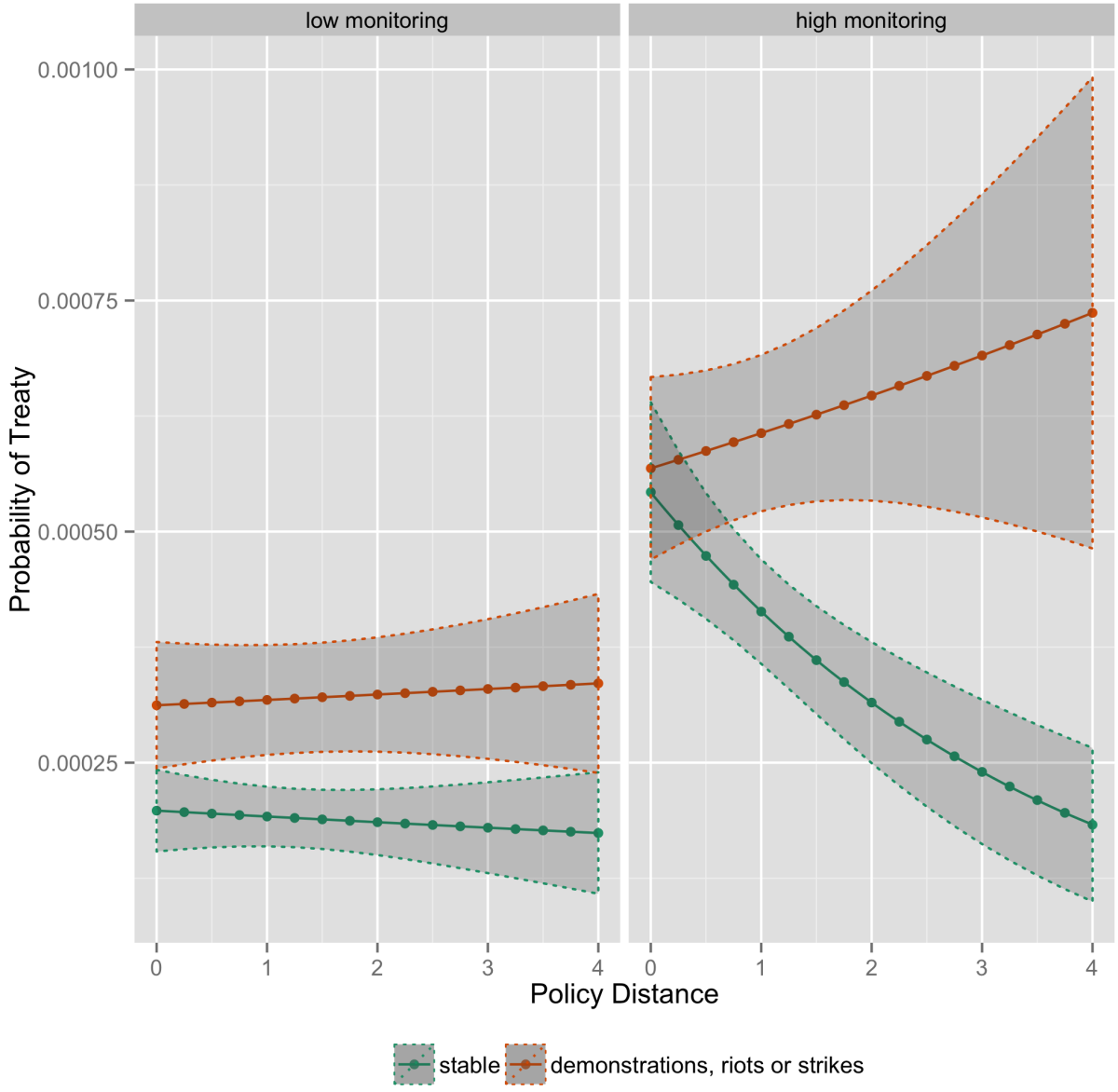


Table 6: Cabinet/Executive Volatility Effects on Agreement Types (multinomial logit)

	M1		M2		M3		M4	
1: Low Monitoring								
CE change	1.225	(0.170)	1.154	(0.163)	1.226	(0.170)	1.220	(0.169)
policy distance	0.964	(0.053)	0.959	(0.054)	0.963	(0.053)	0.963	(0.054)
CE change X policy	1.090	(0.094)	1.105	(0.098)	1.090	(0.094)	1.094	(0.094)
balance of power	0.009***	(0.002)	0.020***	(0.006)	0.009***	(0.002)	0.009***	(0.002)
contiguity	4.745***	(0.767)	5.540***	(0.924)	4.743***	(0.769)	4.755***	(0.783)
recent treaty	13.303***	(2.225)	10.774***	(1.847)	12.852***	(2.189)	13.264***	(2.230)
nuclear weapons	1.490	(0.412)	1.146	(0.305)	1.487	(0.416)	1.435	(0.445)
recent MID	0.883	(0.143)	0.912	(0.144)	0.880	(0.142)	0.845	(0.159)
arms race	2.209***	(0.175)	2.259***	(0.176)	2.199***	(0.174)	2.210***	(0.175)
wealth	1.992***	(0.061)	1.728***	(0.056)	2.009***	(0.063)	1.994***	(0.061)
satellites			3.146***	(0.431)				
treaty experience					0.802	(0.118)		
US–Russia							1.372	(0.521)
India–Pakistan							1.559	(0.425)
2: High Monitoring								
CE change	0.794	(0.099)	0.765*	(0.096)	0.795	(0.099)	0.793	(0.099)
policy distance	0.807**	(0.054)	0.809**	(0.054)	0.805**	(0.054)	0.798**	(0.055)
CE change X policy	1.404***	(0.099)	1.413***	(0.102)	1.404***	(0.099)	1.409***	(0.099)
balance of power	0.009***	(0.002)	0.013***	(0.004)	0.008***	(0.002)	0.009***	(0.002)
contiguity	3.681***	(0.550)	4.101***	(0.628)	3.677***	(0.552)	3.757***	(0.569)
recent treaty	6.175***	(0.819)	5.341***	(0.693)	5.872***	(0.768)	6.220***	(0.822)
nuclear weapons	2.711**	(0.855)	2.117*	(0.655)	2.698**	(0.864)	2.700**	(0.965)
recent MID	1.602**	(0.238)	1.614**	(0.236)	1.592**	(0.239)	1.605**	(0.266)
arms race	1.612***	(0.123)	1.643***	(0.127)	1.598***	(0.123)	1.612***	(0.123)
wealth	1.735***	(0.055)	1.601***	(0.052)	1.757***	(0.057)	1.732***	(0.055)
satellites			2.254***	(0.260)				
treaty experience					0.701**	(0.095)		
US–Russia							1.354	(0.540)
India–Pakistan							0.420**	(0.138)
N	513024.000		513024.000		513024.000		513024.000	

