Loosening the Ties that Bind: A Learning Model of Agreement Flexibility

Barbara Koremenos

The tyranny of an unamendable constitution might lead to revolution; the forced continuance of the legal status quo as between states might make it more difficult to preserve peace.[1]

We all know that future generations will have to live, even on a strictly technological level, in a setting very different from the present one. To imprison them in an iron corset, which could not be adjusted to the changing conditions of history, would in our opinion expose that corset to the danger of bursting. We therefore prefer a steel corset which, being more flexible, could more effectively assure the continuation of the treaty.

—Statement issued by Italian representative Roberto Caracciolo regarding the proposed Nuclear Non-Proliferation Treaty.[2]

How can the corsets of international agreements be made more flexible and therefore more robust? Or, in the words of Jon Elster, how can Ulysses be loosely bound to the mast?[3] This is an important question for the simple reason that international cooperation is important, and international cooperation is often institutionalized in the form of international agreements. Existing international agreements are testament to states' willingness and ability to cooperate despite the international anarchy in which they find themselves. Given the difficulties of cooperation under anarchy documented in the recent international relations literature, understanding how states manage to bring about the formal cooperation embodied in international agreements is of both theoretical and practical interest.

I thank Jeffrey Smith, James Fearon, Andrew Kydd, James Morrow, Charles Glaser, Duncan Snidal, George Bunn, Richard Bilder, Brian Portnoy, Scott Mosier, T. Clifton Morgan, Alan Stam, Jack Child, and workshop participants at the University of Chicago (PIPES), the Brookings Institution, and the Merriam Lab of the University of Illinois, Champagne-Urbana. The comments of editors Peter Gourevitch and David Lake and those of three anonymous referees greatly improved this article.

1. Wilson 1934.
2. U.S. Arms Control and Disarmament Agency 1967.
3. Elster 1989, 49.

International Organization 55, 2, Spring 2001, pp. 289–325
© 2001 by The IO Foundation and the Massachusetts Institute of Technology

https://doi.org/10.1162/002081801511405586 Published online by Cambridge University Press

States can make agreements more desirable in prospect and more robust in practice by varying their provisions for duration and renegotiation. These provisions help states account for the uncertain economic, political, and technological contexts in which agreements are made and (ideally) kept.

If we look around, we find that international agreements differ widely in their duration provisions. The Plaza Agreement on foreign exchange intervention, for example, covered a period of just six weeks. At the other extreme, the Universal Postal Union is of indefinite duration and has already lasted over one hundred years. Renegotiation provisions vary widely as well. The International Coffee Agreement was renegotiated about every six years, whereas the Antarctic Treaty had no renegotiation provisions for the first thirty years and renegotiation at the request of any party after that. Importantly, the variation across agreements in these provisions is not random; rather, as I argue here, it results from the considered choices of the parties to each agreement, in light of their preferences and the unique aspects of the context of each agreement.

Nevertheless, the issues of duration and renegotiation have been almost completely ignored in the political science literature on international relations. They have been wholly neglected in theoretical studies of international cooperation, and there has been surprisingly little discussion of these issues from an empirical point of view. In fact, although some discussions of individual agreements cover the issues of duration and renegotiation for the agreement in question, no work exists that attempts to account for or even describe the observed patterns and variation in agreement duration and renegotiation.[4]

While political scientists have ignored such issues, scholars of international law have not. In fact, in the decades prior to World War II, a few studies were published about revision clauses and other aspects of treaty design. These studies are purely descriptive, however, with no theoretical basis whatsoever.[5]

What I am attempting essentially is to bring theory—in particular, formal international relations theory—to international law. I argue that uncertainty in the international environment—uncertainty that is of varying forms and degrees across

4. One exception is Bilder, who surveys a number of techniques that states can use to help themselves manage the risks of international agreements. Bilder 1981. He identifies and devotes two pages to the technique of "Limiting the Duration of the Agreement." Also, Grieco mentions issues of agreement durability in his concluding chapter (see fn. 28 below). Grieco 1990.

5. Notwithstanding, it might be argued that the international law doctrine of *rebus sic stantibus* (treaties cease to be binding when basic conditions change) addresses precisely the issue I am raising in this analysis. However, as international law scholars are quick to point out, states rarely, if ever, invoke that doctrine because it is considered to be very controversial. Writing about revision clauses in treaties, Wilson argues that they are important because they allow states to withdraw from burdensome obligations "without invoking the *rebus sic stantibus*—or as it has sometimes been called, the 'notorious'—clause." Wilson 1934. Similarly, Kunz writes that "It is particularly important to insist on the complete difference, from a juridical point of view, between the *clausula rebus sic stantibus* and the problem of revision." Kunz 1939, 41. More recently, Jackson states that "The traditional doctrines, such as *rebus sic stantibus* . . . do not begin to accommodate the type of evolution, innovation, and step-by-step change of circumstances that must be addressed by both national and international 'legislation.'" Jackson 1992, 339.

https://doi.org/10.1162/002081801511405 86 Published online by Cambridge University Press

issue contexts—leads states to choose particular duration and renegotiation provisions. These provisions, in turn, affect whether or not states conclude international agreements and whether or not they renege on them.[6]

This study is the first theoretical work to address the nominal (that is, negotiated) length of international agreements.[7] My theoretical work takes as its departure the economic theory of contracts, since agreements are essentially contracts between states, with the key difference that there is no external authority available to enforce them. The economic literature on contract duration formalizes the key trade-off in choosing contract duration between the benefit associated with spreading the (assumed) fixed cost of contracting over additional periods and the loss associated with staying for additional periods in a suboptimal contract.[8] This basic insight is helpful but insufficient to explain the range of duration and renegotiation provisions present in even the small set of agreements I consider here.

In particular, the economic model of contract duration cannot explain why some agreements, such as the Nuclear Non-Proliferation Treaty, are arranged so that after an initial trial period, the agreement may be extended indefinitely. To account for these important cases, I develop an alternative model in which the parties to an agreement learn over time about its costs and benefits to them. This learning reduces their uncertainty about the agreement's effects, with the result that they may eventually become sure enough of these effects to be willing to extend the agreement indefinitely.

My theory identifies key factors affecting the choices of duration and renegotiation provisions in these agreements. The two most important factors are the degree of agreement uncertainty (formally, the variance of the distribution of gains from an agreement) and the degree of noise in the environment (formally, the variance of confounding variables whose effect on outcomes may be confused with that of an agreement). The greater the agreement uncertainty, the more likely states will want to limit the duration of the agreement and incorporate renegotiation. The factor working against renegotiation is noise. The greater the noise, the more difficult it is to learn how an agreement is actually working; hence incorporating limited duration and renegotiation provisions becomes less valuable.

The model provides a framework within which to discuss the Nuclear Non-Proliferation Treaty (NPT). I describe the basic features of the agreement and summarize the available information about how the parties themselves framed the duration and renegotiation issues while negotiating the agreement. I then briefly

6. Moreover, I would argue that renegotiation clauses (as well as other forms of flexibility provisions like escape clauses) help reconcile the tension between two doctrines of international law: *rebus sic stantibus* and *pacta sunt servanda* (treaties should be performed in good faith).

7. Two recent studies by Gaubatz and Bennett examine realized durations of alliances using hazard-rate models. See Gaubatz 1996; and Bennett 1997. These studies ignore the fact that many agreements are initially concluded with a finite duration. They therefore conflate planned agreement terminations with those resulting from a violation of the agreement.

8. The three main papers are Gray 1978; Dye 1985; and Harris and Holmstrom 1987.

https://doi.org/10.1162/002081801511140586 Published online by Cambridge University Press

compare the duration and renegotiation provisions of the NPT to those of two other agreements: the Antarctic Treaty and the Outer Space Treaty.

The remainder of this article is organized as follows. In the second section I develop a formal model in which an agreement characterized by uncertainty may be renegotiated to incorporate new information. The uncertainty is related to the division of gains under the agreement, with the parties resolving this uncertainty over time as they gain experience with the agreement. This form of uncertainty corresponds to the uncertainty experienced by the parties to the international agreement I discuss in detail in the third section. In the fourth section I present other examples of optimally chosen duration and renegotiation provisions. In the fifth section I offer conclusions.

Model: Learning About the Workings of an Agreement

In this section I develop a model of the optimal duration and renegotiation provisions of international agreements. My model is based on the idea that the distributional consequences of a new agreement may not be known with certainty at the time the parties first negotiate the agreement. For example, if an agreement establishes a new institution or a new pattern of trade, the effect of this innovation on the gross national product (GNP) or on the security of each of the parties to the agreement may be known only approximately in advance.[9]

In this model an agreement is like an experience good in which a complete knowledge of the effects of the good is gained only by using it. Over time, by observing the outcomes obtained under the agreement and attempting to distinguish the once-and-for-all effects of the agreement from the normal period-to-period fluctuation in outcomes, the parties come to learn with increasing precision the true distribution of benefits created by the agreement.[10]

9. Young identifies this type of uncertainty when he states that "none of the stakeholders will be able to make confident predictions at the outset regarding the incidence of the benefits likely to flow from alternative institutional arrangements." Young 1991, 288–89. He does not, however, consider the possibility that such uncertainty might make risk-averse states nervous about having their hands tied.

10. There are other forms of uncertainty, and, importantly, states respond to these with alternative flexibility provisions. For example, in the context of economics agreements, the uncertainty is often persistent; states cannot learn once and for all about the division of gains from an agreement because that division is subject to repeated shocks. In such a case, states may follow the example of the G5 finance ministers and have a *series* of finite-duration agreements with renegotiation in between. Koremenos 1999. In a recent study I consider the same environment but with a large number of parties. Koremenos 2000. In such a case renegotiation costs may be prohibitive (since they rise with the number of parties). Hence parties may choose to establish an institution (like the IMF) empowered to adjust the terms of the agreement in response to the repeated shocks. Bordo and Kydland's work on escape clauses during the gold standard addresses a different type of uncertainty: political shocks, such as wars and banking panics. Bordo and Kydland 1995 and 1996. Such uncertainty leads states to incorporate "suspension" mechanisms instead of "adjustment" mechanisms. Downs and Rocke and

https://doi.org/10.1162/002081801511405 86 Published online by Cambridge University Press

How can states make use of what they learn about the distribution of gains from the agreement? They may choose to make their initial agreement of finite duration, and then "re"negotiate a new agreement when the first one comes to a close. If they do so, they can use the information they have gained through their experience under the agreement to realign the division of gains in the renegotiated agreement. Under certain conditions (made precise below) this planned renegotiation and realignment reduces *ex ante* uncertainty and thereby raises the expected utility of the parties. Put another way, careful selection of the duration and renegotiation provisions allows the parties to conclude efficient agreements (ones that increase the size of the pie available to the parties) that otherwise might fail because of distributional problems.

Of course, states may also be uncertain about the absolute level of gains from a potential agreement. In this analysis I focus solely on distributional uncertainty. My justification is both theoretical and empirical. Uncertainty about the absolute level of gains is probably at least as pervasive as distributional uncertainty. However, what would make absolute uncertainty interesting from the viewpoint of cooperation theory is if its presence precluded any agreement at all. In other words, are there many cases in which there was substantial uncertainty about whether an agreement would produce a net gain and hence in which provisions for limited duration and renegotiation made cooperation possible? The one issue area for which absolute uncertainty might play such a role is the environment. Nonetheless, in that issue area, questions of distribution also loom large.[11] In any event, the distribution question seems at least as important empirically and much more interesting theoretically.[12]

Assumptions

The following assumptions underlie the model:

States care about the future—that is, their discount factor is not zero.

States are risk averse.

There is uncertainty about future states of the world.

The costs of making agreements completely contingent are sufficiently large that the parties do not ever choose to do so.[13]

Rosendorff and Milner consider similar models in a trade context. See Downs and Rocke 1995; and Rosendorff and Milner forthcoming.

11. For example, writing about the Convention on Biological Diversity, Raustiala argues that the convention "addressed three (linked) central concerns: the conservation of biodiversity, the promotion of sustainable use, and the equitable sharing of benefits. It is this latter objective, with its clear redistributive implications, that was and remains the cause of much debate." Raustiala 1997, 491.

