International Institutions: Weak Commitments and Costly Signals

Lisa L. Martin
Department of Political Science
University of Wisconsin-Madison
llmartin3@wisc.edu

February 2010

Draft
Two of the major mechanisms by which international institutions might influence state behavior are creating commitments and serving as costly signals. In the context of crisis bargaining, James Fearon (1997) influentially distinguished between these two mechanisms as “tying hands” and “sinking costs.” His analysis clarified the distinction between commitment and signaling, which had often been muddled in the international relations literature. Commitment and signaling are now typically seen as alternative mechanisms, in contexts of bargaining or cooperation, and generally modeled separately.

While it is important, for purposes of conceptual clarity and theory development, to understand the differences between *ex ante* signals and *ex post* incentives that create commitments, in practice actual mechanisms might both sink costs and tie hands. For example, Branislav Slantchev (2005) critiques Fearon’s argument by observing that military mobilization is both costly and changes the chances of victory, so that it is simultaneously a signaling and commitment mechanism. He develops a model based on this insight, and finds that some of Fearon’s major results – such as the lack of a bluffing equilibrium – do not stand up.

By creating costs associated with reneging on agreements, whether reputational or other costs, institutions can create commitments that allow for mutually-beneficial cooperation that could not be sustained outside an institution. However, in many cases institutions cannot change incentives so dramatically that they provide an absolute commitment. For example, military alliances can create incentives to provide mutual assistance in cases of conflict. However, under duress some alliance members find that these incentives are in fact not high enough for them to come to an ally’s assistance, so that they do not live up to their alliance obligations. The rate of reneging on alliance
commitments is in fact quite high, with alliance members reneging on their commitments in times of war about 25% of the time (Leeds 2003). In general, while institutions such as alliances can create commitments, they usually do not have the resources or authority to deter all members from reneging under all circumstances.

I argue, therefore, that we should think of international institutions as “weak commitment” devices. They change cost structures in such a way that they can allow cooperation to emerge that could not emerge in their absence. However, they cannot typically change cost structures sufficiently to deter all reneging on commitments. I develop a model of institutions as commitment devices in the presence of uncertainty about whether they are actually influential enough to prevent reneging on commitments. This model allows us to specify the conditions under which institutions as weak commitment devices can allow cooperation to emerge, and when we will observe states reneging on their commitments.

I then suggest that institutions, beyond changing ex post incentives to enhance commitments, can also be ex ante costly to join and so simultaneously serve as signals, analogous to Slantchev’s argument about military mobilization. I add this costly signaling function to the weak commitment model. This analysis gives rise to expectations about the conditions under which we will observe cooperation, institutions, and reneging; and it has implications for preferences over institutional design.

The first section of this paper provides a brief summary of the literature on institutions as signaling devices. The second develops the weak commitment model and then adds costly signaling to it. The third section focuses on the differences that emerge when signaling is added, discussing the costs and benefits associated with signaling and
who, therefore, is likely to demand that institutions be costly to join. The final section draws out the implications of this analysis for institutional design and further empirical research.

**Institutions as Signaling Devices**

The modern literature on international institutions tends to focus on their commitment properties. Going back to the original work on international regimes (Krasner 1982) and Robert Keohane’s seminal study of institutions (Keohane 1984), scholars have concentrated on the ability of institutions to either coordinate policy or to put in place monitoring and decentralized enforcement mechanisms that raise the costs to reneging on commitments. This tradition has continued in more recent work, as for example in Andrew Guzman’s study of why states comply with treaties (Guzman 2008). Guzman, writing from the perspective of a legal scholar, argues that the standard “3 R” mechanisms of reputation, reciprocity, and retaliation explain why international law can be effective in committing states to live up to its terms. Other international legal scholars, such as John Setear (2002) have also considered the use of institutions as signals, although Setear has focused more on domestic than international institutions in this capacity.

As mentioned in the introduction, Fearon (1997) clarified the previously muddled distinction between commitment (tying hands) and signaling (sinking costs). A number of other authors have further developed the signaling side of the story. James Morrow has modeled alliances as both commitment and signaling devices. Morrow (1994) argues
that alliances enhance commitment by increasing the ability of allies to fight together, and act as signals because they involve sunk peacetime costs. He derives propositions about when alliances will form, when they will be credible, and when they will deter. Morrow (2000) focuses more directly on the signaling properties, asking why “writing down” an alliance has any impact on its ability to deter attacks. Drawing on standard costly signaling models (as I do in this paper), he argues that alliances can only effectively deter if the signals that they send require their members to bear costs. Morrow points to the process of military coordination within alliances as the major source of peacetime costs, as policy coordination itself is costly; and military coordination can leave an ally more vulnerable if it eventually must fight alone (Morrow 2000, 70).