$*p < 0.05, **p < 0.01, ***p < 0.001$

Table 7: Demonstrations, Riots, and Strikes Volatility (DRS events)
Effects on Agreement Types (multinomial logit)

	M1		M2		M3		M4	
1: Low Monitoring								
DRS events	1.576**	(0.219)	1.598***	(0.222)	1.597***	(0.222)	1.574**	(0.220)
policy distance	0.968	(0.072)	0.973	(0.069)	0.967	(0.072)	0.968	(0.072)
DRS events X policy	1.052	(0.085)	1.041	(0.082)	1.050	(0.084)	1.052	(0.085)
balance of power	0.010***	(0.003)	0.021***	(0.006)	0.009***	(0.002)	0.009***	(0.003)
contiguity	4.599***	(0.752)	5.361***	(0.906)	4.597***	(0.755)	4.621***	(0.768)
recent treaty	13.538***	(2.422)	10.931***	(1.998)	12.760***	(2.319)	13.514***	(2.430)
nuclear weapons	1.364	(0.385)	1.037	(0.284)	1.355	(0.389)	1.325	(0.417)
recent MID	0.922	(0.150)	0.959	(0.151)	0.916	(0.150)	0.885	(0.169)
arms race	2.206***	(0.175)	2.252***	(0.176)	2.193***	(0.174)	2.207***	(0.175)
wealth	1.930***	(0.061)	1.671***	(0.055)	1.957***	(0.063)	1.933***	(0.061)
satellites			3.202***	(0.438)				
treaty experience					0.674**	(0.102)		
US–Russia							1.275	(0.494)
India–Pakistan							1.406	(0.392)
2: High Monitoring								
DRS events	1.048	(0.137)	1.067	(0.142)	1.060	(0.139)	1.057	(0.139)
policy distance	0.762**	(0.064)	0.771**	(0.065)	0.761**	(0.064)	0.760**	(0.064)
DRS events X policy	1.401***	(0.138)	1.377**	(0.137)	1.397***	(0.138)	1.395***	(0.138)
balance of power	0.009***	(0.002)	0.013***	(0.004)	0.008***	(0.002)	0.009***	(0.002)
contiguity	3.742***	(0.549)	4.143***	(0.622)	3.735***	(0.551)	3.761***	(0.566)
recent treaty	5.804***	(0.809)	5.031***	(0.680)	5.504***	(0.757)	5.826***	(0.808)
nuclear weapons	2.329**	(0.716)	1.813*	(0.546)	2.316**	(0.723)	2.388*	(0.832)
recent MID	1.636**	(0.253)	1.656***	(0.249)	1.626**	(0.253)	1.685**	(0.288)
arms race	1.609***	(0.123)	1.639***	(0.126)	1.596***	(0.122)	1.608***	(0.123)
wealth	1.686***	(0.053)	1.559***	(0.049)	1.709***	(0.055)	1.684***	(0.053)
satellites			2.236***	(0.259)				
treaty experience					0.689**	(0.094)		
US–Russia							0.987	(0.411)
India–Pakistan							0.446*	(0.144)
N	508666.000		508666.000		508666.000		508666.000	

$*p < 0.05, **p < 0.01, ***p < 0.001$

fine-grained enough to capture shifts in beliefs that may go on both within rivalries as well as between states not yet engaged in direct competition. For example, the United States and the USSR as well as India and Pakistan were engaged in decades-long rivalries, but during those there were important moments of escalation and detente in the relationship that would be overlooked by a binary rivalry measure. Plots of both the UNGA voting ideal point distance measure for just the US–Russia and India–Pakistan dyads over time track relatively closely with what we know about the historical shifts in these relationships. The Lupu treaty commitment distance measure does not capture known escalations and detentes for these dyads.

The measure in Model 7 returns to UN voting, but uses the earlier version of the distance measure, called Affinity, which calculates the degree of similarity between states based on the maximum metric distance between two states using the method of S distance.[27]. The idea for Affinity as a proxy for beliefs about the opponent's type is the same as with the ideal point distance. States that have similar foreign policy preferences are likely to see one another as having aligned incentives as Cooperative types, while those with distant policy preferences are more likely to see one another as having Competitive incentives. Note that Affinity is measured in reverse from the ideal point distance, with higher numbers as closer Affinity, so the result displayed in the table is actually consistent in direction with the ideal point measure. While Affinity was used in many previous studies to capture policy preference similarity, Bailey, Strezhnev, and Voeten detail how the new ideal point measure improves on Affinity, including by allowing greater consistency of the measure over time.[28]

9 Alternate Measures of Domestic Political Shifts

A seemingly promising alternative measure for domestic political shifts was considered and rejected for this analysis. The source of leadership support (SOLS) measure developed by Brett Ashley Leeds, Michaela Mattes, and Jeremy Vogel codes whether or not a new leader draws support from a different base of domestic support.[29] At first glance, this variable seems like a good proxy for the kinds of domestic political shifts that increase uncertainty. If an incoming leader has a new source of domestic support, this leader might be even more likely to be different from the predecessor and more likely to bring about foreign policy change. These cases may generate greater uncertainty about a state's security approaches.

However, upon closer examination, it is clear that this measure does not capture nearly enough of domestic political situations that are theorized as leading to uncertainty about security cooperation incentives. SOLS captures the most significant kind of leadership change,

[27] Signorino and Ritter, "Tau-b or Not Tau-b: Measuring the Similarity of Foreign Policy Positions."

[28] Bailey, Strezhnev, and Voeten, "Estimating Dynamic State Preferences from United Nations Voting Data."

[29] Brett Ashley Leeds, Michaela Mattes, and Jeremy S Vogel, "Interests, institutions, and the reliability of international commitments," *American Journal of Political Science* 53, no. 2 (2009): 461–476; Michaela Mattes, Brett Ashley Leeds, and Royce Carroll, "Leadership Turnover and Foreign Policy Change: Societal Interests, Domestic Institutions, and Voting in the United Nations," *International Studies Quarterly* 59, no. 2 (2015): 280–290.