12. See also fn. 18 below.

13. For a justification of this assumption in the context of the economics literature on incomplete contracts, see the discussion in Hart and Holmstrom 1987.

https://doi.org/10.1162/002081801511140586 Published online by Cambridge University Press

There are costs to negotiating (and renegotiating) agreements.

There are costs to reneging on agreements.

States have shared *a priori* beliefs about the information they do not possess, and they revise their beliefs according to Bayesian logic as their interactions evolve.[14]

All but the last of these assumptions are necessary to produce what we observe in real life, namely, that states conclude both long and short finite-duration agreements as well as indefinite-duration agreements, that most agreements contain no or very few contingent provisions, and that states generally do not renege on the agreements they conclude.[15]

Consider the role each assumption plays. If states did not care about the future, they would not bother to conclude agreements. If states were not risk averse, they would always conclude indefinite agreements in order to avoid paying renegotiation costs and would not care *ex ante* about how much the distribution of gains might differ from that originally agreed upon. If there were no uncertainty about future states of the world, there would be no additional information to incorporate through renegotiation, and all agreements would be indefinite in order to economize on renegotiation costs. If agreements could be made perfectly contingent on the realized state of the world, states would do so and thereby save on renegotiation costs by eliminating the need for renegotiation. If negotiation and renegotiation were costless, states would renegotiate every time new information arrived, and all agreements would be of short duration. If reneging were costless, states would renege often (assuming renegotiation costs were not too high) and would be less likely to adopt finite-duration agreements. Instead, they would use reneging as a form of contingent renegotiation within indefinite-duration agreements.

Basics of the Model

There are two prospective parties to the agreement. In the absence of an agreement, each party obtains an outcome every period. This outcome depends on the particular context, but it could represent something like GNP or some measure of military security. Each party has an expectation about what its outcome will be every period—for example, state 1 expects its GNP to be $1 trillion. This is the party's base outcome. Of course, the actual outcome will rarely, if ever, correspond exactly to the base. The actual outcome will consist of the base plus or minus some amount. For example, state 1's actual GNP might

14. In Knightian terms, the parties face *risk,* not *uncertainty.*

15. For evidence that states conclude both short and long finite-duration agreements as well as indefinite agreements and that most agreements do not have many (if any) contingent provisions, see Koremenos 1999. On the reneging point, see Morgenthau 1978; Downs, Rocke, and Barsoom 1996; Chayes and Chayes 1993; and Henkin 1979.

https://doi.org/10.1162/002081801511140586 Published online by Cambridge University Press

be \$0.9 trillion or \$1.1 trillion. I refer to this unanticipated variation as the outcome shock or noise.

I assume the agreement yields a total gain that is known at the time the agreement is concluded. What is not known at that time is how this gain will accrue to the two parties in practice. I assume that the parties can set the expected value of the two shares in the initial agreement.

The division of the gain agreed upon in the initial agreement reflects the relative bargaining power of the two parties. For example, suppose that states 1 and 2 have equal bargaining power and they conclude a joint research venture that will result in a total profit of \$25 billion. What cannot be known in advance is exactly how whatever technology emerges from the venture will benefit industry in each of the two states. Initially, each state invests an equal amount, and the parties set the expected gain to be the same for both states, \$12.5 billion.

The basic problem facing the parties to an agreement in this model is to sort out the effects of the agreement from other random fluctuations in outcomes. For example, suppose that, after the joint venture is concluded, state 1's GNP is \$1.05 trillion. How can state 1 know how much (if any) of the \$50 billion increase in GNP results from the joint venture and how much results from an agricultural boom spawned by favorable weather? The answer is that it cannot know exactly, but it can learn over time.

The states face a choice between an agreement of indefinite duration and one finite-duration agreement followed by an agreement of indefinite duration. In the simple two-period case I consider formally, the choice becomes one two-period agreement with no renegotiation or two one-period agreements with renegotiation in-between to realign the distribution of gains.

Renegotiation takes place whenever a finite duration agreement comes to an end. Thus the reservation outcome for both parties in the renegotiation consists of the no-agreement outcome. Essentially, the parties are in the same situation with renegotiation as in the original negotiation except that they have learned something about the realized distribution of gains from the agreement in the interim. Once the parties choose an indefinite-duration agreement, no further renegotiation takes place.

I assume that if and when the parties renegotiate the agreement, they incorporate an adjustment factor that makes the expected gain to each of the parties the same as it was in the original agreement. This adjustment factor takes account of the information gained about the realized value of the distribution of gains during the periods since the original agreement was concluded.

For example, suppose that, after a number of years, states 1 and 2 learn that state 1's industry is actually reaping significantly more benefits from the research venture than state 2's industry. If the parties originally agreed to a finite-duration agreement followed by renegotiation, then when they renegotiate at the end of the initial agreement, they will adjust the investment schedule so that state 1 invests more and state 2 invests less. This change will roughly bring the actual distribution of gains in line with what was expected when the agreement was first concluded. Assuming

https://doi.org/10.1162/002081801511405 86 Published online by Cambridge University Press

that the same expected division of gains is the result of every renegotiation is another way of saying that the relative bargaining power of the two parties is assumed to be constant over time.

In general, we would expect the bargaining power of the two states to change over time as their economic fortunes change. In particular, we might expect the realized division of gains under the agreement to affect the bargaining power of the two states if and when an agreement is renegotiated. Thus rather than returning to the initial expected division of gains in the renegotiated agreement, the states would agree to a new expected division of gains that would be more favorable to the party whose realized gain exceeded its original expected gain. Adding changes in bargaining power to the model in this form increases the variance of the outcomes conditional on renegotiation because the renegotiation no longer tries to undo completely the difference between the expected and realized gain. This, in turn, enlarges the set of cases in which no renegotiation would be chosen by risk-averse states.

I do not incorporate the effects of changes in bargaining power into the model for four reasons. First, and most important, given that the duration problem has been wholly neglected in the literature thus far, I choose to focus exclusively on it and keep other elements of the context (including the bargaining component) as simple as possible.[16] Second, allowing changes in bargaining power does not affect the comparative statics presented later. It changes the location of the cutpoint at which parties switch from one form of agreement to another, but the general comparative static results remain.[17] Third, in some cases the resources affected by the agreement in question are small relative to GNP so that the actual effect of the realized division of gains under the agreement on the parties' bargaining power would be small. Fourth, for pairs of states involved in multiple agreements, the agreement shocks will tend to average out, so that the states' relative bargaining power remains roughly the same.

In the context of international relations, there is no external authority available to enforce agreements. In other words, states can renege. In this model, reneging is equivalent to abandoning the agreement. I assume that parties that renege suffer a cost. I also assume that the parties can negotiate a new agreement in the period following the abandonment of the old.

Hence the basic intuition of my model is that the parties will integrate planned renegotiation into international agreements when the value to them of reducing the *ex ante* variance of the outcome stemming from agreement uncertainty is large relative to the cost of renegotiating. This reduction in *ex ante* variance is achieved by realigning the division of gains at the time of the renegotiation to be closer to the

16. This is especially important given that bargaining theory has not yet produced results that are robust. For example, results reported by Fearon disappear as soon as the war of attrition model is replaced with a Rubinstein alternating offers model. Fearon 1998. Moreover, even within particular models, results depend greatly on very specific assumptions, such as the time between offers.

17. I elaborate this point later.

https://doi.org/10.1162/002081801511140586 Published online by Cambridge University Press

original division by incorporating an adjustment factor into the agreement. Note that the adjustment factor is chosen by the parties within the model; it is not a parameter of the model for which comparative static results can be obtained. Put differently, it is an endogenous and not an exogenous variable.

Notation

Formally, assume that there are two prospective parties to the agreement, $n = 1, 2$. Let their outcomes in each period t in the absence of the agreement be given by

$$Y_{1,t} = b_1 + u_{1,t}$$

$$Y_{2,t} = b_2 + u_{2,t},$$

where the outcome $Y_{n,t}$ depends on the particular context but could represent something like GNP, where b_1 and b_2 are the expected values of the outcome measure, and where $u_{1,t}$ and $u_{2,t}$ represent variation in the outcome over time, independent of the agreement—noise. I assume that $u_{1,t}$ and $u_{2,t}$ have mean zero and are independently and identically distributed across periods and across parties with probability density function $f(u)$. Without loss of generality, I normalize b_1 and b_2 to zero.

With respect to the agreement, let the total gain from the agreement be a fixed amount g, known to both parties.[18] Denote the expected values of the shares determined in the bargaining process by m for party 1 and by $(g - m)$ for party 2. I assume that the actual realization $(m + \varepsilon)$ is a random variable with probability density function $h(m + \varepsilon)$ where $E(m + \varepsilon) = m$. Thus in the presence of the agreement, the outcomes of the two parties in period t are given by

$$Y_{1,t} = m + \varepsilon + u_{1,t}$$

$$Y_{2,t} = g - (m + \varepsilon) + u_{2,t},$$

18. The simplifying assumption of a fixed, known g can be relaxed without changing any implications of the model. A more general model would make g random, with the parties then facing the more difficult problem of untangling both the total gain and the distribution of gains from the normal noise in the outcome. With a random g, agreements might be concluded in which the expected value of g *ex ante* was positive but the realized value was negative. This provides an additional motivation for having a finite-duration agreement. Instead of renegotiating, the parties will simply not conclude additional agreements if they learn that g is probably negative. In many agreement contexts, states clearly care about both the distribution of gains and the total gain. I have chosen to focus here on the distribution of gains in the interest of parsimony. None of the comparative statics in my model depend on the assumption of a fixed g, but allowing g to be a random variable would substantially increase the notational burden and the formal complexity of the model.

https://doi.org/10.1162/002081801511405 86 Published online by Cambridge University Press

with associated expected values

$$E(Y_{1,t}) = m$$

$$E(Y_{2,t}) = g - m.^{19}$$

Note that ε has no t subscript because it represents the one-time agreement uncertainty. It is drawn at the time an agreement is concluded and stays the same after that. In contrast, the u's, which represent the outcome shocks or noise, do have t subscripts, as new u's are drawn for both parties each period. Thus the u's embody persistent noise.

Initial negotiation costs are k_1. Renegotiation costs are k_2. The adjustment factor incorporated into the agreement at the renegotiation stage is a, and the cost paid by parties that renege is c.

In sum, states are often unsure about how agreements will work in practice. The difference between how they expect an agreement to work and how it actually works is represented by ε, the random component of the distribution of gains from the agreement. The parties know the distribution from which ε is drawn but must learn about the particular value of ε for their agreement. The variance of the distribution from which ε is drawn represents their degree of agreement uncertainty.

In each period, each of the two parties to an agreement receives some outcome Y. In the absence of an agreement, the outcome consists of a known base b and a period-specific random component u. Since the parties know b, in the absence of an agreement they can figure out u.

When there is an agreement, the situation changes. The outcome Y now consists of three components: b, u, and either $(m + \varepsilon)$ or $g - (m + \varepsilon)$. The terms involving ε represent the gains from the agreement. Like u, these terms are random. Unlike u, these terms are fixed; they do not vary from period to period. The basic problem facing the parties is to sort out the effects of the agreement, ε, from the normal noise in the outcome, u. Over time, the parties can learn about the value of $(m + \varepsilon)$ or $(m - \varepsilon)$ realized under the agreement. That is, over time they can distinguish the effect of the agreement on their outcomes from the period-to-period variation due to u.

Two-period Game

For simplicity, I assume throughout that the parties have identical utility functions and bargaining power and that the agreement yields a positive gain.

19. Note that without period-specific shocks, determining the value of ε would take only a single period. With only a single common shock, $u_t = u_{1,t} = u_{2,t}$, the exact value of ε could be determined in two periods.

https://doi.org/10.1162/002081801511405 86 Published online by Cambridge University Press

The bargaining outcome in both periods is exogenous and satisfies the Nash bargaining solution.[20]

Timeline. At the beginning of period 1, the two parties, $n = 1, 2$, play a Nash demand game in which they choose the expected division of gains. The parties' Nash demand game strategies consist of $\{m_n{:}m_n \in [0, g]\}$. If $m_1 = m_2$ and $m_n \in [0, g]$, the parties continue the negotiations. In all other cases, the parties conclude no agreement.