Jana von Stein (2005, 2008) has also studied the signaling or screening properties of international agreements, directing her efforts toward careful statistical modeling of ratification and compliance decisions. Von Stein (2005) examines Article VIII of the IMF Treaty, which commits states that sign it to avoid certain currency practices such as interference with payments and discrimination among foreign currencies. Simmons (2000) had previously asked which states complied with their commitments under Article VIII, and found that the primary determinant was the ease with which they were able to comply.

Von Stein builds on Simmons’ work by explicitly modeling the selection process by which states choose to enter Article VIII. She finds that Art. VIII seems to act purely as a screening device. That is, it separates out those states that would find it difficult to comply with its terms, and adds no further commitment beyond the screening function.
She concludes that “the international legal commitment has little constraining power independent of the factors that lead states to sign” (von Stein 2005, 611). In other words, the only effect of Art. VIII comes through the *ex ante* costs of entering it, which serve to screen out those states that will not be able to comply at low cost. Von Stein does not go into much detail about the nature of these costs or the source of their discriminatory power, but the evidence clearly points to a signaling function. The model in this paper allows for such an *ex ante* costly signaling function, while combining it with a weak commitment capacity. Von Stein (2008) considers ratification of international environmental agreements, in particular that UN Framework Convention and the Kyoto Protocol. Again, selection into these agreements is a prominent part of the causal mechanism, suggesting a signaling function.

Other authors have touched on the signaling properties of international institutions and agreements. Of note, Alexander Thompson (2006, 2009) discusses how the structure of the UN Security Council allows it to send effective signals when a state is attempting to coerce another. The use of the Security Council allows coercing states to strategically transmit information, so that the level of international support for the use of force becomes closely tied to Security Council approval. Yoram Haftel (2007) considers bilateral investment treaties (BITs), asking whether they serve primarily as credible commitments or signals. Developing countries sign BITs hoping that they will increase the flow of foreign direct investment (FDI). Haftel’s statistical analysis suggest that the commitment function of BITs may be more important than their signaling function, as FDI flows respond only to BITs that are mutually ratified, not to BITs that have merely been signed. Hyde (2006) analyzes international election monitoring as a costly signal,
showing why it rapidly evolved into a widely-observed norm. Finally, James Walsh (2007) considers the general question of whether states are engaged in signaling games, focusing on the relationship between the United States and Soviet Union under Gorbachev. He argues that Gorbachev was sending costly signals to the United States, but that understanding the sources and consequences of these signals requires integrating domestic political conflict into the analysis.

It is also of relevance to note discussions of institutions as signals in settings other than international politics. A number of authors have looked at democratic institutions as sources of costly signals (Mansfield and Pevehouse 2008; Schultz 1999; Hafner-Burton, Mansfield, and Pevehouse nd). Moving more into the realm of economics, Friedel Bolle’s work on corporate governance systematically studies governance institutions as signaling devices (Bolle and Braham 2003; Bolle 2002). Bolle emphasizes the potentially negative normative consequences of relying on corporate governance as a signal, as unless the signaling costs are precisely calibrated they can decrease social welfare. In the model in this paper, I likewise find that “fine-tuning” the costs of signals is essential if they are to function effectively. Baglioni (2008) builds on Bolle’s work by considering corporate governance as both a signal and a commitment device.

Model

In this section, I develop two simple games of incomplete information. Both involve a state, A, that decides whether to join an institution. A second state, B, then decides whether to cooperate with A. If B cooperates, A then chooses whether to
cooperate or to renege. A can be either a reliable or an unreliable type, distinguished by the fact that the reliable type receives a higher payoff from cooperation. For the unreliable type, the benefits of cooperation are not high enough to prevent reneging. B pays a price if it chooses to cooperate and then A reneges. I first develop a model where an institution functions as an *ex post* commitment device, then add an *ex ante* signaling function to the institution.