Table 8: Alternate Beliefs Measures, CE Change (generalized ordered logit)

	M6		M7		M8	
1: Low Monitoring						
CE change	1.199	(0.127)	1.439***	(0.077)	1.929***	(0.213)
Treaty Distance	0.361***	(0.038)				
CE change X treaty distance	0.993	(0.144)				
Enduring rivalry			0.849	(0.177)		
CE change X rivalry			1.465	(0.325)		
Affinity					1.652***	(0.244)
CE change X affinity					0.494***	(0.087)
	[– control variables not reported–]					
2: High Monitoring						
CE change	0.976	(0.123)	1.319***	(0.085)	2.216***	(0.256)
Treaty Distance	0.469***	(0.062)				
CE change X treaty distance	1.199	(0.200)				
Enduring rivalry			0.510**	(0.119)		
CE change X rivalry			1.616	(0.479)		
Affinity					2.024***	(0.344)
CE change X affinity					0.324***	(0.061)
	[– control variables not reported–]					
N	516053		688027		523096	

$*p < 0.05, **p < 0.01, ***p < 0.001$

Table 9: Alternate beliefs measures, DRS Change (generalized ordered logit)

	M6		M7		M8	
1: Low Monitoring						
DRS events	1.136	(0.124)	1.793***	(0.087)	2.423***	(0.373)
Treaty Distance	0.273***	(0.035)				
DRS events X treaty distance	1.758***	(0.269)				
Enduring rivalry			1.157	(0.236)		
DRS events X enduring rivalry			0.438***	(0.108)		
Affinity					1.987***	(0.397)
DRS events X affinity					0.513**	(0.121)
	[– control variables not reported–]					
2: High Monitoring						
DRS events	0.940	(0.147)	1.703***	(0.124)	2.610***	(0.522)
Treaty Distance	0.328***	(0.056)				
DRS events X treaty distance	2.463***	(0.543)				
Enduring rivalry			1.048	(0.267)		
DRS events X enduring rivalry			0.347***	(0.088)		
Affinity					2.163**	(0.524)
DRS events X affinity					0.391**	(0.116)
	[– control variables not reported–]					
N	513848		631323		517640	

$*p < 0.05, **p < 0.01, ***p < 0.001$

which in democracies is normal changes in power by political party, but in autocracies may essentially be a regime change. As the data user manual states, they "code single-party systems as not experiencing changes in the source of a leader's support for the duration of that party's rule." This means there are no SOLS changes throughout the single-party leadership of the duration of the USSR or North Korea. While big domestic political changes surely raise uncertainty, changes at a less extensive level, particularly in single-party and personalist leader regimes would likewise raise uncertainty in a way that is relevant for my theory, and this would be overlooked by the SOLS measure. Empirically, using the SOLS measure means that there is very little variation in the domestic political shift variable for some states, such as the USSR. However, we know from case study evidence that changes in Soviet leadership were carefully observed by the United States and used to make assessments about Soviet intentions.

10 Socialization and Learning Results

Treaty experience: I test the alternative hypothesis that a state's experience with arms control outside the current dyad has an effect on the probability of cooperation with the current partner. As discussed in the main text, this captures whether socialization into formal treaty-based cooperation approaches, or the setting of best practices standards, has an effect on new cooperation. I create two treaty experience variables. "One treaty experience" is coded 1 if one state in the dyad participated in a treaty with another partner in the last ten years, and 0 otherwise. "Both treaty experience" is coded 1 if both states in the dyad have participated in a treaty with another partner in the last ten years, and 0 otherwise. The time frame is permissive, allowing for any socialization effects to be only slowly adopted.

In order to separate socialization about appropriate behavior or norms of international cooperation from rational learning about the adversary's incentives or capabilities, I create a second variable for past treaty experience with the particular dyad partner. This is essentially an extension of the "ongoing treaty" variable which is already included as control because recent treaties may affect current cooperation prospects for multiple reasons. I create the variable "dyad treaty 10yr" which is coded 1 if the dyad signed an agreement in the last ten years and 0 otherwise. In the models below, I've also included the five year dyadic treaty variable for comparison, which is the same as "ongoing treaty" in the base model. Tables 10 and 11 provide the results.

11 Other Robustness Checks: Rare Events and Time Period Controls

There might be a concern that intrusive monitoring and verification are modern inventions and such tools were not considered as an option in earlier cooperation. Historical evidence on treaty negotiations suggests this is not the case; earlier negotiations considered the prospect of intrusive inspections, sometimes called "supervision" before WWII, even if it was chosen less often. However, to address concerns that there may be differences across time periods with respect to how arms control information provisions were used, I ran a robustness check

Table 10: Socialization and Learning, CE Change Effects

	M10		M11		M12		M13	
1: Low Monitoring								
CE change	0.988	(0.101)	0.991	(0.100)	0.992	(0.102)	0.990	(0.102)
policy distance	0.878*	(0.046)	0.886*	(0.045)	0.880*	(0.046)	0.877*	(0.046)
CE change X policy distance	1.241***	(0.075)	1.237***	(0.073)	1.240***	(0.075)	1.242***	(0.075)
balance of power	0.009***	(0.002)	0.014***	(0.003)	0.010***	(0.002)	0.009***	(0.002)
contiguity	4.052***	(0.536)	3.530***	(0.482)	4.036***	(0.533)	4.050***	(0.539)
nuclear weapons	2.036*	(0.583)	2.315**	(0.652)	1.988*	(0.567)	2.029*	(0.589)
recent MID	1.214	(0.154)	1.199	(0.160)	1.214	(0.153)	1.207	(0.155)
arms race	1.835***	(0.110)	1.880***	(0.114)	1.838***	(0.110)	1.823***	(0.109)
wealth	1.837***	(0.048)	1.741***	(0.051)	1.852***	(0.048)	1.857***	(0.050)
dyad treaty 5yr	8.722***	(1.031)			7.901***	(0.976)	8.359***	(0.976)
dyad treaty 10yr			11.569***	(1.471)				
one treaty experience					0.789**	(0.058)		
both treaty experience							0.745**	(0.074)
2: High Monitoring								
CE change	0.780*	(0.092)	0.797*	(0.092)	0.785*	(0.093)	0.783*	(0.094)
policy distance	0.844**	(0.053)	0.842**	(0.051)	0.846**	(0.053)	0.844**	(0.054)
CE change X policy distance	1.380***	(0.088)	1.368***	(0.084)	1.376***	(0.088)	1.379***	(0.088)
balance of power	0.011***	(0.003)	0.017***	(0.005)	0.012***	(0.003)	0.011***	(0.003)
contiguity	3.366***	(0.490)	2.984***	(0.451)	3.372***	(0.493)	3.371***	(0.494)
nuclear weapons	2.070*	(0.612)	2.399**	(0.671)	2.038*	(0.599)	2.051*	(0.614)
recent MID	1.530**	(0.224)	1.489**	(0.213)	1.532**	(0.224)	1.517**	(0.225)
arms race	1.528***	(0.117)	1.580***	(0.124)	1.518***	(0.119)	1.515***	(0.118)
wealth	1.715***	(0.054)	1.617***	(0.057)	1.722***	(0.056)	1.736***	(0.056)
dyad treaty 5yr	5.465***	(0.682)			5.036***	(0.688)	5.178***	(0.653)
dyad treaty 10yr			8.575***	(1.492)				
one experience					0.833	(0.087)		
both experience							0.712*	(0.096)
N	513024.000		513024.000		513024.000		513024.000	