If the parties continue the negotiations, they then enter the agreement-type choice stage wherein each party must choose among the following: no agreement (*NA*); one two-period agreement, which is the analog of a nonrenegotiated agreement (*NR*); and two one-period agreements, the analog of a renegotiated agreement (*R*). If both parties choose *NR*, the parties enter into a nonrenegotiated agreement. If both parties choose *R*, the parties enter into a renegotiated agreement. Otherwise, the parties conclude no agreement (*NA*). Note that nine possible strategy profiles can result at this stage, $\{NA, NR, R\} \times \{NA, NR, R\}$, and only two, (*NR*, *NR*) and (*R*, *R*), result in an agreement.

Next, nature draws $u_{1,1}$ and $u_{2,1}$ and, if there is some form of agreement, ε. The outcomes for period 1, $Y_{1,1}$ and $Y_{2,1}$, are observed by both players.

If the parties concluded a two-period agreement in the first period, then at the beginning of period 2 they choose whether to abide by the duration provision stipulated in the agreement. Similarly, if the parties agreed to two one-period agreements with renegotiation in between, they must decide whether to proceed with the renegotiation.

If the parties negotiated a one-period agreement in the initial period and elect to proceed with renegotiation, or if one party reneges on a two-period agreement, the parties negotiate a new agreement. They play a Nash demand game in which they choose the expected division of gains, where their action set again consists of $\{m_n{*}{:}m_n{*} \in [0, g]\}$. If $m_1{*} = m_2{*}$ and $m_n{*} \in [0, g]$, the parties conclude a one-period agreement. Otherwise, the parties conclude no agreement.

Next, nature draws $u_{1,2}$ and $u_{2,2}$. At this point the parties receive their payoffs, and the game ends.

Equilibrium. The equilibrium concept employed is perfect Bayesian equilibrium. I employ an incomplete-information solution concept because, even though the preferences of both parties are common knowledge, there is uncertainty about the physical consequences of any concluded agreement. This gets translated as uncertainty regarding preferences over possible agreements. In this particular setup, each party is the "opposite type" from the other; as each party learns about its own type, it also learns about the other's type. We can think of party 1 as type $+\varepsilon$ and party

20. This cooperative solution corresponds to the Rubinstein alternating-offers noncooperative solution when δ is close to 1. See Osborne and Rubinstein 1990.

https://doi.org/10.1162/002081801511405586 Published online by Cambridge University Press

2 as type $-\varepsilon$. Additionally, I impose the following restriction: The set of punishments used by the parties (that is, the costs a party pays after reneging) must be renegotiation-proof. Appendix A provides a characterization of an equilibrium of the game.

Comparative Statics

I focus here on the two most important implications of my model. Both have to do with the effects of changing the amount and type of uncertainty faced by the parties to an agreement. First, consider the effects of changes in the degree of agreement uncertainty, represented by the variance of ε (the shock to the distribution of gains under the agreement):

Hypothesis 1: All else equal, for risk-averse parties an increase in agreement uncertainty (the variance of ε) increases the value of renegotiation and therefore makes the parties more likely to choose a renegotiated agreement (two one-period agreements) than a nonrenegotiated agreement (one two-period agreement).[21]

To see the intuition here, it helps to think of the effect of an increase in the variance of the agreement shock ε in two ways: absolutely and relative to the variance of the noise, u. To see the absolute effect, consider the special case where the variance of the noise, u, is zero. In this special case, an increase in the degree of agreement uncertainty would still increase the value of renegotiation. The more variable the agreement shocks, the more that risk-averse states gain in expected utility from being able to undo them through renegotiation.

To see the relative effect, return to the general case where the variance of u is not zero. In the general case, states can learn more about the realized value of ε when it is more easily distinguished from u. Increasing the variance of ε while holding the variance of u constant does just that—it makes it easier to distinguish the agreement shock from the noise. Put somewhat differently, increasing the variance of ε relative to that of u makes the first-period outcomes more informative about ε. Because renegotiation is more valuable when states have better information about the realized value of ε at the time they renegotiate, an increase in the variance of ε again increases the value of renegotiation.

Now consider the effect of an increase in the degree of agreement uncertainty on whether states choose to conclude any agreement. For risk-averse parties, any increase in the variance of the outcomes under the agreement reduces the expected utility (at the time of the decision whether to conclude an agreement) under either type of agreement relative to no agreement. This is obvious in the case of a nonrenegotiated agreement, but it is also the case for a renegotiated agreement, since the adjustment mechanism does not undo the effects of the agreement shocks in every state of the world. Note, however, that as long as the distribution of the

21. It also makes it more likely that the parties will choose no agreement at all.

https://doi.org/10.1162/002081801511405 86 Published online by Cambridge University Press

agreement shock is such that the probability that a state actually loses from an agreement is zero, the parties will always choose some form of agreement as long as the negotiation costs are not too large.

The second important implication of my model concerns the effects of changes in the variance of u, the factors outside the agreement ("noise" in this context) that affect the outcome of interest:

Hypothesis 2: All else equal, for risk-averse parties an increase in noise (the variance of u) decreases the value of renegotiation and therefore makes the parties more likely to choose a nonrenegotiated agreement (a two-period agreement) than a renegotiated agreement (two one-period agreements).

The intuition here concerns the value of renegotiating at the end of the first period. As the noise increases, the amount of information the first-period outcomes provide about the value of ε decreases. The less information the parties have about ε, the less value they place on being able to reset the division of gains under the agreement, and therefore the less value they place on renegotiating. In other words, an increase in the noise decreases the information content of the first-period realizations, with the result that the parties learn less about the true value of the agreement shock. This, in turn, means that the parties cannot do as good a job of realigning the distribution of gains, which decreases the (*ex ante*) value of renegotiation.[22] Note that this is precisely the reverse of what happens under hypothesis 1 when the relative variance of ε increases.

Figures 1 and 2 illustrate the two comparative static hypotheses using simulated choices from a discretized version of the two-period model. In the simulations, the base outcome, b, is set to 20.[23] The gain from the agreement, g, is set equal to 8 and is assumed to be divided equally in expectation between the two parties so that $m = 4$. I fix the values of the discount factor δ at 0.9; and the costs of negotiation, renegotiation, and reneging, k_1, k_2, and c, at 1.0, 0.5, and 1.0, respectively. I use a cube root utility function (that is, the utility from a given outcome is its cube root), which implicitly sets the level of risk aversion for the parties. Appendix B describes the discretized model and its solution in greater detail.

Conditional on these values for the utility function and the model parameters, I calculate the utility of each state for a large number of values for the variances of

22. Note that in addition to its effect on the relative attractiveness of a renegotiated agreement, an increase in the variance of u tends to decrease the expected utility associated with every agreement-type choice (including no agreement) for risk-averse parties by increasing the variance of the realized outcomes.

23. I use a positive base value to ensure that the utility associated with each possible realization is always positive. This ensures that I can calculate utility values even with utility functions such as the cube root. I could change the base value to some larger number, such as 100 or 1,000, without changing any of the substantive results. Recall that in the model the base is normalized to zero.

https://doi.org/10.1162/002081801511405586 Published online by Cambridge University Press

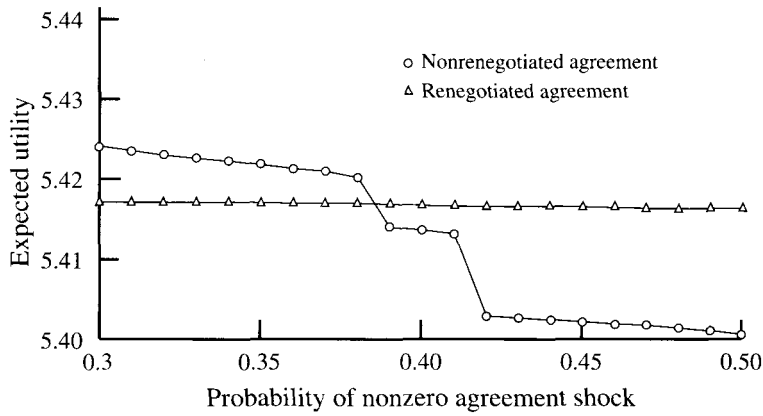


FIGURE 1. *Increasing the variance of the agreement shock*

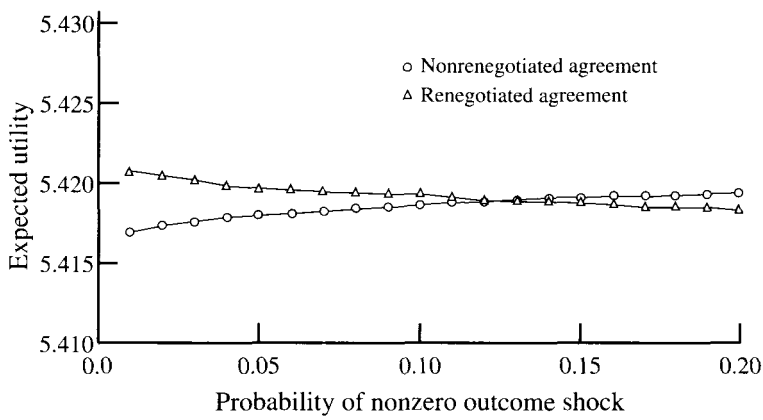


FIGURE 2. *Increasing the variance of the noise*

the noise in the base outcome, u, and the one-time agreement shock, ε. In each case, I assume that u_1, u_2, and ε take on the values of –2 , 0, or 2.

In Figure 1, I illustrate the effects of increasing the variance of the agreement shock, ε. I start with probabilities of (0.3,0.4,0.3) for the three values of ε and then symmetrically increase the probabilities of the two nonzero values until I end up with probabilities of (0.5,0.0,0.5). In other words, I increase the likelihood that the parties will receive a nonzero agreement shock. The probability of each nonzero value is shown on the horizontal axis of the figure. The probabilities for u remain constant at (0.3,0.4,0.3) for all of the cases in Figure 1. The two lines in Figure 1 trace out the expected utility for both parties (they are identical) associated with a nonrenegotiated agreement (the line with the circles), and a renegotiated agreement

https://doi.org/10.1162/002081801511405586 Published online by Cambridge University Press

(the line with the triangles).[24] Thus moving from left to right in the figure shows the effects on the expected utilities of the different agreement types of an increase in the degree of agreement uncertainty, holding the noise constant. It can be seen that initially the states receive a higher expected utility from a nonrenegotiated agreement (one two-period agreement) but that as the variance of ε increases, the expected utility of a renegotiated agreement (two one-period agreements) eventually comes to dominate.[25] As a result, the states in my model change their agreement-type choice when the degree of agreement uncertainty exceeds a certain level.[26]

Figure 2 does the same thing, but this time allowing the probabilities of the values of u to vary from (0.0,1.0,0.0) to (0.2,0.6,0.2) while holding the probabilities of the values of ε constant at (0.3,0.4,0.3). As you move to the right, which represents an increase in the degree of noise because the probabilities of the nonzero values are increasing, the states' preferences change from wanting to have a renegotiated agreement (two one-period agreements) to wanting to have a nonrenegotiated agreement (one two-period agreement).[27] This reflects the decreasing value of renegotiation as the variance of u increases, which causes u to "drown out" the information about ε implicit in the parties' first-period outcomes. In other words, as the environment becomes noisier, it is harder for states to learn.[28]

24. Figures 1 and 2 both omit the value of no agreement. In each case it always lies between 5.1 and 5.2. In Figure 1 it is flat since it is unaffected by the variance of ε. In Figure 2 it declines with the variance of u.

25. The sudden drops in the expected utility of the nonrenegotiated agreement in Figure 1 result from states deciding to renege in particular states of the world when the variance of the agreement shock reaches a certain level.

26. If the bargaining power of the parties were affected by the realized distribution of gains as discussed earlier, this would reduce the expected utility of a renegotiated agreement for all values of the variance of the agreement shock. This, in turn, would lead the two lines in Figure 1 to cross to the right of where they do now, implying that the states would choose not to renegotiate in some cases where they otherwise would have.