*Commitment model.* If international institutions function solely as commitment devices, they impose a cost on members that do not live up to the terms of their commitments – i.e., that renege. Other members of an institution find such reneging costly. Consider, for example, members of the WTO deciding whether to admit a potential new member such as China. The WTO has in place monitoring and enforcement mechanisms that impose costs on members if they violate WTO rules. These mechanisms, if strong enough, should induce members to live up to their commitments. However, reneging nevertheless does sometimes occur, and it imposes costs on other members as their access to markets in the violating country is limited.

To model this situation in a simple manner, consider a country A that is deciding whether to join an institution such as the WTO. A is aware that the institution has the capacity to impose costs if A joins and then reneges on commitments. A second country, B, has to decide whether to cooperate with A, whether or not A joins the institution. B is uncertain of the benefits that A will reap from cooperation; in particular, whether the benefits are high enough to prevent A from reneging on its commitments. (I assume that A has complete information about its own payoffs.) With probability p, A is a “reliable”
type, meaning that its payoffs from cooperation are high. Figure 1 shows the sequence of moves in this game.

[Fig. 1 about here]

If B chooses not to cooperate with A, they both receive a payoff of 0. If A and B both cooperate, B gets a payoff of $b_b$. A reliable A gets a payoff of $b_r$ from cooperation, and an unreliable A a payoff of $b_u$, with $b_r > b_u$. If B chooses to cooperate but A then reneges, B receives a payoff of $-d$ and A a payoff of $a$; $a > b_r > b_u > 0$. This ordering of payoffs indicates that, in the absence of an institution, the benefits of reneging are high enough that even a reliable A would reneg in the absence of an institution, imposing cost d on B. However, if A has joined the institution and then reneges, it pays a cost of $-c$, indicating the punishment cost imposed by the institution.

The equilibrium of this game will depend, in part, on how large the cost c is that the institution can impose on members that renege on their commitments. If c is very large, the institution would be a strong commitment device. In a strong commitment equilibrium, $b_u > a - c$, so that the prospect of punishment by the institution induces cooperation even from an unreliable A. In this case, it is an equilibrium for B to cooperate only if A joins the institution, knowing that institutional punishment will prevent A from reneging. Because even an unreliable A gets some benefits from cooperation, a strong commitment institution will lead to cooperation, leaving all states better off.

However, international institutions often do not have adequate enforcement capacity to prevent all reneging.\footnote{Why members do not always endow institutions with sufficient enforcement capacities to induce a strong commitment equilibrium is an interesting question. The answer likely has to do with greater uncertainty in} In a weak commitment equilibrium, which is of most
interest for purposes of this paper (and likely for the analysis of most actual institutions),
the punishment for reneging is sufficient to induce cooperation on the part of a reliable A
but not an unreliable type. That is, a weak commitment equilibrium exists if \( b_r > a - c > b_u > 0 \). Under these conditions, it is an equilibrium for both reliable and unreliable A to
join the institution. B will not be able to update its beliefs about A’s type, and so will cooperate if \( p > d/(b_b + d) \), meaning that the expected payoffs from cooperation are greater than zero.

A weak commitment equilibrium thus has the following properties: No cooperation emerges without institutions, because all A types would renege in the absence of institutional enforcement mechanisms (by assumption). All A types choose to join the institution; and B then cooperates probabilistically with A. A reliable A benefits from a weak commitment institution because it gains at least probabilistic cooperation. An unreliable A also benefits, because it (probabilistically) is able to induce B to cooperate and then will renege. State B’s expected payoff from a weak commitment institution is positive if the probability that A is reliable is high, or if the benefits of cooperation to B (\( b_b \)) are large. However, when B does cooperate, it runs the risk realizing the reneging payoff \(-d\).

Overall, pure commitment institutions that provide only weak commitments do allow cooperation to emerge under some conditions, by overcoming the temptation to renege for some types of states. However, they also have some undesirable properties.

the environment than I assume in this model. In particular, the benefits of cooperation (for both A and B) are likely to be stochastic rather than fully predictable and constant, as for example in Downs and Rocke (1997). If in one period realized benefits of cooperation are subject to a negative shock, but the punishment for reneging remains high, states may find it best to opt out of the institution entirely, leading to the collapse of cooperation. Thus, as Downs and Rocke argue, we should see some “imperfection” optimally built into institution. That is, states do not create strong commitment institutions because they are concerned that those strong enforcement capacities could undermine cooperation over the long term.
In particular, they allow unreliable states to bluff, joining the institution and then suckering other members into cooperation from which they will renege. Existing members of the institution can benefit, but only at the cost of exposing themselves to risk. In addition, because there is no *ex ante* cost to joining the institution, potential new members who are reliable cannot distinguish themselves from unreliable potential members, and so do not always gain the cooperation of existing members.