$*p < 0.05, **p < 0.01, ***p < 0.001$

time, $time^2$ and $time^3$ vars not reported, all models use dyad clustered standard errors, odds ratios reported

Table 11: Socialization and Learning, DRS Change Effects

	M10		M11		M12		M13	
DRS	1.269*	(0.130)	1.237*	(0.123)	1.283*	(0.131)	1.285*	(0.132)
policy distance	0.857*	(0.057)	0.867*	(0.055)	0.859*	(0.057)	0.856*	(0.057)
DRS X policy distance	1.212**	(0.085)	1.200**	(0.081)	1.207**	(0.084)	1.209**	(0.085)
balance of power	0.009***	(0.002)	0.014***	(0.003)	0.010***	(0.002)	0.009***	(0.002)
contiguity	4.043***	(0.526)	3.518***	(0.474)	4.024***	(0.523)	4.040***	(0.530)
nuclear weapons	1.823*	(0.514)	2.072**	(0.575)	1.782*	(0.500)	1.811*	(0.519)
recent MID	1.238	(0.163)	1.225	(0.168)	1.238	(0.161)	1.229	(0.163)
arms race	1.842***	(0.110)	1.889***	(0.115)	1.846***	(0.110)	1.829***	(0.110)
wealth	1.782***	(0.048)	1.695***	(0.050)	1.795***	(0.048)	1.807***	(0.049)
dyad treaty 5yr	8.508***	(1.013)			7.744***	(0.955)	8.056***	(0.954)
dyad treaty 10yr			11.621***	(1.530)				
one treaty experience					0.800**	(0.059)		
both treaty experience							0.682***	(0.068)
2: High Monitoring								
DRS	1.094	(0.139)	1.112	(0.143)	1.089	(0.142)	1.104	(0.141)
policy distance	0.846*	(0.062)	0.844*	(0.059)	0.844*	(0.063)	0.843*	(0.063)
DRS X policy distance	1.277**	(0.106)	1.252**	(0.099)	1.279**	(0.107)	1.276**	(0.106)
balance of power	0.010***	(0.003)	0.015***	(0.004)	0.011***	(0.003)	0.009***	(0.003)
contiguity	3.518***	(0.512)	3.081***	(0.462)	3.526***	(0.517)	3.509***	(0.514)
nuclear weapons	1.763	(0.517)	2.087**	(0.571)	1.743	(0.508)	1.755	(0.522)
recent MID	1.483*	(0.230)	1.466*	(0.220)	1.490*	(0.232)	1.480*	(0.233)
arms race	1.536***	(0.119)	1.599***	(0.127)	1.525***	(0.120)	1.532***	(0.120)
wealth	1.690***	(0.050)	1.598***	(0.053)	1.696***	(0.051)	1.709***	(0.052)
dyad treaty 5yr	5.000***	(0.669)			4.672***	(0.686)	4.742***	(0.631)
dyad treaty 10yr			8.200***	(1.608)				
one treaty experience					0.865	(0.094)		
both treaty experience							0.750	(0.112)
N	508666.000		508666.000		508666.000		508666.000	

$*p < 0.05, **p < 0.01, ***p < 0.001$

time, $time^2$ and $time^3$ vars not reported, all models use dyad clustered standard errors, odds ratios reported

for the monitoring type models which includes a dummy variable for the post–Cold War time period. These models already use relatively modern time frames – since UN voting is a measure, all years are after 1945. The post–Cold War line is the most compelling time period distinction that applies to all states. The results show that including a time period dummy does not dramatically alter the magnitude of the main explanatory variables and they maintain the same level of significance as the base model. For rare events, I ran a rare events logit and Firth method logit for both of the H3 variables. The results are very similar to the regular logits already reported. The differences are so minor that they do not appear in the three-decimal-place view of coefficients or odds ratios, and are not reported here.

Tables 12 and 13 provide the results. All models were run with the full set of control variables; only the main explanatory variables and the time period dummies are reported in the table.

Table 12: Time Periods (logit)

	CE model		CE model w/ FE		DRS model		DRS model w/ FE	
CE change	1.360***	(0.070)	1.216***	(0.055)				
pre 1945	3.969***	(0.362)	1.001	(0.105)	3.471***	(0.321)	1.014	(0.108)
post Cold War	1.223**	(0.085)	2.250***	(0.144)	1.257***	(0.084)	2.812***	(0.198)
DRS					1.703***	(0.083)	1.278***	(0.067)
[– control variables not reported–]								
N	688,027		62,664		631,323		43,857	

$*p < 0.05, **p < 0.01, ***p < 0.001$

Clustered standard errors on the non-fixed effects models; Cold War time period is the excluded category

Table 13: Time Periods (generalized ordered logit)

	base CE model		time period model		base DRS model		time period model	
1: Low Monitoring								
CE change	0.988	(0.101)	0.981	(0.102)				
policy distance	0.878*	(0.046)	0.908	(0.047)	0.857*	(0.057)	0.892	(0.059)
CE change X policy distance	1.241***	(0.075)	1.244***	(0.076)				
post Cold War			1.297***	(0.082)			1.352***	(0.082)
DRS					1.269*	(0.130)	1.311**	(0.133)
DRS X policy distance					1.212**	(0.085)	1.206**	(0.085)
	[– control variables not reported–]							
2: High Monitoring								
CE change	0.780*	(0.092)	0.775*	(0.092)				
policy distance	0.844**	(0.053)	0.892	(0.057)	0.846*	(0.062)	0.904	(0.066)
CE change X policy distance	1.380***	(0.088)	1.383***	(0.089)				
post Cold War			1.445***	(0.153)			1.518***	(0.168)
DRS					1.094	(0.139)	1.120	(0.143)
DRS X policy distance					1.277**	(0.106)	1.270**	(0.104)
	[– control variables not reported–]							
N	513024.000		513024.000		508666.000		508666.000	