27. The fact that the expected utility associated with a nonrenegotiated agreement is increasing in the variance of the outcome shock (the noise) over most of the range shown in Figure 2 may seem puzzling given that the states are risk-averse. This pattern results from the effect of the variance of the noise on the probability of states of the world in which one state reneges and imposes on the other state a cost larger than the benefit it gets from doing so. As the variance of the noise increases, these states of the world become less likely. Over this range, the positive effect of reducing the probabilities of these states of the world on the overall expected value outweighs the negative effect of the increasing variance in outcomes.

28. It is important to note that the implications of my model are consistent with certain neorealist views of international cooperation. While my model incorporates neoliberal assumptions (that is, states care about absolute gains), incorporating the neorealist assumption that states care about relative gains would actually *enlarge* the set of cases for which states would choose to incorporate renegotiation provisions into their international agreements. In fact, Grieco states that "If two states are worried or uncertain about relative achievement of gains, each will prefer a less durable cooperative arrangement, for each will want to more readily be able to exit from the arrangement in the event that gaps in gains favor the other." Grieco 1990, 228. (I thank an anonymous referee for pointing me to this passage in Grieco's work.) In terms of Figure 1, adding in concerns about relative gains would lower the expected value of a nonrenegotiated agreement and raise the value of a renegotiated agreement. These movements result from the fact that concerns about relative gains magnify the utility gains and losses associated with any given departure from the agreed-upon division of gains. The net result is that the lines in Figure 1 would then cross to the left of where they do in a world where states care only about absolute gains,

https://doi.org/10.1162/002081801511405586 Published online by Cambridge University Press

Up to this point, I have considered a two-period version of my model solely for simplicity in exposition and analysis. The real world, of course, has more than two periods. A more general version of my model lengthens the time horizon to infinity. The basics of the analysis stay the same, but states now may face two choices. The first is an agreement-type choice between no agreement, one infinite-duration agreement, and one finite agreement followed by an infinite-duration agreement. If states choose to renegotiate the agreement, they must then make a second choice regarding the timing of the renegotiation. The degree of noise in the environment will determine the optimal timing, with more noise leading to a longer period before renegotiation so that the parties have more time to learn about the true distribution of gains. This case of a finite-duration agreement followed by renegotiation and an indefinite-duration agreement is of particular interest because the NPT adopts this form.

Nuclear Non-Proliferation Treaty

I first describe the substance of the NPT, including as much information as is publicly available about its duration and renegotiation provisions and how they were arrived at; I then discuss the agreement in light of my model.

This case study makes the following points. First, it reveals the empirical importance of this structure of duration and renegotation provisions. The NPT is arguably one of the more important international agreements of this century. An understanding of its provisions necessarily informs any analysis of its successes and failures and also contributes to the effective design of future agreements.

Second, at an even more basic level, the case demonstrates the importance that the parties to major international agreements assign to duration and renegotiation provisions. They spend time debating them during the negotiation process, they implicitly (and sometimes explicitly) recognize the trade-off captured by my model, and they choose these provisions in a way that my model shows to make sense.

Third, the case study provides an example of how to operationalize the key variables in the model. The variables in the model, like many variables considered by political scientists more generally, are more difficult to operationalize than, for example, earnings are for an economist. The case pays particular attention to operationalizing the distribution of gains from the agreement and the uncertainty that initially surrounded this distribution.

Fourth, the case study demonstrates that my model provides a powerful framework for organizing and systematizing discussions of the duration and renegotiation provisions of international agreements. By identifying the key variables that determine these provisions, it guides the investigator in sifting through what is, in the

which means that adding in concerns about relative gains *increases* the set of cases in which the states choose to renegotiate.

https://doi.org/10.1162/002081801511405586 Published online by Cambridge University Press

case of the NPT, a very large volume of information. It also highlights key areas on which the parties disagreed and shows how the solutions they chose solved the problems they faced.

Background and Substance of the Agreement

The NPT was signed in 1968 and entered into force in 1970 for a period of twenty-five years. In 1995 the parties to the treaty reconvened and decided to extend the treaty indefinitely. The NPT arose out of fears on the part of the existing nuclear-weapon states (NWS) during the early Cold War era that the spread of nuclear weapons to a substantial number of additional countries would be both dangerous and destabilizing. As Thomas Graham recounts, "In 1968, the United States Atomic Energy Commission foresaw a world that might have as many as twenty-eight nuclear powers. The danger that such a world would pose cannot be overstated." Quoting a Swiss official, he continues, "Between two nuclear powers it's a game of chess, among four, it's bridge, among a dozen, it would be poker, roulette, or any of those games controlled by chance."[29] In response to this concern, the United States and the Soviet Union, along with a number of other countries, undertook to establish a treaty prohibiting the further proliferation of nuclear weapons.

The treaty has the following main provisions: Article 1 prohibits NWS from transferring nuclear explosives to any recipient regardless of whether that recipient is a party to the NPT and from otherwise assisting a non-nuclear-weapon state (NNWS) in developing such weapons. Article 2 places obligations on the NNWS not to receive or manufacture nuclear explosive devices. Article 3 requires that the signatories negotiate either individually or collectively full-scope safeguards agreements with the International Atomic Energy Association (IAEA). Articles 4 and 5 provide reassurance to the NNWS that they will be able to enjoy the peaceful uses of nuclear energy and nuclear explosions without discrimination, that is, the NWS are obliged to provide both technological and material assistance to the NNWS. Article 6 demands progress by the existing nuclear powers on controlling the arms race.[30]

Duration and Renegotiation: The Role of Uncertainty

Choosing the duration and renegotiation provisions of the NPT provoked an intense debate. The treaty negotiations lasted from 1962 to 1968. As late as 1967, the United States and the Soviet Union (the original drafters of the treaty) were pressing for a treaty with an unlimited duration, whereas the Germans and the Italians were emphasizing the impossibility of accepting such a duration. Because of their

29. Graham 1989, 662.
30. Treaty on the Non-Proliferation of Nuclear Weapons 1968.

https://doi.org/10.1162/002081801511405586 Published online by Cambridge University Press

uncertainty about the distribution of gains that would result from the NPT, many NNWS mirrored Germany and Italy in being wary about tying their hands for an unlimited period in an uncertain world.

First, uncertainties surrounded the security consequences of the treaty. The NNWS felt great uncertainty about the effort that the NWS would put into nuclear disarmament.[31] Closely related to this is a paradox inherent in the text of the NPT. If the NWS really did reduce their nuclear stockpiles as Article 6 commits them to do, the extended deterrence they provide to their NNWS allies would become less credible, and these states would have a stronger incentive to join the nuclear club.[32] In regard to the nonnuclear NATO countries, George Bunn and Charles N. Van Doren note that the "countries that were most advanced in civilian nuclear technology and that relied on an alliance with the United States to deter possible attack by the Soviet Union spoke out ... against an NPT of longer duration than their alliances might turn out to be."[33] Jenson content-analyzed speeches made during a 1968 UN General Assembly debate on the proposed treaty in an effort to ascertain and categorize reservations. He reports that 62 percent of the speakers expressed concern regarding security guarantees, wondering how NNWS would be protected under the NPT.[34]

Another source of uncertainty about the distribution (and level) of security benefits under the NPT centered on which countries would end up participating in the regime. The overall level of gains increases with the number of countries that join, and the distribution of gains depends heavily on the geographic distribution of signators.

Second, the NNWS were concerned about the effect of the NPT on their economic prosperity and on their technological development. They were apprehensive that the treaty might restrict their ability to make peaceful use of nuclear energy. The treaty would have to draw a line between the use of nuclear energy for peaceful, civilian purposes and the use of nuclear energy for military purposes. Much uncertainty existed among the NNWS about whether this line could be effectively drawn and what the distributional effects would be of drawing it. Speaking before the Bundestag in 1967, Foreign Minister Brandt declared that his "government and others are also seeking to insure that the nonproliferation treaty does not further widen the already existing technological gap between the nuclear powers and the non-nuclear countries."[35]

31. A broader formal model of duration and renegotiation choice in the NPT could include NWS effort as an unobserved variable and actual arms reductions as an observed variable increasing in effort but not completely determined by it. The NNWS, which might condition their willingness to extend the treaty on observed disarmament, would then essentially be in a principal-agent relationship with the NWS.

32. See, for example, the discussion in Smith 1987.

33. Bunn and Van Doren 1992, 5. These NNWS did not propose making the NPT duration contingent on the continued membership of the United States in NATO. This suggests that the barriers to contingency in agreements that are often invoked in theory are also present in fact.

34. Jenson 1974, 2.

35. U.S. Arms Control and Disarmament Agency 1967, 49.

https://doi.org/10.1162/002081801511405 86 Published online by Cambridge University Press

Many potential NPT members also worried about how economically costly the IAEA monitoring would turn out to be. George Quester details concerns raised by the Japanese nuclear power industry about the potential costs associated with the extensive on-site monitoring required by the NPT, including the possibility of shutting down plants in order to allow verification of nuclear fuel information. Quester also notes that many states (including Japan) worried that "the IAEA inspectorate may become imperiously bureaucratic, demanding greater access even where no increase whatsoever is thereby achieved in safeguards reliability . . . due to personal vanity, institutional imperialism, or excessive legalism."[36] These concerns affect both the distribution and level of gains given concerns that the IAEA would be dominated by those NWS whose experts performed the inspections; those states could vary the inspection costs to suit their political and economic agenda.

Another area of uncertainty about the distribution of economic gains concerned the NPT's failure to fully address the relationship between parties to the NPT who supply nuclear technology and NNWS not party to the treaty who purchase it. The NNWS that joined the treaty worried that NNWS not party to the treaty would be able to obtain nuclear technology with fewer restrictions than signatory NNWS.

The potential decline of the U.S. commercial nuclear industry added still another source of uncertainty. As Roger K. Smith notes, this decline was expected, but it would have been difficult to know in advance when other suppliers would emerge and who they would be.[37] These changes would affect states on both sides of the market for peaceful nuclear technology.

Third, uncertainties surrounded the political benefits and costs that would result from such an agreement. Because the NPT would make the NNWS importers of peaceful nuclear technology, they were uncertain about whether this situation would give the NWS political leverage they could exploit. A similar concern was expressed about fuel supplies. Moreover, the effect of treaty adherence on the political power and prestige of the NNWS was uncertain. Prestige might follow from the acquisition of nuclear weapons or it might follow instead from a state's willingness to accede to the treaty. As Lloyd Jenson states: "If there is general acceptance of the NPT, the few states refusing to join are likely to be just that much more criticized." Only time would tell how universal the treaty (and the norm embodied in it) would become.[38]

Other political concerns revolved around specific regional issues. According to Quester, Italy worried that European unification would "be impossible once the NPT has endorsed France and Britain as nuclear states and relegated Germany and Italy permanently to the position of nonweapons status."[39] That is, Italy worried that the distribution of gains from the NPT would skew the distribution of power in Europe in ways that would make European integration difficult if not impossible. On the other side of the globe, Japan had concerns over how the NPT would affect its

36. Quester 1973, 105, 212.
37. Smith 1987.
38. Jenson 1974, 38.
39. Quester 1973, 7.

https://doi.org/10.1162/002081801511405586 Published online by Cambridge University Press

ability to react to a broader U.S. withdrawal from Southeast Asia in the wake of the Vietnam War.

In sum, the parties to the NPT faced substantial uncertainty at the time of its inception about its effects on their security and on their economic and political well being. Some of this uncertainty, such as that surrounding how many countries would eventually join the regime, concerned the overall level of gains from the agreement. However, much of the political and security uncertainty, and all of the economic uncertainty, concerned the distribution of gains (and possibly losses) from the NPT. For example, conditional on any given number of countries joining the regime, which specific countries joined had large effects on the distribution of security gains from the agreement. In the economic realm, concerns about how well peaceful uses of nuclear power could be separated from military ones, competition in the market for peaceful nuclear power technology, and IAEA inspection costs are all distributional.