*Signaling model.* The dynamics of this game change substantially when a new member has to pay an up-front cost to join the institution. In this case, the institution serves as a costly signal, and under some conditions reliable potential members can now effectively distinguish themselves from unreliable types, allowing more mutually-beneficial cooperation to emerge. I will continue to focus on payoffs that lead to a weak commitment equilibrium, and the sequence of moves remains the same as in Fig. 1. However, now if A joins the institution it pays cost z on entry. I use Perfect Bayesian Equilibrium as the solution concept for this game, and characterize the equilibria of this signaling and commitment game in Table 1.

[Table 1 about here]

We can focus on two parameters to describe the equilibria: the B’s prior belief that A is reliable (p) and the size of the signaling cost (z). A low reliability situation exists if p<d(b+d) (left column in Table 1); otherwise we have a high reliability situation (right column). Signaling costs are low if b_r>a-c>z; they are moderate if b_r>z>a-c; they are high if z>b_r>a-c. The rows of Table 1 indicate different levels of signaling cost.
Consider first the situation where the costs of entering the institution are very high (bottom row). In this case, entry costs are so high that they outweigh the benefits of cooperation for even a reliable A. A pooling equilibrium then emerges, in which A does not join the institution and B does not cooperate. (Remember that, in the absence of an institution, even a reliable A succumbs to the temptation to renege.) Thus, if members of an institution miscalculate and set entry costs too high, they will deter even potential cooperating members from joining.

Consider next the situation of moderate entry costs, the middle row of the table. In this case, a separating equilibrium emerges. Reliable A’s, with their higher benefits of cooperation, will choose to bear this cost, but unreliable types will not. State B can now fully update its beliefs about A’s type, and will choose to cooperate with A if it joins the institution, but not otherwise. Signaling costs in this “sweet spot” have many desirable properties. They allow the fullest extent of cooperation between reliable states, and eliminate any potential for unreliable states to bluff and sucker existing members. State B is not subject to risk in this situation, because it can fully distinguish reliable from unreliable types.

Finally, consider the top row of the table, where signaling costs are low. In this case, the equilibrium that emerges depends on B’s prior belief about A’s type. If B believes that A is likely reliable (high p), a pooling equilibrium will emerge. Because B believes that A is reliable, B will cooperate whenever A joins the institution. And because the cost of joining the institution is low, even unreliable A’s will be willing to pay it. In this situation, an unreliable A will bluff and sucker B. However, if B believes that the probability that A is reliable is low, we find a semi-separating equilibrium. In this case,
B will not always cooperate because it faces a high chance of being suckered. Instead, if A chooses to join the institution, B’s only equilibrium response is to cooperate probabilistically (as indicated in Table 1). If A does not join the institution, B does not cooperate. Reliable A’s always join the institution; unreliable types join with some probability less than 1 (as indicated in the table).

Thus, setting entry costs low creates a number of inefficiencies. Reliable types can no longer fully distinguish themselves from unreliable types, so they cannot always induce B to cooperate, even when they pay entry costs. Unreliable types are also able to bluff, so that B sometimes has to bear the cost of A’s reneging. States designing institutions therefore face a dilemma when they make *ex ante* demands of new members. If these demands are set too high, they will deter entry and lose out on potentially beneficial cooperation. On the other hand, if they set entry costs too low, they will allow unreliable new members into the institution and allow themselves to be suckered, at least probabilistically.

*Comparing Signaling and Pure Weak Commitment.* From states’ perspectives, what difference does adding a signaling component to a pure commitment institution with weak enforcement capacities make? Tables 2, 3, and 4 get at this question by considering the difference in payoffs for our three types of states: unreliable potential members, reliable potential members, and existing members of the institution (state B). The cells in these tables indicate the difference in expected payoffs that occurs when an *ex ante* costly signal is added to a weak commitment institution.