$*p < 0.05, **p < 0.01, ***p < 0.001$

Online Appendix Part III

Data on arms control agreements used in Jane Vaynman, *Enemies in Agreement: Political Volatility and the Design of Arms Control.* Cambridge: Cambridge University Press, 2026.[1]

Table 1: Adversarial Agreements Dataset

Code	Year	Treaty Name	Information Provisions	
			general level	specific type
1817.1	1817	Rush-Bagot Agreement	low	none
1824.1	1824	Convention Between the United States of America and His Majesty the Emperor of All the Russias	low	none
1829.1	1829	Treaty of Adrianople	low	none
1829.2	1829	Treaty of Peace Between Peru and Colombia at Guayaquil	low	none
1841.1	1841	Limitation of Egyptian Arms by Turkey	low	none
1841.2	1841	London Straits Convention	low	none
1850.1	1850	The Clayton-Bulwer Treaty	low	none
1856.1	1856	General Treaty for the Re-Establishment of Peace (Treaty of Paris)	high	observation
1863.1	1863	Treaty of Peace Between Colombia and Ecuador at Pinsaqui	low	none
1868.1	1868	Declaration of St. Petersburg	low	none
1871.1	1871	Armistice for Franco-Prussian War	low	reporting
1881.1	1881	Treaty Between Argentine and Chile-Straits of Magellan	low	none
1885.1	1885	General Act of the Conference of Berlin	low	none
1888.1	1888	Constantinople Convention of the Suez Canal	high	verification
1899.1	1899	Hague Declaration (IV,2) Concerning Asphyxiating Gases	low	none
1899.2	1899	Hague Declaration (IV,3) Concerning Expanding Bullets	low	none
1902.1	1902	Convention Between Chile and the Argentine Republic on the Limitation of Naval Armaments	low	reporting
1905.1	1905	Karlstad Convention - Final Separation of Sweden and Norway	high	observation
1905.2	1905	The Treaty of Portsmouth	low	none
1907.1	1907	Hague Convention (IV) Respecting the Laws and Customs of War on Land	low	none
1907.2	1907	Hague Convention (IX) Concerning Bombardment by Naval Forces in Time of War	low	none
1907.3	1907	Hague Convention (V) Respecting the Rights and Duties of Neutral Powers and Persons in Case of War on Land	low	none
1907.4	1907	Hague Convention (VIII) Relative to the Laying of Automatic Submarine Contact Mines	low	none

[1]This list excludes three treaties signed prior to 1816: 1807 Tilsit Treaty, 1808 Franco-Prussian Treaty, and 1809 Peace Treaty between France and Austria. Code refers to the unique identifier for each treaty used in the data and replication files, available at Harvard Dataverse https://doi.org/10.7910/DVN/EUFGGL.

Code	Year	Treaty Name	Information Provisions	
			general level	specific type
1907.5	1907	Hague Convention (XIII) Concerning the Rights and Duties of Neutral Powers in Naval War	low	none
1918.1	1918	Treaty of Brest-Litovsk	low	none
1919.1	1919	Convention for the Control of the Trade in Arms and Ammunition	high	verification
1919.2	1919	Covenant of the League of Nations	low	none
1919.3	1919	Treaty of Versailles	high	verification
1919.4	1919	Treaty of Neuilly	high	verification
1919.5	1919	Treaty of Saint-Germain-en-Laye - the Allied and Associated Powers and Austria	high	verification
1920.1	1920	Peace Treaty of Sèvres	high	verification
1920.2	1920	Treaty Concerning the Archipelago of Spitsbergen (Spitsbergen Treaty)	low	none
1920.3	1920	Treaty of Peace Between Russia and Estonia	high	verification
1920.4	1920	Treaty of Trianon	high	verification
1921.1	1921	Convention Relating to the Non-Fortification and Neutralisation of the Åland Islands	low	none
1921.2	1921	Peace Treaty of Riga	low	none
1921.3	1921	Treaty and Supplementary Declaration Between Four Powers	low	none
1922.2	1922	Treaty Between All Nine Powers in Matters Concerning China	low	none
1922.3	1922	Washington Conference on the Limitation of Armament	low	reporting
1922.4	1922	Washington Treaty in Relation to the Use of Submarines and Noxious Gases in Warfare	low	none
1923.1	1923	Convention for the Limitation of Armaments Adopted at the Conference on Central American Affairs	low	reporting
1923.2	1923	Gondra Treaty-Treaty to Avoid or Prevent Conflicts Between the American States	low	reporting
1923.3	1923	The Convention Relating to the Regime of the Straits and Turkey	high	observation
1925.1	1925	Protocol for the Prohibition of the Use in War of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare (1925 Geneva Protocol)	low	none
1928.1	1928	Pact of Paris for the Renunciation of War as an Instrument of National Policy (Kellogg–Briand Pact)	low	none
1930.1	1930	International Treaty for the Limitation and Reduction of Naval Armament (London Naval Treaty)	low	reporting
1931.1	1931	Turkey-USSR-Naval Protocol	low	reporting
1933.1	1933	Anti-War Treaty of Non-Aggression and Conciliation	low	reporting
1934.1	1934	The Treaty of Taif	low	none
1935.1	1935	Anglo-German Naval Agreement	low	reporting
1936.1	1936	Convention Regarding the Regime of the Straits (Montreux Convention)	low	reporting
1936.2	1936	London Conference-Treaty for the Limitation of Naval Armament	low	reporting
1945.2	1945	Charter of the United Nations	low	reporting