This kind of uncertainty is the cornerstone of hypothesis 1. In this case, the agreement shock, ε, can be broken down into three components: security, economic, and political. All three are characterized by high variance.[40]

Additionally, it is important to note that many aspects of the uncertainty about the distribution of gains from the NPT were of the one-time character that underlies my model. For example, to the extent that the advent of the NPT set in motion a transition to a stable equilibrium in terms of membership in the regime, all that states have to do is to wait and find out what that equilibrium will be to determine the realized distribution of gains. Similarly, once a stable set of institutions arises to govern the transfer of peaceful nuclear technology by members to nonmember NNWS, the member NNWS can determine the effect of this aspect of the NPT on their economic well being.

Duration and Renegotiation: The Compromise

The Italian representative to the negotiating committee, Roberto Caracciolo, proposed a "certain flexibility in the provisions of the treaty relating to duration, amendments, and the right of withdrawal"; reacting to Article 7 of the draft treaty, which called for a treaty of unlimited duration, he stated that, "it is not the lot of man, to pledge eternity. Moreover, if we look back across our thousands of years of history, we see very few noninstitutional treaties that have simply survived the vicissitudes of one generation, let alone achieved immortality. Therefore we fear that to affirm a principle so remote from reality may introduce into the treaty an

40. Hypothesis 1 is essentially a comparative static result, which cannot be tested with this case study. The case study does, however, show the plausibility of the model. Later, I briefly compare the NPT to the Outer Space Treaty for which the variance of the agreement shock was arguably low. Additionally, hypothesis 1 is put forth as a conjecture in the Rational Designs project (Koremenos, Lipson, and Snidal forthcoming) and is the most strongly supported conjecture among the eight case studies.

https://doi.org/10.1162/002081801511140586 Published online by Cambridge University Press

element of weakness rather than of strength."[41] Reacting to this same article, a Swiss aide-memoire stated that "to subscribe to such a commitment seems hardly conceivable in a field where development is as rapid and unpredictable as that of nuclear science and its technical, economic, political, and military implications. Consequently, it would be preferable that the treaty should be concluded for a definite period, at the end of which a review conference would decide about its renewal."[42]

The states that favored an unlimited duration argued that if the parties knew for certain that the treaty would end on a specified date in the future, they would feel pressure to obtain nuclear weapons by that date, thereby undermining the treaty. Furthermore, they argued that it was important that the treaty be guaranteed a life-span of sufficient length to enable it to serve as a foundation upon which other nuclear disarmament measures could be built.

The parties began the path toward compromise when Caracciolo submitted the following to replace the "unlimited duration" paragraph in the draft treaty: "This treaty shall have a duration of *X* years and shall be renewed automatically for any party which shall not have given, six months before the date of expiry of the treaty, notice of its intention to cease to be party to the treaty." He stated that "the proposed amendment could be regarded as an acceptable compromise between the idea of unlimited duration and that of a fixed term. It provides that the treaty shall always remain in force for those who do not denounce it, and at the same time allows those who are not satisfied with its operation to withdraw after a certain number of years."[43] A month later the Italian amendment was revised: "The present treaty shall have a duration of *X* years. It shall be automatically extended for terms equal to its initial duration for those governments which, subject to six months notice, shall not have made known their intention to withdraw."[44]

The final resolution of the duration and renegotiation issues was a compromise between the Soviet and U.S. desires for an indefinite duration agreement and the preferences of the NNWS for a finite learning period during which they could determine whether the agreement as it worked in practice was in their interest. As Bunn and Van Doren explain, "The Italian proposal for a specific period, plus successive automatic renewals, was the most detailed and the most important before the American and Soviet co-chairmen when they drafted the present language of Article 10.2." They continue: "What the Co-Chairmen drafted—the present Article 10.2—called for an extension conference after 25 years to decide whether to extend the Treaty 'indefinitely . . . or for an additional fixed period or periods.' It thus accepted the basic Italian idea of a first period of years at the end of which there would be an opportunity for renewal. To the Italian option for an indefinite number

41. U.S. Arms Control and Disarmament Agency 1967, 527–29.
42. Ibid., 573.
43. Ibid., 529.
44. Bunn and Van Doren 1991, 7.

https://doi.org/10.1162/002081801511140586 Published online by Cambridge University Press

of consecutive renewal periods, the drafters added other options, including indefinite renewal and one fixed period."[45]

The drafters also altered the renewal mechanism proposed by the Italians. Instead of individual notices of intentions to withdraw (and therefore automatic subsequent withdrawal), the duration issue would be decided at a multilateral extension conference after twenty-five years.

Why did the NWS and NNWS have different preferences over duration? This can easily be explained within the context of my model. It is reasonable to assume that the distribution of possible values of the agreement shock ε was skewed. Specifically, there was virtually no chance that the NWS could "lose big" from the NPT. The worst that could happen for them would be to find out that nuclear weapons had no use and hence were a complete waste of resources. In contrast there was a chance that the NNWS could "lose big" from adhering to an agreement that demanded they forsake nuclear weapons in a world in which other states—both members and nonmembers of the NPT—possessed them. Such a skewed distribution would imply different preferences over duration even if the expected gains of the NWS and NNWS were equal.[46]

Resolving the Uncertainty Through Learning

The sources of economic, political, and security outcomes are ambiguous. For example, if a state experiences a low growth rate in a given year, the role played by its decision not to develop its own nuclear industry may not be immediately apparent. The costs and benefits of nuclear energy as a source of power were quite uncertain in the late 1960s and early 1970s. Hence it would take time for states to sort out the effects of the NPT on their well being from those of other factors. Put another way, the level of noise (the cornerstone of hypothesis 2) is high.

If my model is relevant to this case, we should see evidence that the states involved did indeed initiate learning processes to help them distinguish the value of the agreement shock from the noise in the environment. Again, the real world is seldom two periods, and the NPT context is no exception. The multiperiod variant of hypothesis 2 addresses both the incidence and the timing of renegotiation.[47]

The parties to the NPT planned review conferences every five years at which they could cooperatively take stock of how the treaty was working in practice. This interval was chosen in part because it would take approximately that long to produce arms control agreements. Referring to the Third Review Conference, Hon. Lewis A. Dunn, of the U.S. Arms Control and Disarmament Agency, stated:

45. Ibid., 7–8.

46. For simplicity, the formal model and the simulations presented in the second section assume the agreement shock, ε, has a symmetric distribution. Importantly, there is nothing in the model that precludes the agreement shock from having a skewed distribution.

47. Koremenos 1999.

https://doi.org/10.1162/002081801511140586 Published online by Cambridge University Press

> The main task at the Review Conference will be an article-by-article review of the operation of the Treaty. Woven throughout that review and ensuing debate there will be five major questions. First, has the Treaty strengthened the security of the parties by helping prevent the further spread of nuclear explosives? Second, how well has the Treaty facilitated cooperation in the peaceful uses of nuclear energy? Third, what has been done to bring the nuclear arms race to an end? Fourth, what can realistically be done to strengthen the NPT? And most important of all, weighing each of these considerations, has the NPT been a success?[48]

The evidence indicates that the uncertainty about many aspects of the true distribution of gains persisted years after the NPT was signed. At the 1980 Review Conference, the debate centered on Article 4 (which focuses on technological exchange) and Article 6 (which focuses on arms control). Disagreements between the NWS and the NNWS regarding how these Articles were working in practice prevented the conference participants from agreeing on and hence issuing a final document. In other words, given the level of noise, the parties had not yet been able to determine the true distributional effects of the NPT. By 1985, much more of the uncertainty surrounding Article 4 had been resolved, as the following statement by Dunn illustrates:

> Since the 1980 Review Conference:
>
> - all new or amended agreements for cooperation with non-nuclear weapon states entered into by the United States have been with parties to the NPT or the Treaty of Tlatelolco;[49]
> - virtually all US exports of enriched uranium . . . were to NPT parties; . . .
> - all of US-funded IAEA "Footnote A" (extrabudgetary) technical assistance projects—111 projects for more than $4.5 million—have been for developing country Non-Proliferation Treaty parties.[50]

Essentially, the passage of time resolved much of the uncertainty over the distribution of economic gains from the NPT. Regarding the issue that raised so many concerns during the original negotiations, the parties successfully separated the civilian and military uses of nuclear technology so that trade in legitimate nuclear materials flourished under the NPT within a network of IAEA safeguards agreements. In practice, there has not been discrimination against the NNWS regarding technological progress in nuclear energy. With respect to Article 5, which addresses the issue of peaceful nuclear explosions, by the mid-1970s, it was determined by the NWS that the anticipated benefits from such explosions were not forthcoming.

Furthermore, over the two-and-a-half decades following the signing of the NPT, the parties witnessed the development of alternative suppliers of nuclear technology,

48. U.S. House of Representatives 1985, 65.
49. The Treaty of Tlatelolco establishes Latin America as a nuclear-weapons free zone.
50. U.S. House of Representatives 1985, 69–70.

https://doi.org/10.1162/002081801511405 86 Published online by Cambridge University Press

such as France and Germany. Taken together, these developments represented the stabilization of the regime regarding the distribution of peaceful nuclear technology to NNWS. Under the regime that eventually emerged, it was clear that the NNWS that joined the NPT would not suffer economic harm along this dimension.[51]

In addition to the concerns that played a major role at the review conferences, many other aspects of the uncertainty surrounding the distribution of gains from the NPT were largely or completely resolved during the initial trial period. In terms of security, the NPT greatly reduced the spread of nuclear weapons compared with what would likely have occurred without it. During the trial period, membership in the NPT increased to the point of being almost global.

In terms of the distribution of political gains (and losses), it became clear that concerns that the NPT would prevent European integration were groundless. What turned out to matter for European integration was not Britain's bombs but Germany's GNP. Time also rendered moot Japan's worries about its ability to react to a broad U.S. pullout from Asia.

In 1995 after four review conferences, the 163 parties to the treaty gathered in New York to decide whether the NPT would continue in force indefinitely or be extended for an additional fixed period or periods. Interviews with conference participants suggest that essentially all of the parties came to the conference favoring extension, a fact that itself provides powerful evidence of learning. Debate centered on whether extension would be indefinite or for a series of twenty-five-year periods.[52] In the end, a consensus resolution extended the NPT indefinitely. The NWS had gained what they expected to in terms of maintaining their power and influence (as Panofsky and Bunn note, "possession of nuclear weapons and permanent membership in the UN Security Council remain identical"),[53] and the NNWS had learned how the NPT worked for them in practice.

Additional Agreements

Antarctic Treaty of 1959

While the Antarctic Treaty and the NPT might at first glance appear unrelated, they share a common structure in their duration and renegotiation provisions. My model shows what it is about their contexts that drives this common structure.

51. Nye presents additional examples of learning and uncertainty resolution in his discussion of policies relating to the nuclear fuel cycle and attempts to control aspects of the cycle related to nuclear weapons development. Nye 1981.

52. The interviews appear in Welsh 1995. The debate regarding the extension provision was largely among the NNWS, since the NWS all favored indefinite extension. Ultimately, the Canadian argument that indefinite duration would cause the NWS "to be permanently held accountable to Article VI on disarmament" carried the day.

53. Panofsky and Bunn 1994, 9.

https://doi.org/10.1162/002081801511140586 Published online by Cambridge University Press

The Antarctic Treaty of 1959 governs international behavior in Antarctica in the presence of overlapping territorial claims.[54] The treaty stipulates that Antarctica should forever be used for peaceful purposes exclusively, prohibits measures of a military nature, prohibits nuclear explosions and disposal of radioactive waste, guarantees freedom of scientific research and promotes scientific exchange of both personnel and results, and establishes a comprehensive system of on-site inspection to ensure observance of the treaty. Perhaps most importantly, under the treaty the parties agree not to press or surrender their divergent views regarding territorial sovereignty.[55]

The Antarctic Treaty is of indefinite duration, but it has two regimes for renegotiation and amendment. For the first thirty years, the parties constrain themselves to learn about the workings of the agreement as written by making it very difficult to either modify or withdraw from the agreement. During this time, the treaty permits only Pareto improvements—those that can be unanimously agreed upon—to the agreement. After the first thirty years, renegotiation can occur essentially on demand. A majority can secure an amendment that redistributes the gains from the agreement. Those who gained less than they expected can threaten to leave unless the distribution of gains is adjusted. At the same time, too large a redistribution will cause the original winners to leave. Thus, shocks to the realized distribution of gains can be adjusted for within the context of the treaty itself, once sufficient time has passed to clarify the treaty's effects.