[Tables 2, 3, and 4 about here]
Table 2 considers unreliable potential new members: states that would derive relatively little benefit from cooperation, and therefore are likely to renege on their commitments. Such states can only lose from demanding an entry cost. At best, they are indifferent. This occurs when state B believes that A is likely unreliable. In this case, no cooperation would emerge under a weak commitment institution, and little changes when costly signals are introduced. On the other hand, when B believes there is a good chance that A is reliable, an unreliable A unambiguously loses when it has to send a costly signal. Without signaling, in this circumstance unreliable A’s were able to bluff, inducing B into cooperating and then reneging on B. If the cost to join the institution is moderate or high, an unreliable A will not pay it and loses the opportunity to fool B. If the signaling cost is low, an unreliable A will be able to bluff, but is still slightly worse off than in the no-signaling situation because it has to pay the entry cost.

Table 3 looks at the situation from the prospective of a reliable potential member, a state A that will gain relatively high benefits from cooperation. The story is more complicated here, with a reliable A sometimes gaining from the opportunity to signal and sometimes losing. When B’s prior belief is that A is probably unreliable, a reliable A gains from the opportunity to send a costly signal, because it can now differentiate itself unreliable types and gain B’s cooperation. If the cost of the signal is set too high A will not gain these benefits, and is indifferent between a signaling institution and a pure weak commitment one.

However, when B believes that A is likely reliable, a reliable A can only lose from the addition of a signaling component, although the degree of loss varies depending on the cost of entry. When entry costs are low or moderate, so is a reliable A’s loss. As
in the weak commitment case, A cooperates with B; but it now has to “burn money” by paying the *ex ante* cost. If entry costs are high, a large loss results for a reliable A, as it will not bear this cost and cooperation will not materialize. Thus, we would expect that potential new members who anticipate relatively high benefits from cooperation would have varying attitudes toward demands that they pay an entry cost. If they see that this would allow them to differentiate themselves from unreliable states, they will happily pay these costs. But if existing members already believe that A is reliable, A will be opposed to any additional demands to pony up before entering the institution.

Finally, Table 3 considers the difference in payoffs for state B, an existing member of the institution. For empirical purposes, this table is likely the most useful, as existing institutional members will set the entry costs for new members. Table 3 therefore leads directly to hypotheses about institutional design. As indicated, existing members can gain from the introduction of costly signals, lose, or find themselves indifferent. The introduction of entry costs is in general helpful to existing members when they believe that there is a high probability that A is unreliable. As A more likely becomes reliable, entry costs could actually decrease B’s welfare, unless B is able to hit the sweet spot that creates a separating equilibrium.

If B sets entry costs too high, no cooperation will emerge. When potential members are believed unreliable, this makes no difference to B, as it would not have cooperated anyway. But when potential members are more likely reliable, high entry costs hurt existing members of the institution by keeping out new members and making cooperation impossible. Likewise, if B sets entry costs too low, the results are mixed. This can lead to a small gain in welfare for B, as occasional cooperation emerges with
reliable A’s; but this effect is partially offset by the fact that B will now open itself to possible exploitation by unreliable A’s. If B sets entry costs low and it is likely that A is reliable, the outcome from B’s perspective is the same as for pure commitment institutions – all A’s join and B always cooperates – so B is indifferent. Only when B sets entry costs at the appropriate intermediate level – not so high as to deter reliable A’s from entering, but high enough that unreliable A’s will not bear the cost – does B unambiguously gain from giving A the opportunity to send a costly signal.

**Implications**

This model, while quite stark and simple, gives rise to a number of empirical implications for the study of international institutions. Some implications regard states’ decisions to join existing institutions, and others the behavior of states once they are in an institution. A number of implications also arise about different types of states’ preferences over institutional design. In this section I will summarize some of the more prominent implications for empirical research, emphasizing those referring to institutional design.

First, consider pure weak commitment institutions – those that impose some cost for reneging, but that do not have sufficient enforcement capabilities to deter all reneging; and that do not impose an *ex ante* cost for joining the institution. When such institutions exist, there is no reason for states not to join them. The institution will then succeed in generating cooperation among reliable members (those who derive relatively high benefits from cooperation), but not among unreliable ones.
Weak commitment hypothesis 1: When institutions are costless to join, they will have a large membership.

Weak commitment hypothesis 2: When an institution is costless to join, its members will frequently renge on their commitments.