Code	Year	Treaty Name	Information Provisions	
			general level	specific type
1947.1	1947	Treaty of Peace With Bulgaria	low	reporting
1947.2	1947	Treaty of Peace With Finland	low	reporting
1947.3	1947	Treaty of Peace With Hungary	low	reporting
1947.4	1947	Treaty of Peace With Italy	high	verification
1947.5	1947	Treaty of Peace With Romania	low	reporting
1949.1	1949	Geneva Convention (IV) Relative to the Protection of Civilian Persons in Time of War	low	none
1949.2	1949	Israel-Egypt Armistice Agreement	high	verification
1949.3	1949	Israel-Jordan Armistice Agreement	high	observation
1949.4	1949	Israel-Lebanon Armistice Agreement	high	observation
1949.5	1949	Israel-Syria Armistice Agreement	high	observation
1953.1	1953	Korean War Armistice Agreement	high	verification
1954.1	1954	Convention for the Protection of Cultural Property in the Event of Armed Conflict (Cultural Property Protection Convention)	low	none
1954.2	1954	First Indochina War Demilitarized Zone-Geneva Conference	high	verification
1955.1	1955	Treaty for the Re-Establishment of an Independent and Democratic Austria (Austrian State Treaty)	low	reporting
1956.1	1956	Statute of the International Atomic Energy Agency	high	verification
1959.1	1959	Antarctic Treaty	high	observation
1963.1	1963	Memorandum of Understanding Between the USA and the USSR Regarding the Establishment of a Direct Communications Link (Hot Line Agreement)	low	reporting
1963.2	1963	Treaty Banning Nuclear Weapon Tests in the Atmosphere, in Outer Space and Under Water (Partial Test Ban Treaty, PTBT)	low	none
1964.1	1964	Unilateral Statements by the USA, the USSR and the UK Regarding the Reduction of Fissionable Materials Production	low	none
1966.1	1966	Tashkent Declaration (India-Pakistan)	low	none
1967.1	1967	Treaty for the Prohibition of Nuclear Weapons in Latin America and the Caribbean (Treaty of Tlatelolco), as Amended	high	verification
1967.2	1967	Treaty on Principles Governing the Activities of States, in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies (Outer Space Treaty)	high	observation
1968.1	1968	Treaty on the Non-Proliferation of Nuclear Weapons (NPT)	high	verification
1968.2	1968	United Nations Security Council Resolution 255 on Security Assurances to Non-Nuclear-Weapon States	high	observation
1971.1	1971	Agreement Between the USA and the USSR on Measures to Reduce the Risk of Outbreak of Nuclear War (US-Soviet Nuclear Accidents Agreement)	low	reporting
1971.3	1971	Treaty on the Prohibition of the Emplacement of Nuclear Weapons and Other Weapons of Mass Destruction on the Seabed and the Ocean Floor and in the Subsoil Thereof (Seabed Treaty)	high	observation

Code	Year	Treaty Name	Information Provisions	
			general level	specific type
1972.1	1972	Agreement Between the USA and the USSR on the Prevention of Incidents on and Over the High Seas (Incidents at Sea Agreement, INCSEA)	low	reporting
1972.2	1972	Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on Their Destruction (Biological Weapons Convention)	low	reporting
1972.3	1972	India-Pakistan Simla Agreement	low	reporting
1972.4	1972	Interim Agreement Between the USA and the USSR on Certain Measures With Respect to the Limitation of Strategic Offensive Arms (SALT I Agreement)	high	observation
1972.5	1972	Treaty Between the USA and the USSR on the Limitation of Anti-Ballistic Missile Systems (ABM Treaty)	high	observation
1973.1	1973	Agreement Between the USA and the USSR on the Prevention of Nuclear War	low	reporting
1973.2	1973	Agreement on Ending the War and Restoring Peace in Vietnam	high	verification
1973.3	1973	Protocol to the 1972 Incidents at Sea Agreement	low	none
1974.1	1974	Agreement on Disengagement of Forces Between Israel and Syria	high	verification
1974.2	1974	Declaration of Ayacucho	low	none
1974.3	1974	Israel-Egypt Separation of Forces Agreement	high	verification
1974.4	1974	Joint Statement by the USA and the USSR on Strategic Offensive Arms (Vladivostok Accord)	low	reporting
1974.5	1974	Protocol to the 1972 ABM Treaty	low	reporting
1974.6	1974	Treaty Between the USA and the USSR on the Limitation of Underground Nuclear Weapon Tests (Threshold Test Ban Treaty, TTBT)	high	observation
1974.7	1974	United Nations General Assembly Resolution 3314 (XXIX) on the Definition of Aggression	high	observation
1975.1	1975	Convention on Registration of Objects Launched Into Outer Space (Registration Convention)	low	reporting
1975.2	1975	Document on Confidence-Building Measures and Certain Aspects of Security and Disarmament, Included in the Final Act of the Conference on Security and Co-Operation in Europe (Helsinki CBM Document)	low	reporting
1975.3	1975	Interim Agreement Between Israel and Egypt	high	verification
1975.4	1975	United Nations General Assembly Resolution 3472 B (XXX), Declaration on Nuclear-Weapon-Free Zones	high	observation
1976.1	1976	Agreement Between France and the USSR on the Prevention of the Accidental or Unauthorized Use of Nuclear Weapons	low	reporting
1976.3	1976	Treaty Between the USA and the USSR on Underground Nuclear Explosions for Peaceful Purposes (Peaceful Nuclear Explosions Treaty, PNET)	high	verification
1977.1	1977	Agreement Between the UK and the USSR on the Prevention of Accidental Nuclear War	low	reporting

Code	Year	Treaty Name	Information Provisions	
			general level	specific type
1977.11	1977	Convention of the OAU for the Elimination of Mercenarism in Africa	low	reporting
1977.2	1977	Convention on the Prohibition of Military or Any Other Hostile Use of Environmental Modification Techniques (ENMOD Convention)	low	reporting
1977.3	1977	Protocol (I) Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (1977 Protocol I)	low	reporting
1979.1	1979	Agreement Governing the Activities of States on the Moon and Other Celestial Bodies (Moon Agreement)	high	observation
1979.2	1979	Peace Treaty Between Israel and Egypt	high	verification
1979.3	1979	Soviet Statement on the Backfire Bomber	low	none
1979.4	1979	Treaty Between the USA and the USSR on the Limitation of Strategic Offensive Arms (SALT II Treaty)	high	observation
1980.1	1980	Convention on the Physical Protection of Nuclear Material	low	reporting
1981.1	1981	Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects (CCW, or Inhumane Weapons Convention)	low	reporting
1982.1	1982	United Nations Convention on the Law of the Sea (UNCLOS)	low	reporting
1983.1	1983	Agreement Between Israel and Lebanon	high	verification
1985.1	1985	South Pacific Nuclear Free Zone Treaty (Treaty of Rarotonga)	high	verification
1986.1	1986	Document of Stockholm Conference	high	verification
1986.2	1986	USSR and United Kingdom Agreement on INCSEA	low	reporting
1987.1	1987	Agreement Between the USA and the USSR on the Establishment of Nuclear Risk Reduction Centers	low	reporting
1988.1	1988	Agreement Between Pakistan and India on the Prohibition of Attack Against Nuclear Installations and Facilities	low	reporting
1988.2	1988	Agreement Between the USA and the USSR on Notifications of Launches of Intercontinental Ballistic Missiles and Submarine-Launched Ballistic Missiles	low	reporting
1988.4	1988	Treaty Between the USA and the USSR on the Elimination of Their Intermediate-Range and Shorter-Range Missiles (INF Treaty)	high	verification
1988.5	1988	USSR and Federal Republic of Germany Agreement on INCSEA	low	reporting
1989.1	1989	Agreement Between the USA and the USSR on Reciprocal Advance Notification of Major Strategic Exercises	low	reporting
1989.2	1989	Agreement Between the USA and the USSR on the Prevention of Dangerous Military Activities (DMA Agreement)	low	reporting

