International Law and International Relations

Edited by Beth A. Simmons
and Richard H. Steinberg

CAMBRIDGE www.cambridge.org/9780521861861

actual outcome will consist of the base plus or minus some amount. For example, state 1's actual GNP might 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 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.[10] 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

10 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.

comparative static results remain.[11] 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 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

[11] I elaborate this point later.

$u_{1,t}$ 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.[12] 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

$$Y_{1,t} = m + \varepsilon + u_{1,t}$$
$$Y_{2,t} = g - (m + \varepsilon) + u_{2,t},$$

with associated expected values

$$E(Y_{1,t}) = m$$
$$E(Y_{2,t}) = g - m.$$ [13]

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.

[12] 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.

[13] 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.

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. The bargaining outcome in both periods is exogenous and satisfies the Nash bargaining solution.[14]

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

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

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*) × {*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 2 as type $-\varepsilon$. Additionally, I impose the following restriction: The set of punishment 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).[15]

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

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

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.[16] Note that this is precisely the reverse of what happens under hypothesis 1 when the relative variance of ε increases.

Figures 15.1 and 15.2 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.[17] 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. ***

[16] 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.

[17] 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.

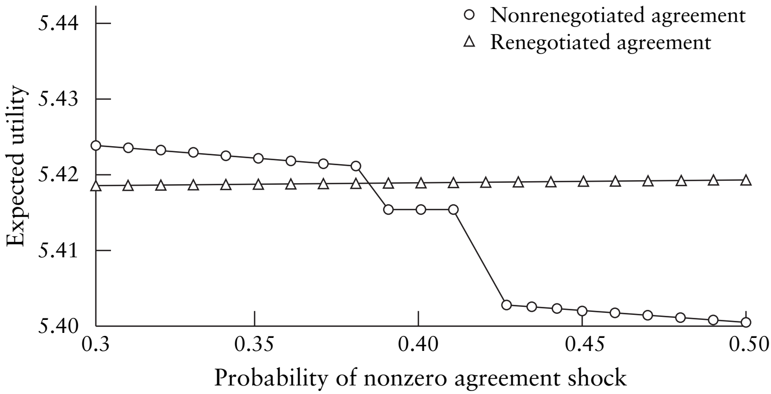

FIGURE 15.1. Increasing the variance of the agreement shock.

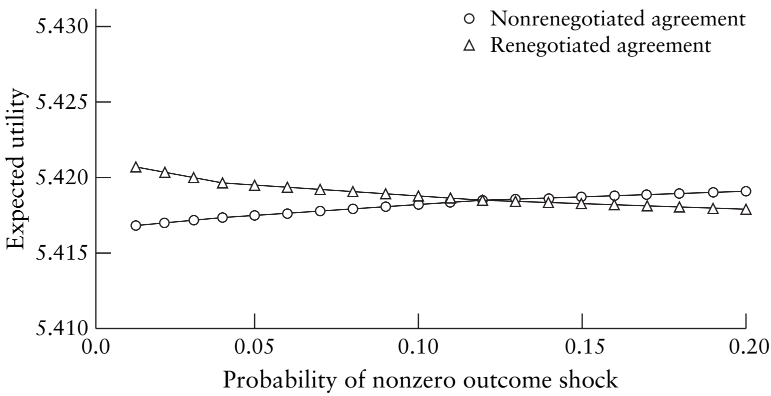

FIGURE 15.2. Increasing the variance of the noise.

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 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 15.1, 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 15.1. The two lines in Figure 15.1 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 (the line with the triangles) ***.[18] 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.[19] As a result, the states in my model change their agreement-type choice when the degree of agreement uncertainty exceeds a certain level.[20]

Figure 15.2 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).[21] This reflects the decreasing value of renegotiation as the variance of u increases, which causes u to

[18] Figures 15.1 and 15.2 both omit the value of no agreement. In each case it always lies between 5.1 and 5.2. In Figure 15.1 it is flat since it is unaffected by the variance of ε. In Figure 15.2 it declines with the variance of u.

[19] The sudden drops in the expected utility of the nonrenegotiated agreement in Figure 15.1 result from states deciding to renege in particular states of the world when the variance of the agreement shock reaches a certain level.

[20] 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 15.1 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.

[21] 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 15.2 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.

"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.[22]

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

* * *

This case study makes the following points. First, it reveals the empirical importance of this structure of duration and renegotiation provisions. The NPT is arguably one of the more important international agreements of this century. An understanding of its provisions necessarily informs any

[22] 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 15.1, 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 15.1 would then cross to the left of where they do in a world where states care only about absolute gains, which means that adding in concerns about relative gains *increases* the set of cases in which the states choose to renegotiate.