For purposes of this paper, I am more interested in the properties of signaling institutions. In these institutions, because they sometimes screen out unreliable potential members, reneging should be less common than in pure commitment institutions. In addition, the frequency of reneging should be a function of ex ante signaling costs. When such costs are low, reneging should be fairly common, although not quite as frequent as in pure commitment institutions. When signaling costs are moderate, they should fully screen out unreliable members, so that no reneging is observed. Finally, when entry costs are set too high, they are prohibitive and keep out even reliable members. We should not observe reneging, but we should also expect to see stagnant institutions with small membership.

Signaling hypothesis 1: When entry costs are low, members should renge on their commitments fairly often.

Signaling hypothesis 2: When entry costs are moderate, we should observe no reneging.

Signaling hypothesis 3: When entry costs are very high, institutional membership should remain small, but we should not observe reneging.

Of most interest to me, the model generates implications for states’ preferences over institutional design, in particular over ex ante entry costs. First, consider the preferences of potential new members. Those who derive low benefits from cooperation
but who could profit from suckering existing members into cooperation and then reneging will be opposed to any *ex ante* entry costs, as they can only lead to a decrease in utility.

*Potential entrant hypothesis 1*: Potential members who would derive low benefits from cooperation will oppose entry barriers.

In contrast, potential new members who would gain more from cooperation will have context-dependent preferences over entry costs. If the overall probability that a new entrant is reliable is low, those who would in fact be reliable will have a strong preference for the introduction of costly signaling. Such signaling costs will allow them to distinguish themselves from unreliable potential members. However, if the overall probability that a new member is reliable is relatively high, then reliable members have little to gain from the introduction of signaling costs, as they would have been able to gain the cooperation of existing members even in the absence of entry costs.

*Potential entrant hypothesis 2a*: When there are many potential new members in the population who would gain little from cooperation, those who anticipate large benefits from cooperation will favor the use of *ex ante* costs.

*Potential entrant hypothesis 2b*: When there are many potential new members in the population who would gain much from cooperation, they will be opposed to the use of *ex ante* costs.

Perhaps the most intriguing and immediately testable implications of the model involve the preferences of existing members of the institution over the introduction of signaling costs. Existing members have control over institutional design, so their preferences should be reflected in changes in the institution itself. In addition, we have
access to debates of existing members about institutional reform, and so may be able to observe the negotiating stances that states take to see if they line up with the model’s implications.

The model implies that it is of great importance that existing members of an institution get the level of signaling costs “just right.” If they set costs too low, they either leave existing members indifferent (if there are many reliable potential members in the population) or lead to only a small gain (if there are few reliable potential members in the population). If they set entry costs too high, existing members will either experience a loss (if there are many reliable potential members in the population) or be indifferent (few reliable potential members in the population). In contrast, moderate entry costs will serve as an effective screening device, separating reliable from unreliable potential members.

What does it mean to get signaling costs “just right?” One important implication of the model is that the appropriate level of signaling costs is determined by the costs and benefits of cooperation for reliable potential members. That is, the following factors are irrelevant to the determination of appropriate signaling costs: the benefits of cooperation for existing members; the costs to them if a new member reneges; and the proportion of reliable potential members in the population. Only the payoffs of potential members should enter into the optimal determination of entry costs. Existing members must aim to set signaling costs so that they are lower than the benefits of cooperation for reliable new members (otherwise they would deter entry); but higher than the benefits of reneging on deals (otherwise they would not screen out unreliable new members). These insights allow us to state the following hypotheses:
Existing member hypothesis 1: Existing members of an institution should express substantial concern that signaling costs be set at a level that is neither too high nor too low.

Existing member hypothesis 2: We should observe little conflict of interest among existing members about the appropriate level of entry costs.

Existing member hypothesis 3: As the benefits of cooperation for potential members rise, signaling costs should go up.

Existing member hypothesis 4: As the benefits of reneging for potential members rise, signaling costs should go up.

In the rest of this paper, I begin to explore empirical applications of the weak commitment and signaling model. I consider the institution of peacekeeping in civil conflicts and the expansion of the WTO. I also briefly consider extending the analysis to the expansion of the EU and of NATO.

Empirical Applications

Peacekeeping. Peacekeeping is the quintessential “weak commitment” institution. Peacekeepers enter a country when both sides in a civil war are willing to accept their intervention. The purpose of peacekeepers – as opposed to the occasional “peace enforcement” mission – is to monitor a ceasefire and other terms of a peace agreement. They typically carry only light arms, and are not present in sufficient numbers to impose large costs on either side if they decide to being fighting anew. Thus the question of why peacekeepers have any effect has been a puzzle.
Timothy Carter (2003) began developing an answer to this puzzle by arguing that
the agreement between sides in a civil war to allow peacekeepers to enter itself serves as
a signal to the UN about both sides’ intentions. Peace agreements are difficult to
negotiate in civil wars, and the UN has a preference to see a clear agreement before
sending in peacekeepers, so that the necessary conditions can be met for a separating
equilibrium to exist.