Code	Year	Treaty Name	Information Provisions	
			general level	specific type
1989.3	1989	International Convention Against the Recruitment, Use, Financing and Training of Mercenaries	low	reporting
1989.4	1989	USSR and Canada Agreement on INCSEA	low	reporting
1989.5	1989	USSR and France Agreement on INCSEA	low	reporting
1989.6	1989	USSR and Italy Agreement on INCSEA	low	reporting
1990.1	1990	Agreement Between the USA and the USSR on Destruction and Non-Production of Chemical Weapons and on Measures to Facilitate the Multilateral Convention on Banning Chemical Weapons (US-Soviet Chemical Weapons Agreement)	high	verification
1990.2	1990	Charter of Paris for a New Europe	high	observation
1990.3	1990	Federal Republic of Germany and Poland Agreement on INCSEA	low	reporting
1990.4	1990	Treaty on Conventional Armed Forces in Europe (CFE Treaty), as Amended by the 1999 Agreement on Adaptation	high	verification
1990.5	1990	USSR and Netherlands Agreement on INCSEA	low	reporting
1990.6	1990	USSR and Spain Agreement on INCSEA	low	reporting
1990.7	1990	Protocol to TTBT	high	verification
1990.8	1990	Vienna Document	high	verification
1991.1	1991	Agreement Between India and Pakistan on Prevention of Air Space Violations and for Permitting Over Flights and Landings by Military Aircraft	low	reporting
1991.11	1991	United Nations Security Council Resolution 687 Imposing Arms Restrictions on Iraq and Establishing the UN Special Commission on Iraq (UNSCOM)	high	verification
1991.12	1991	United Nations Security Council Resolution 715 on the Monitoring of Iraqi Compliance With Arms Restrictions	high	verification
1991.14	1991	USSR and Canada Agreement on the Prevention of Dangerous Military Activities	low	reporting
1991.15	1991	USSR and Greece Agreement on INCSEA	low	reporting
1991.16	1991	Agreement Between Russia and Greece on Prevention of Dangerous Military Activities	low	reporting
1991.17	1991	Agreement Between Romania and Hungary on the Establishment of an Open Skies Regime	high	verification
1991.2	1991	Agreement Between India and Pakistan on the Advance Notice of Military Exercises	low	reporting
1991.3	1991	Cartagena Declaration on Renunciation of Weapons of Mass Destruction	low	none
1991.4	1991	Guidelines for Conventional Arms Transfers Agreed by the Permanent Members of the UN Security Council	low	none
1991.5	1991	Gulf War Ceasefire	high	verification
1991.6	1991	Protocol on Environmental Protection to the 1959 Antarctic Treaty (Madrid Protocol)	high	observation
1991.7	1991	Soviet President's Announcement Regarding Unilateral Reductions of Nuclear Weapons	low	none

Code	Year	Treaty Name	Information Provisions	
			general level	specific type
1991.8	1991	Treaty Between the USA and the USSR on the Reduction and Limitation of Strategic Offensive Arms (START I Treaty)	high	verification
1991.9	1991	UN General Assembly Resolution 46/36 on Transparency in Armaments	high	observation
1992.1	1992	1992 Helsinki Document of the Conference on Security and Co-Operation in Europe	low	reporting
1992.11	1992	Tashkent Document	high	verification
1992.12	1992	Treaty on Open Skies	high	observation
1992.13	1992	Memorandum Between Russian Federation and China on Mutual Reduction of Troops in the Border Area	low	reporting
1992.2	1992	Concluding Act of the Negotiation on Personnel Strength of Conventional Armed Forces in Europe (CFE-1A Agreement)	high	verification
1992.4	1992	Guidelines for Sensitive Missile-Relevant Transfers (Missile Technology Control Regime, MTCR), (Revised 1987 Guidelines)	low	none
1992.5	1992	Joint Declaration by Pakistan and India on the Complete Prohibition of Chemical Weapons	low	none
1992.6	1992	Joint Declaration by South and North Korea on the Denuclearization of the Korean Peninsula	high	verification
1992.9	1992	Protocol to the 1991 START I Treaty (Lisbon Protocol)	high	verification
1993.1	1993	Agreement on the Maintenance of Peace Along the Line of Actual Control in the India-China Border	high	verification
1993.2	1993	Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on Their Destruction (Chemical Weapons Convention)	high	verification
1993.3	1993	Treaty Between the USA and Russia on Further Reduction and Limitation of Strategic Offensive Arms (START II Treaty)	high	verification
1994.1	1994	Agreed Framework Between the USA and North Korea	high	verification
1994.2	1994	Agreement Between the USA and Russia to De-Target Strategic Nuclear Missiles, Contained in the Moscow Declaration of the US and Russian Presidents	low	none
1994.4	1994	Code of Conduct on Politico-Military Aspects of Security, Included in the Conference on Security and Co-Operation in Europe Budapest Decisions	low	none
1994.5	1994	Joint Declaration by the UK and Russia on the De-Targeting of Nuclear Missiles	low	none
1994.6	1994	Russian Federation and Republic of Korea Agreement on INCSEA	low	reporting
1994.7	1994	Agreement Between Russia and China on Prevention of Dangerous Military Activities	low	reporting
1995.1	1995	Declaration of Santiago	low	reporting