The duration of the Antarctic Treaty was an important issue during the negotiations. According to the leader of the U.S. delegation, because of concerns about their territorial claims, "the Chileans were adamant on having a definite time limit." The U.S. Secretary of State "pointed out that he thought some clause providing for revision in the future (which would therefore in effect counter any idea of a perpetual treaty) was desirable."[56] In hearings before the U.S. Senate Foreign Relations Committee, it was stated that thirty years "was the final adjustment arrived at and it was felt that if after 30 years of operation a majority, including a majority of the consultative parties, thought that there ought to be a change and that change had such merit, then a party should be able to release itself from the treaty if the change did not come into force."[57]

According to M. J. Peterson, the parties with territorial claims were more willing to accept the compromise on sovereignty when the treaty was given a definite term. She states: "The possibility of holding a review conference in thirty years meant that any participant could reopen the issue later if it so desired. . . . This interval

54. The twelve signatory states are Argentina, Australia, Belgium, Chile, France, Japan, New Zealand, Norway, South Africa, the Soviet Union, the United Kingdom, and the United States.

55. Antarctic Treaty 1959.

56. U.S. Department of State 1991, 609.

57. U.S. Senate 1960.

https://doi.org/10.1162/002081801511405586 Published online by Cambridge University Press

corresponded to the interval most of the experts believed was needed to learn enough to assess the resource potential of the continent."[58]

I argue that my model applies to the Antarctic Treaty. The uncertainty surrounding the treaty's effects was largely distributional, centering on three sets of opposing groups: the United States and the Soviet Union (locked in the Cold War power struggle), the territorial claimants and nonclaimants (division of territory is a zero-sum endeavor), and the technologically advanced states with the ability to extract Antarctic resources and less-developed states lacking that ability (given fixed resources, distributional considerations predominate). In all three cases, the level of uncertainty was high (hypothesis 1).

The conflicts animating all three sets of opposing groups related to the treaty's effect on the value of current and potential territorial claims. For example, one debate in the United States centered on whether the treaty would strengthen the Soviet Union's influence in the region, its latent territorial claims, and its access to Antarctic resources.[59] The United States was also concerned that the treaty might improve the position of past claimants at its expense.[60] The opposition of Chile and Argentina to an indefinite initial duration arose from concerns that, without the opportunity to review the treaty after a finite period, they would lack the means to halt any erosion of their territorial claims. It is important to note that the concerns of the United States and those of past claimants Chile and Argentina constituted worries about the same shock, with the United States concerned that the shock would turn out one way and Argentina and Chile concerned it would turn out the other. The uncertainty that all the parties felt about the treaty's effects on actual or potential territorial claims has the "one time" aspect that leads to agreements with finite trial periods in my model.

Uncertainty also surrounded the conflict between those states that could extract resources in the short run and those that could not. In particular, how would scientific activity in Antarctica affect the legal basis for future claims regarding mineral resource exploitation? Most experts believed that scientific activity would have implications for resource jurisdiction if a minerals regime came into being, but its precise importance was not clear.

Importantly, the historical evidence suggests that the parties viewed the first thirty years of the treaty as a trial period during which to learn about the distribution of gains under the agreement.[61] In terms of the model, the variance of agreement shock is large, but the uncertainty it embodies is mostly of a one-time character. Thus hypothesis 1 indicates that the parties should choose a renegotiated agreement rather than one of initially indefinite duration. At the same time, noise is important here as well, since short-run flare-ups in the Cold War or transitory domestic political unrest in Argentina or Chile could, in the short run, make the treaty's long-run equilibrium

58. Peterson 1988, 72.
59. Boczek 1984.
60. Taubenfeld 1961.
61. For evidence on this point, see Koremenos 1999.

https://doi.org/10.1162/002081801511140586 Published online by Cambridge University Press

unclear. In a two-period world hypothesis 2 then argues against a renegotiated agreement. However, as I argued earlier, in the real world of more than two periods, the effect of the variance of the noise will often be to determine the duration of the trial period in a renegotiated agreement, rather than to prevent one from being concluded at all. Here the relatively high variance of the noise in the political, economic, and military environments surrounding the treaty leads to a relatively long initial duration, just as it did in the case of the NPT.

What did the parties in fact learn during the trial period? All of the parties learned its value in fostering scientific progress and cooperation. With respect to distributional issues, much was learned as well. The Antarctic Treaty was one of the first treaties signed by both the United States and the Soviet Union during the Cold War. The treaty's provisions represented a radical departure from traditional notions of sovereignty by permitting open inspection of every ship and installation on the continent. In 1959 no one knew if the provisions would actually be followed. They were. As of 1993, over 120 uneventful inspections had transpired. Overall, the two super powers learned that participation in the Antarctic Treaty did not disrupt the balance of power between them. The claimant states learned that adherence to the treaty reduced the need for them to adopt defensive and/or confrontational positions with respect to their claims. They also came to realize that the treaty offered them the best protection for their claims given the advent of super power interest in Antarctica. Finally, the uncertainty about resource extraction was precluded by later agreements focusing on that issue.

The outcome of the parties' learning during the trial period was general satisfaction with the treaty. That the parties viewed it as a success is made clear by their allowing it to continue without modification after 1991 and by statements made prior to 1991.[62]

The Outer Space Treaty

While the Antarctic Treaty resembles the NPT in terms of its duration and renegotiation provisions, the Outer Space Treaty does not.[63] The Outer Space Treaty has an implicit indefinite duration and no renegotiation provision. The treaty forbids the placement of weapons of mass destruction into orbit, onto celestial bodies, or in outer space. It mandates that all celestial objects be free of military bases, installations, fortifications, or military maneuvers and that they be used only for peaceful purposes. Outer space "is not subject to national appropriation"; thus all states will be allowed to participate equally in outer space research and exploration if they so desire, and all are required to inform the UN Secretary-General and the

62. See, for example, the statements in Beck 1986, 167–68, and his citation of a 1984 UN study.

63. The official name of the Outer Space Treaty is Treaty of Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies, 1967, *UN Treaty Series* 610 (2):8843.

https://doi.org/10.1162/002081801511140586 Published online by Cambridge University Press

scientific community of the nature and results of such activities. It entered into force in 1967 for ninety-one states, including the United States and the Soviet Union.

In the midst of the Cold War, after Sputnik raised the specter of Soviet military superiority, the Western powers sought to demilitarize space. Initially, the Soviet Union wanted linkage to other disarmament issues—in particular, elimination of U.S. foreign bases that were home to short- and medium-range missiles. The Western powers rejected linkage, which delayed an agreement for some time. Eventually, as U.S. space capability increased (thereby increasing the gains to the Soviets from the treaty) and after observing the success (and public relations benefits) of the Limited Test Ban Treaty and the Antarctic Treaty, the Soviet stance changed, which led after some relatively quick negotiations to the signing of the Outer Space Treaty.

The Outer Space Treaty's substantive provisions were modeled on those of the Antarctic Treaty. However, unlike the Antarctic Treaty and the NPT, the agreement lacks flexibility.[64] I argue that the uncertainty about the distribution of gains was much lower in the case of the Outer Space Treaty than in the Antarctic Treaty or the NPT and that this difference accounts for the absence of a renegotiation provision in the Outer Space Treaty (hypothesis 1).

The Outer Space Treaty essentially extended well-known principles of international law to a new context. The parties were all familiar with how these principles worked—for example, in the case of the high seas. No resource issues were looming in the background of the Outer Space Treaty, as there were in the Antarctic case; rather, the treaty governed issues of demilitarization.[65] There were also no long-standing, mutually incompatible territorial claims to be worked around. And, importantly, the number of major players with, or likely to acquire, the capability to put weapons in space was much smaller (unlike the NPT). Moreover, because the era of space exploration was just beginning, the major players had not yet spent much on the research and development of space-based weapons, so losses in the form of sunk costs were small and easily determined.

Finally, the fundamental benefit to the major players in the Outer Space Treaty was the reduction in uncertainty regarding the future development and deployment of space-based weapons. Neither the United States nor the Soviet Union wanted to wake up one day to find nuclear weapons orbiting overhead. In analyzing Article IV, which proscribes the placement of nuclear weapons and other weapons of mass destruction in orbit around the earth and on celestial bodies, Bess C. M. Reijnen argues that the United States and the Soviet Union intended to maintain the balance of power in place at that time.[66] The benefits from locking in the status quo in terms of its effect on reducing uncertainty are clear in advance, with both sides clearly gaining to the extent that they are risk averse.

64. Article XV of the treaty states that any treaty party may propose amendments. Importantly, however, states may choose not to accept and be bound by an amendment.

65. See fn. 68.

66. Reijnen 1992. For more on this point, see U.S. Department of State 1960, 2:976–77.

https://doi.org/10.1162/002081801511140586 Published online by Cambridge University Press

Further evidence that the states strongly desired to reduce uncertainty about the distribution of benefits comes from the Soviet Union's insistence that international organizations not have the same status as states under the treaty. Peter Malanczuk states that, during negotiations, the Soviets were opposed to placing international organizations on the same footing as states.[67] The Soviet Union had a lot fewer such organizations and knew that it would gain a lot less than the United States if such organizations became equal partners. Put somewhat differently, the Soviets were uncertain about future control of international organizations (more concerned about this than about their control of other states given the dominance of Western states in most international organizations) and so wanted their influence in the treaty reduced. The United States ultimately acceded to this desire.

All of the factors just discussed conspired to reduce the uncertainty about the distributions of gains under the treaty. As hypothesis 1 suggests, when uncertainty about the distribution of gains is low, so is the value of renegotiation; states opt instead for a nonrenegotiated agreement. Comparing the Antarctic Treaty and the NPT with the Outer Space Treaty illustrates this comparative static result in practice.[68]

Conclusion

The credibility of commitments in the face of uncertainty requires a trade-off between flexibility and constraint. I focus on a particular form of uncertainty—a one-shot uncertainty surrounding the division of gains from an agreement—and on a particular kind of flexibility—the combination of limited duration followed by a single renegotiation.

Some might argue that we do not observe much reneging on international agreements empirically and draw the implication that pretty much any duration and renegotiation provisions would do in a given context. In my model, this is clearly false. I calculate the lost utility from failure to choose the optimal provisions. In the case of small deviations, the lost utility takes the form of unrealized potential gains from the agreement. In the case of large deviations, it takes the form of reneging or failure to initiate the agreement at all. Empirically, a selection process similar to that described by George Downs, David Rocke, and Peter N. Barsoom regarding state

67. Malanczuk 1997.

68. It is important to note that the Moon Agreement of 1979 (officially called the Agreement Governing the Activities of States on the Moon and Other Celestial Bodies) does address issues characterized by great distributional uncertainty, for example, the exploration and exploitation of natural resources, and does have a provision for renegotiation. Reijnen writes, "A provision unique in space law is the one that a review conference [after a period of ten years] *shall* (obligatory) not only consider the question of potential revision of the [Moon Agreement] but also the question of implementation of article 11 . . . on the international regime . . . to govern the exploitation of the natural resources of the Moon." Reijnen 1992, 317.

https://doi.org/10.1162/002081801511405 86 Published online by Cambridge University Press

compliance in international agreements is at work with respect to duration and renegotiation provisions.[69] The reason we do not observe much reneging in actual agreements is in part because their duration and renegotiation provisions have been chosen in ways that act to minimize this costly behavior. For example, the reason we do not observe agreements failing because of uncontrollable economic circumstances is that agreements in areas subject to such disruptions will tend to be of short duration—short enough that states experiencing sudden losses will stay with the agreement until it is renegotiated rather than renege.