Code	Year	Treaty Name	Information Provisions	
			general level	specific type
1995.3	1995	NPT Review and Extension Conference Decision on Principles and Objectives for Nuclear Non-Proliferation and Disarmament	high	observation
1995.6	1995	Protocol IV to the 1981 Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects (Protocol on Blinding Laser Weapons)	low	none
1995.7	1995	The Dayton Peace Accords	high	verification
1995.8	1995	Treaty on the Southeast Asia Nuclear Weapon-Free Zone (Treaty of Bangkok)	high	verification
1996.1	1996	OSCE Lisbon Document	high	observation
1996.11	1996	Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies	low	reporting
1996.12	1996	Agreement Between Russia and China on Establishing a Direct Secret Phone Line	low	reporting
1996.3	1996	Agreement Between India and China on Confidence-Building Measures in the Military Field Along the Line of Actual Control in the India-China Border Areas	low	reporting
1996.4	1996	Agreement Between Russia, Kazakhstan, Kyrgyzstan, Tajikistan (as a Joint Party) and China on Confidence Building in the Military Field in the Border Area Agreement	high	observation
1996.5	1996	Agreement on Sub-Regional Arms Control Concerning Yugoslavia (Serbia and Montenegro), Bosnia and Herzegovina, and Croatia (Florence Agreement)	high	verification
1996.6	1996	Amended Protocol II on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices to the 1981 Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious	high	observation
1996.7	1996	Comprehensive Nuclear Test-Ban Treaty (CTBT)	high	verification
1996.8	1996	Israel-Lebanon Ceasefire Understanding	high	observation
1996.9	1996	Treaty on the African Nuclear-Weapon-Free Zone (Treaty of Pelindaba)	high	verification
1997.1	1997	Agreed Statements and Common Understandings Between Russia, the USA, Belarus, Kazakhstan and Ukraine Relating to the 1972 ABM Treaty	low	reporting
1997.13	1997	NATO-Russia Founding Act on Mutual Relations, Cooperation and Security	low	reporting
1997.14	1997	Protocol to the 1993 START II Treaty	low	reporting
1997.2	1997	Agreement Between Russia, Kazakhstan, Kyrgyzstan and Tajikistan (as a Joint Party) and China on the Mutual Reduction of Armed Forces in the Border Area	high	verification

Code	Year	Treaty Name	Information Provisions	
			general level	specific type
1997.3	1997	Agreement Between Russia, the USA, Belarus, Kazakhstan and Ukraine on Confidence-Building Measures Related to Systems to Counter Ballistic Missiles Other Than Strategic Ballistic Missiles	low	reporting
1997.4	1997	Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on Their Destruction (APM Convention)	high	verification
1997.7	1997	Inter-American Convention Against the Illicit Manufacturing of and Trafficking in Firearms, Ammunition, Explosives, and Other Related Materials	low	reporting
1997.9	1997	Joint Statement by the USA and Russia on Parameters on Future Reductions in Nuclear Forces	low	none
1998.1	1998	Declaration of a Moratorium on the Importation, Exportation and Manufacture of Light Weapons in West Africa (ECOWAS Moratorium)	low	reporting
1998.2	1998	European Union Code of Conduct on Arms Exports	low	reporting
1998.3	1998	European Union Joint Action Concerning the Spread of Small Arms and Light Weapons	low	reporting
1998.4	1998	Joint Statement by the USA and Russia on the Exchange of Information on Missile Launches and Early Warning	low	reporting
1998.5	1998	Political Declaration of Mercosur, Bolivia and Chile as Zone of Peace	low	none
1998.7	1998	United Nations Security Council Resolution 1209 on Illicit Arms Flow	low	reporting
1998.8	1998	US-China MMCA Agreement	low	none
1999.1	1999	Code of Conduct for the Implementation of the 1998 ECOWAS Moratorium	high	verification
1999.2	1999	Inter-American Convention on Transparency in Conventional Weapons Acquisitions	low	reporting
1999.3	1999	Memorandum of Understanding Between India and Pakistan Relating to Confidence-Building Measures	high	observation
1999.7	1999	United Nations Security Council Resolution 1284 Establishing the UN Monitoring, Verification and Inspection Commission (UNMOVIC) on Iraq	high	verification
1999.8	1999	Vienna Document of the Negotiations on Confidence- and Security-Building Measures (1999 Vienna CSBM Document)	high	verification
2000.1	2000	Agreement Between the USA and Russia Concerning the Management and Disposition of Plutonium Designated as No Longer Required for Defense Purposes and Related Cooperation (Plutonium Management and Disposition Agreement, PMDA)	low	reporting
2000.2	2000	Joint Statement by the USA and Russia on the Strategic Stability Cooperation Initiative	high	observation
2000.3	2000	Memorandum of Agreement Between the USA and Russia on the Establishment of a Joint Center for the Exchange of Data From Early Warning Systems and Notifications of Missile Launches	high	observation

Code	Year	Treaty Name	Information Provisions	
			general level	specific type
2000.4	2000	Nairobi Protocol	high	observation
2000.5	2000	Nuclear Suppliers Group Guidelines for Nuclear Transfers (Guidelines for Nuclear Transfers, Revised 1977 London Guidelines)	low	reporting
2000.6	2000	Nuclear Suppliers Group Guidelines for Transfers of Nuclear-Related Dual-Use Equipment, Materials, Software, and Related Technology (nuclear Dual-Use Guidelines, Revised 1992 Warsaw Guidelines)	low	reporting
2000.7	2000	OSCE Document on Small Arms and Light Weapons	high	observation
2000.8	2000	Russian Federation and Portugal Agreement on INCSEA	low	reporting
2001.1	2001	Protocol Against the Illicit Manufacturing of and Trafficking in Firearms, Their Parts and Components and Ammunition (Firearms Protocol), Supplementing the 15 November 2000 United Nations Convention Against Transnational Organized Crime	low	reporting
2001.2	2001	Protocol on the Control of Firearms, Ammunition and Other Related Materials in the Southern African Development Community (SADC) Region	high	observation
2001.3	2001	United Nations Programme of Action to Prevent, Combat and Eradicate the Illicit Trade in Small Arms and Light Weapons in All Its Aspects	low	reporting
2001.4	2001	Agreement Between Russia and Czech Republic on Prevention of Dangerous Military Activities	low	reporting
2002.1	2002	Declaration by Heads of State and Government of NATO Member States and the Russian Federation on NATO-Russia Relations	high	observation
2002.2	2002	Joint Declaration on the New Strategic Relationship by the USA and Russia	high	observation
2002.3	2002	Treaty Between the USA and Russia on Strategic Offensive Reductions (Moscow Treaty, SORT)	low	reporting
2002.4	2002	Kazakhstan and China Agreement on Prevention of Dangerous Military Activities	low	reporting
2005.1	2005	Agreement Between India and Pakistan on Pre-Notification of Flight Testing of Ballistic Missiles	low	reporting
2006.1	2006	Central Asian Nuclear-Weapon-Free Zone (CANWFZ)	low	reporting
2006.2	2006	Ecowas Convention on Small Arms and Light Weapons, Their Ammunition and Other Related Materials	high	observation
2007.1	2007	Indo-Pakistan Agreement on Reducing Risk From Accidents Relating to Nuclear Weapons	low	reporting
2008.1	2008	Convention on Cluster Munitions	low	reporting
2009.1	2009	Russia-China Ballistic Missiles Launch Notification	low	reporting
2010.1	2010	New START Treaty	high	verification