My analysis also responds to some recent game-theoretic work in international relations. James D. Fearon points out a weakness in current theories of international cooperation that focus primarily on the enforcement of international agreements while ignoring the bargaining that generates the agreements in the first place. Fearon's model integrates the bargaining over the terms of an agreement into the cooperation problem. This formulation reveals that the same shadow of the future that allows self-enforcing agreements also makes reaching an agreement more difficult by increasing the distributional effects of the selection of the initial equilibrium.[70]

I provide a theoretical justification for the fact that states often integrate planned renegotiation into international agreements. Building in renegotiation at the start reduces the distributional impacts of the initial equilibrium selection but does not reduce the shadow of the future that supports enforcement of the agreement as a whole since violations in the initial period can still be punished by noncooperation in future periods. Of course, allowing renegotiation adds the possibility that one party may drop out at the end of a particular finite-duration contract. Indeed, in the case of the International Coffee Agreement, the United States did just that. Nonetheless, if the probability of continuation is sufficiently high, finite-duration contracts linked by renegotiation may represent the real-world solution to Fearon's theoretical dilemma.

Finally, my analysis extends the neoliberal international relations literature beyond its current focus on the general issue of how cooperation can emerge. Both Robert Axelrod and Kenneth Oye suggest devices such as lengthening the shadow of the future, practicing reciprocity, and improving recognition capabilities; Stephen D. Krasner looks at the role of international regimes in promoting and maintaining cooperation; and Robert O. Keohane argues that regimes reduce the transactions costs associated with international cooperation.[71] This literature has opened up the central questions of international politics. It has done so, however, only by moving well away from any detailed analysis of specific institutional arrangements or

69. Downs, Rocke, and Barsoom 1997.

70. Fearon 1998. Fearon's work as well as Morrow's show how distributional differences can undermine cooperation in significant ways. Morrow 1994. These works are in part a response to Grieco and Krasner, who have rightly argued that neoliberals tend to emphasize enforcement issues and ignore distributional issues. See Grieco 1988; and Krasner 1991.

71. See Axelrod 1984; Oye 1986; Krasner 1983; and Keohane 1984.

https://doi.org/10.1162/002081801511140586 Published online by Cambridge University Press

questions of institutional design. In other words, this literature has failed to investigate the precise mechanisms through which cooperation can emerge.[72] There is no inherent reason, however, why the broader political issues cannot be considered simultaneously with the specific institutional arrangements designed to address them in ways that illuminate both the broader relationships and the institutionalization itself. My goal in investigating duration and renegotiation provisions has been to deepen our understanding of international cooperation by asking about specifics.

Appendix A: Equilibrium of the Two-period Game

This appendix provides a characterization of an equilibrium of the game described earlier. Additional notation is as follows: δ is the (common) discount factor and β_t denotes the parties' common beliefs at period t. Let $Q_1 = 1$ if party 1 reneges ("quits") on the duration provisions (and hence the agreement) and let $Q_1 = 0$ otherwise; let Q_2 be similarly defined for party 2. Let $Q = 1$ if $Q_1 = 1$ or $Q_2 = 1$, and let $Q = 0$ otherwise. $\Pr(Q_n = x)$ is the probability that $Q_n = x$.

An equilibrium of the two-period game is characterized by the following:

The parties' common beliefs in period 1, $\{\beta_1{:}f(u),\ h(m + \varepsilon)\}$;[73]
their Nash demand game strategies in period 1, $\{m_n{:}m_n \in [0, g]\}$;
their agreement-type strategies in period 1, $\{AT_n{:}[0, g] \times (NA, NR, R)\}$;
their common beliefs in period 2, $\{\beta_2{:}h(m + \varepsilon) \mid Y_{1,1}, Y_{2,1}; f(u)\}$;
their reneging strategies in period 2, $\{Q_n{:}[0, g] \times \{(NR, NR), (R, R)\} \times \Re^2 \times (0, 1)\}$; and
their Nash demand game strategies in period 2, $\{m_n^*{:}[0, g] \times \{(NR, NR), (R, R)\} \times \Re^2 \times (0, 1) \times [0, g]\}$.[74]

The parties' strategies and beliefs are consistent with one another and fulfill the following conditions:

1. At the Nash demand game stage in period 1, the parties' strategies satisfy the Nash bargaining solution, (that is, $m = g/2$).
2. At the agreement-type choice stage, each party maximizes its expected utility by choosing among the following: *NA, NR, R*. Their expected utilities are defined as follows:

72. Likewise, the tools of game theory have been directed mainly at abstract questions that emphasize cooperation rather than institutional design as the dependent variable.

73. Recall that $f(u_{1,1}) = f(u_{2,1}) = f(u_{1,2}) = f(u_{2,2}) = f(u)$.

74. For notational convenience, henceforth I replace m^* with $m + a$. Recall a is the adjustment factor incorporated at the renegotiation stage of a renegotiated agreement, and $E(m + \varepsilon + a \mid Y_{1,1}, Y_{2,1}) = m$. In other words, given my assumption of constant bargaining power, the parties strive for the original expected division of gains when they conclude their new agreement, but this time around they have more information, which, of course, they use.

https://doi.org/10.1162/002081801511405 86 Published online by Cambridge University Press

a. In the absence of an agreement, the expected utilities of the two parties are

$$\text{Party 1: } E(U_{NA}) = E[U(u_{1,1}) + \delta U(u_{1,2})]$$

$$\text{Party 2: } E(U_{NA}) = E[U(u_{2,1}) + \delta U(u_{2,2})].$$

b. In the presence of a single, two-period agreement (a nonrenegotiated agreement), their expected utilities are

$$\begin{aligned}\text{Party 1: } E(U_{NR}) = E[&U((m + \varepsilon) - k_1 + u_{1,1}) \\ &+ \delta(\Pr(Q = 0)U((m + \varepsilon) + u_{1,2}) \\ &+ \Pr(Q_1 = 1)U((m + \varepsilon) + a - k_2 - c + u_{1,2}) \\ &+ \Pr(Q_2 = 1)U((m + \varepsilon) + a - k_2 + u_{1,2}))].\end{aligned}$$

$$\begin{aligned}\text{Party 2: } E(U_{NR}) = E[&U((g - (m + \varepsilon)) - k_1 + u_{2,1}) \\ &+ \delta(\Pr(Q = 0)U((g - (m + \varepsilon)) + u_{2,2}) \\ &+ (\Pr(Q_1 = 1)U((g - (m + \varepsilon)) - a - k_2 + u_{2,2}) \\ &+ (\Pr(Q_2 = 1)U((g - (m + \varepsilon)) \\ &- a - k_2 - c + u_{2,2}))],\end{aligned}$$

where

$$\begin{aligned}\Pr(Q_1 = 1) = \Pr(&E[U((m + \varepsilon) + a - k_2 - c + u_{1,2})|Y_{1,1}, Y_{2,1}] \\ &> E[U((m + \varepsilon) + u_{1,2})|Y_{1,1}, Y_{2,1}]);\end{aligned}$$

$$\begin{aligned}\Pr(Q_2 = 1) = \Pr(&E[U((g - (m + \varepsilon)) - a - k_2 - c + u_{2,2})|Y_{1,1}, Y_{2,1}] \\ &> E[U((g - (m + \varepsilon)) + u_{2,2})|Y_{1,1}, Y_{2,1}]);\end{aligned}$$

and

$$\Pr(Q = 1) = \Pr(Q_1 = 1) + \Pr(Q_2 = 1);^{75}$$

and where, given my assumption of constant bargaining power, a satisfies

$$E(m + \varepsilon + a|Y_{1,1}, Y_{2,1}) = m.$$

75. Both parties would never want to renege simultaneously given that $g > 0$ and $c \geq 0$. If $g > 0$ and $c \geq 0$, at least one of the parties will always be happy with a nonrenegotiated agreement at the end of the first period (that is, its actual gain will be greater than or equal to its expected gain) since g is split between the two parties.

https://doi.org/10.1162/002081801511140586 Published online by Cambridge University Press

c. In the presence of two one-period agreements (a renegotiated agreement), their expected utilities are

$$\text{Party 1: } E(U_R) = E[U((m + \varepsilon) - k_1 + u_{1,1}) + \delta U((m + \varepsilon) + a - k_2 + u_{1,2})].$$

$$\text{Party 2: } E(U_R) = E[U((g - (m + \varepsilon)) - k_1 + u_{2,1}) + \delta U((g - (m + \varepsilon)) - a - k_2 + u_{2,2})].^{76}$$

3. The parties use Bayes' rule to obtain a posterior density for $h(m + \varepsilon)$ conditional on the outcomes in period 1.

4. At the reneging stage, each party maximizes its expected utility by choosing either to quit the agreement or not. Party 1 will quit if

$$E[U((m + \varepsilon) + a - k_2 - c + u_{1,2})|Y_{1,1}, Y_{2,1}] > E[U((m + \varepsilon) + u_{1,2})|Y_{1,1}, Y_{2,1}];$$

otherwise, it will not quit. Party 2 will quit if

$$E[U((g - (m + \varepsilon)) - a - k_2 - c + u_{2,2})|Y_{1,1}, Y_{2,1}] > E[U((g - (m + \varepsilon)) + u_{2,2})|Y_{1,1}, Y_{2,1}];$$

otherwise, it will not quit.

5. At the Nash demand game stage in period 2, the parties' strategies satisfy the following:

Recall $m^* = m + a$. The parties choose m^* s.t.

$$E(m + \varepsilon + a|Y_{1,1}, Y_{2,1}) = g/2.$$

Appendix B: Construction of Figures 1 and 2

This appendix describes the construction of the solutions to the discretized version of the two-period model that underlies Figures 1 and 2. The equations indicating the expected utility for each of the parties under no agreement, one two-period agreement, and two one-period agreements appear at the end of the appendix. In the simulations for Figures 1 and 2, I allow the outcome shock, u, to take on the values –2, 0, and 2. The agreement shock, ε, can take on these same three values so that $I = J = L = Z = 3$. As described in the text, in Figure 1 I start with probabilities of (0.3,0.4,0.3) for the three values of ε and then symmetrically

76. Given my assumption that states can negotiate a new agreement in period 2 following reneging by one state, a state will renege if it believes the difference between its payoff under a new agreement and its payoff under the second period of the two-period agreement exceeds the sum of the reneging and "re"negotiation costs. States have no incentive to renege on an agreement that includes renegotiation since doing so does not improve their bargaining power and so does not affect their outcomes in period 2.

https://doi.org/10.1162/002081801511405586 Published online by Cambridge University Press

increase the probabilities of the two nonzero values until I end up with probabilities of (0.5,0.0,0.5). The probabilities for u remain constant at (0.3,0.4,0.3). In Figure 2 I start with probabilities of (0.0,1.0,0.0) for the three values of u and then symmetrically increase the probabilities of the two nonzero values until I end up with probabilities of (0.5,0.0,0.5). In this case the probabilities for ε are held constant at (0.3,0.4,0.3).

The base outcome b, not shown in the equations at the end of the appendix, is set to 20. Including the base outcome has no effect on the results but allows the use of fractional-root utility functions by keeping all of the outcome values positive. The gain, g, is set equal to 8 and is assumed to be equally divided between the two parties in expectation, so that $m = 4$. I employ a cube-root utility function so that U equals the cube root of the outcome in each period. Contracting costs, k_1, for the original agreement are set to 1.0, renegotiation costs, k_2, are set to 0.5, and the costs of reneging, c, are set to 1.0.

The behavior of the adjustment parameter, a, depends on the realized outcomes of the two parties. In the discrete version of the model, some combinations of outcomes uniquely identify the agreement shock, ε. In these cases, states that have chosen to renegotiate can choose the adjustment factor, a, to completely undo the shocks. Other pairs of outcomes do not uniquely identify the agreement shock because they can be obtained through more than one combination of agreement and outcome shocks. For example, if both states observe an outcome of 24 in the first period, this can occur in three ways: (1) the outcome and agreement shocks are all zero; (2) state 1's outcome shock is 2, the agreement shock is –2, and state 2's outcome shock is –2, so that the outcome shocks cancel the agreement shock; or (3) state 1's outcome shock is –2, the agreement shock is 2, and state 2's outcome shock is –2, so that the outcome shocks again cancel the agreement shock. In such cases, the optimal adjustment factor is a weighted average of the optimal adjustment factor associated with each combination of outcome and agreement shocks that is consistent with the observed outcomes, where the weights are probabilities of each combination.

In those cases in which the parties can infer the value of ε from the observed values of the outcomes in period 1, the party for which the agreement shock equals –2 may choose to renege. The party receiving the +2 agreement shock has no incentive to renege since doing so would make it worse off.

In the simulation each party decides unilaterally whether to renege in each observationally distinct state of the world that can arise at the end of period 1. Each party compares the expected cost of reneging, $c + k_2$, to the benefit, which is 2 under the assumption that reneging is followed by negotiation of a new single period agreement for period 2, in which g is divided equally between the two parties. If the benefits of reneging exceed the costs for one party, that party reneges. When this occurs, the other party bears the additional renegotiation cost, k_2.

The almost all-or-nothing learning described in this appendix is a result of the discrete structure adopted for the simulations. It has no effect on the qualitative features of the model. In the continuous case, the parties would learn something from every realized pattern of outcomes but never enough to completely resolve the uncertainty after a single period.

Equations for Calculating the Expected Utilities

This section contains the equations for calculating the expected utilities for party 1; party 2 is symmetric.

In the absence of an agreement, the expected utility is calculated as follows:

https://doi.org/10.1162/002081801511140586 Published online by Cambridge University Press

$$E(U_{NA}) = \sum_i \sum_j \Pr(u_{1,1} = u_i)\Pr(u_{1,2} = u_j)[U(u_i) + \delta U(u_j)].$$

In the presence of a single, two-period agreement, the expected utility is calculated as follows:

$$\begin{aligned} E(U_{NM}) = \sum_i \sum_j \sum_l \sum_s \Pr(u_{1,1} = u_i)\Pr(u_{1,2} = u_j)\Pr(\varepsilon = \varepsilon_l)\Pr(u_{2,1} = u_s) \\ \times [U(u_i + (m + \varepsilon_l) - k_1) + \delta[D_1 U(u_j + (m + \varepsilon_l) + a - k_2 - c) \\ + D_2 U(u_j + (m + \varepsilon_l) + a - k_2) + (1 - D_1 - D_2)U(u_j + (m + \varepsilon_l))]], \end{aligned}$$

where

$$D_1 = 1[E(u_{1,2} + (m + \varepsilon_l) + a - k_2 - c)|Y_{1,1}, Y_{2,1}) > E(u_{1,2} + (m + \varepsilon_l)|Y_{1,1}, Y_{2,1})],$$

$$\begin{aligned} D_2 = 1[E(u_{2,2} + (g - (m + \varepsilon_l)) - a - k_2 - c|Y_{1,1}, Y_{2,1}) \\ > E(u_{2,2} + (g - (m + \varepsilon_l))|Y_{1,1}, Y_{2,1})], \end{aligned}$$

and a solves

$$E(m + \varepsilon + a|Y_{1,t}, Y_{2,t}) = m.$$

In the presence of two one-period agreements, the expected utility is calculated as follows:

$$\begin{aligned} E(U_R) = \sum_i \sum_j \sum_l \Pr(u_1 = u_i)\Pr(u_2 = u_j)\Pr(\varepsilon = \varepsilon_l) \\ \times [U(u_i + (m + \varepsilon_l) - k_l) + \delta U(u_j + (m + \varepsilon_l) + a - k_2)], \end{aligned}$$

where a solves

$$E(m + \varepsilon_l + a|Y_{1,t}, Y_{2,t}) = m.$$

References

Antarctic Treaty. 1959. *United Nations Treaty Series* 402:71 (5778).

Axelrod, Robert. 1984. *The Evolution of Cooperation.* New York: Basic Books.

Beck, Peter. 1986. *The International Politics of Antarctica.* New York: St. Martin's Press.

Bennett, D. Scott. 1997. Testing Alternative Models of Alliance Duration, 1816–1984. *American Journal of Political Science* 41 (3):846–78.

Bilder, Richard B. 1981. *Managing the Risks of International Agreement.* Madison: University of Wisconsin Press.

Boczek, Boleslaw A. 1984. The Soviet Union and the Antarctic Regime. *American Journal of International Law* 78 (4):834–58.

https://doi.org/10.1162/002081801511140586 Published online by Cambridge University Press

Bordo, Michael D., and Finn E. Kydland. 1995. The Gold Standard as a Rule: An Essay in Exploration. *Explorations in Economic History* 32 (4):423–64.

———. 1996. The Gold Standard as a Commitment Mechanism. In *Modern Perspectives on the Gold Standard,* edited by Tamin Bayoumi, Barry Eichengreen, and Mark P. Taylor, 55–100. Cambridge: Cambridge University Press.

Bunn, George, and Charles N. Van Doren. 1991. Options for Extension of the NPT: The Intentions of the Drafters of Article X.2. In *Options and Opportunities: The NPT Extension Conference of 1995,* edited by George Bunn, Charles N. Van Doren, and David Fisher. Programme for Promoting Nuclear Nonproliferation Study No. 2. Highfield, Southampton: University of Southampton.

———. 1992. *Two Options for the 1995 NPT Extension Conference Revisited.* Washington, D.C.: Lawyers' Alliance for World Security.

Chayes, Abram, and Antonia Handler Chayes. 1993. On Compliance. *International Organization* 47 (2):175–205.

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

Downs, George W., David M. Rocke, and Peter N. Barsoom. 1996. Is the Good News About Compliance Good News About Cooperation? *International Organization* 50 (3):379–407.

Dye, Ronald A. 1985. Costly Contract Contingencies. *International Economic Review* 26 (1):233–50.

Elster, Jon. 1989. *Nuts and Bolts for the Social Sciences.* Cambridge: Cambridge University Press.

Fearon, James D. 1998. Bargaining, Enforcement, and International Cooperation. *International Organization* 52 (2):269–305.

Gaubatz, Kurt Taylor. 1996. Democratic States and Commitment in International Relations. *International Organization* 50 (1):109–39.

Graham, Thomas, Jr. 1989. The Duration of the Nuclear Non-Proliferation Treaty: Sudden Death or New Lease on Life? *Virginia Journal of International Law* 29 (3):661–77.

Gray, Jo Anna. 1978. On Indexation and Contract Length. *Journal of Political Economy* 86 (1):1–18.

Grieco, Joseph M. 1988. Anarchy and the Limits of International Cooperation: A Realist Critique of the Newest Liberal Institutionalism. *International Organization* 42(3):485–507.

———. 1990. *Cooperation Among Nations: Europe, America, and Non-tariff Barriers to Trade.* Ithaca, N.Y.: Cornell University Press.

Harris, Milton, and Bengt Holmstrom. 1987. On the Duration of Agreements. *International Economic Review* 28 (2):389–406.

Hart, Oliver, and Bengt Holmstrom. 1987. The Theory of Contracts. In *Advances in Economic Theory,* edited by Truman F. Bewley, 71–155. Cambridge: Cambridge University Press.

Henkin, Louis. 1979. *How Nations Behave: Law and Foreign Policy.* 2d ed. New York: Columbia University Press.

Jackson, John H. 1992. Status of Treaties in Domestic Legal Systems: A Policy Analysis. *American Journal of International Law* 86 (2):310–40.

Jensen, Lloyd. 1974. *Return from the Nuclear Brink: National Interest and the Nuclear Nonproliferation Treaty.* Lexington, Mass.: Lexington Books.

Keohane, Robert O. 1984. *After Hegemony: Cooperation and Discord in the World Political Economy.* Princeton, N.J.: Princeton University Press.

Koremenos, Barbara. 1999. On the Duration and Renegotiation of International Agreements. Ph.D. diss., University of Chicago.

———. 2000. Bending but Not Breaking: Flexibility in International Financial and Monetary Agreements. Working Paper 1.73. Berkeley: University of California, Center for German and European Studies.

Koremenos, Barbara, Charles Lipson, and Duncan Snidal, eds. Forthcoming. Rational Designs: Explaining the Form of International Institutions. *International Organization* 55 (4). Special issue.

Krasner, Stephen D. 1991. Global Communications and National Power: Life on the Pareto Frontier. *World Politics* 43 (3):336–66.

Krasner, Stephen D., ed. 1983. *International Regimes.* Ithaca, N.Y.: Cornell University Press.

https://doi.org/10.1162/002081801511405 86 Published online by Cambridge University Press

Kunz, Josef L. 1939. The Problem of Revision in International Law. *American Journal of International Law* 33 (1):33–55.

Malanczuk, Peter. 1997. Actors: States, International Organisations, Private Entities. In *Outlook on Space Law over the Next 30 Years: Essays Published for the 30th Anniversary of the Outer Space Treaty,* edited by Gabriel Lafferranderie and Daphne Crowther, 23–36. Boston: Kluwer Law International.

Morgenthau, Hans J. 1978. *Politics Among Nations: The Struggle for Power and Peace.* 5th ed. rev. New York: Knopf.

Morrow, James D. 1994. Modeling the Forms of International Cooperation: Distribution Versus Information. *International Organization* 48 (3):387–423.

Nye, Joseph S. 1981. Maintaining a Nonproliferation Regime. *International Organization* 35 (1):15–38.

Osborne, Martin J., and Ariel Rubinstein. 1990. *Bargaining and Markets.* San Diego, Calif.: Academic Press.

Oye, Kenneth A., ed. 1986. *Cooperation Under Anarchy.* Princeton, N.J.: Princeton University Press.

Panofsky, Wolfgang K. H., and George Bunn. 1994. The Doctrine of the Nuclear Weapon States and the Future of Non-Proliferation. *Arms Control Today* 24 (6):3–9.

Peterson, M. J. 1988. *Managing the Frozen South: The Creation and Evolution of the Antarctic Treaty System.* Berkeley: University of California Press.

Quester, George. 1973. *The Politics of Nuclear Proliferation.* Baltimore, Md.: Johns Hopkins University Press.

Raustiala, Kal. 1997. Domestic Institutions and International Regulatory Cooperation: Comparative Responses to the Convention on Biological Diversity. *World Politics* 49 (4):482–510.

Reijnen, Bess C. M. 1992. *The United Nations Space Treaties Analyzed.* Gif-sur-Yvette Cedex, France: Editions Frontieres.

Rosendorff, B. Peter, and Helen V. Milner. Forthcoming. The Optimal Design of International Trade Institutions: The Sources and Impact of Escape Clauses. *International Organization* 55 (4).

Smith, Roger K. 1987. Explaining the Non-Proliferation Regime: Anomalies for Contemporary International Relations Theory. *International Organization* 41 (2):253–81.

Taubenfeld, Howard J. 1961. A Treaty for Antarctica. *International Conciliation* (531):245–322.

Treaty on the Non-proliferation of Nuclear Weapons. 1968. *United Nations Treaty Series* 729:161 (10485).

U.S. Arms Control and Disarmament Agency. 1967. *Documents on Disarmament 1967.* Washington, D.C.: ACDA.

U.S. Department of State. 1960. *Documents on Disarmament, 1945–1959.* Washington, D.C.: U.S. Government Printing Office.

———. 1991. *Foreign Relations of the United States.* Vol. 2, 1958–60. Washington, D.C.: U.S. Government Printing Office.

U.S. House of Representatives. 1985. Committee on Foreign Affairs. *Third Review Conference of the Treaty on the Non-Proliferation of Nuclear Weapons: Hearing Before the Subcommittees on Arms Control, International Security, and Science and on International Economic Policy and Trade.* 99th Cong, 1st sess., 1 August.

U.S. Senate. 1960. Committee on Foreign Relations. *The Antarctic Treaty: Hearings.* 86th Cong., 2d sess., 14 June.

Welsh, Susan B. 1995. Delegate Perspectives on the 1995 NPT Review and Extension Conference. *The Nonproliferation Review* 2 (3):1–24.

Wilson, Robert R. 1934. Revision Clauses in Treaties Since the World War. *American Political Science Review* 28 (5):901—909.

Young, Oran R. 1991. Political Leadership and Regime Formation: On the Development of Institutions in International Society. *International Organization* 45 (3):281–308.

https://doi.org/10.1162/002081801511405 86 Published online by Cambridge University Press