More recently, Page Fortna and I have pushed further on the idea that
peacekeeping may serve primarily as a signal rather than as a commitment device. In
Fortna and Martin (2009), we adapt the model presented above to the situation of fighting
between a government and a rebel group. In the model, we assume that the rebel group is
not sure whether the government is the type that will live up to the terms of a peace deal.
The government can choose to negotiate a peace agreement with the rebel group; to invite
peacekeepers in; or to continue fighting. The government pays a cost, in terms of
violation of sovereignty and intrusive foreign troops monitoring its actions, if it allows
peacekeepers to enter the country. Thus, peacekeeping has the potential to act as a costly
signal of the government’s intention to comply with the terms of a peace accord.

The model allows for three possible outcomes: continued fighting; a peace accord
(or truce) without the involvement of peacekeepers; or the intervention of peacekeepers.
It leads to a series of predictions about when we are most likely to see each of these three
outcomes. For example, the more a government values peace, the more likely it is to
allow peacekeepers to enter. Similarly, the greater the sovereignty costs associated with
peacekeeping, the less likely we are to observe that it occurs. Some results are more
counterintuitive. For example, consider the rebels’ prior beliefs that the government is a
reliable type (the analogue to p in the above model). As rebels become more certain that the government is reliable, the more likely we are to observe peacekeeping relative to continued fighting. However, an increased belief in reliability should also lead to a decreased chance of observing peacekeeping relative to peace agreements without peacekeeping, as rebels are more likely to be willing to accept agreements without peacekeepers. As in the model above, peacekeeping that involves a moderate ex ante cost will be the most effective at allowing rebels groups to distinguish between reliable and unreliable governments. A separating equilibrium will emerge when the upfront costs of peacekeeping are below the benefits of peace for a reliable government, but higher than the benefits of peace for an unreliable government.

Fortna and Martin subject the hypotheses derived from this model to a series of empirical tests, based on data from civil wars between 1989 and 1997 (64 cases). Because there are three possible outcomes, the appropriate method for analyzing these data is multinomial logit. We derive a series of proxies for the parameters of the model, for example arguing that the duration of war is positively correlated with the benefits of peace and that democracy is positively correlated with rebels’ prior beliefs about the reliability of the government.

We find substantial support for the predictions of the model. When considering the relative incidence of peacekeeping and continued fighting, the results are strong. The duration of war has a significantly positive coefficient, as predicted, while democracy also has a significantly positive coefficient. Other proxies also perform as expected and all but one meet standard tests of statistical significance. The multinomial logit specification also provides tests of the relative incidence of peacekeeping and peace
agreements without peacekeeping. The results on this dimension are not quite as strong, but still promising. For example, the duration of war has the predicted positive effect, but democracy no longer has a statistically significant effect. Overall, we conclude that the signaling model provides substantial insight into the demand for peacekeeping, and helps us to understand its dynamics in a way that thinking of it purely as a commitment device could not. In future work, we intend to expand the set of civil wars in the dataset and to extend the model to peacekeeping in interstate wars.

The WTO. The GATT/World Trade Organization provides an excellent setting for testing the central implications of the signaling model. Over time, the demands made of countries that wish to join the GATT/WTO have changed substantially. The GATT often had strikingly low barriers to entry. For example, post-colonial states were guaranteed, under Article XXVI:5(c), accession to the GATT with essentially no bargaining or other costs involved (Copelovitch and Ohls 2009). This situation changed in 1995 with the creation of the WTO, as no similar provision was made for former colonies or dependencies. Since 1995, accession negotiations have tended to be drawn-out and contentious, especially in prominent cases such as the accession of Russia (Dyker 2004) and China.

A major turning point in the process of accession came during the Uruguay Round of negotiations that led to establishment of the WTO. The Uruguay Round Agreements themselves do not provide any explicit rules for the process of accession, stating only that accession is open to any country as long as that country agrees on terms with WTO member states (Kavass 2007: 455). In practice, the WTO secretariat has
developed detailed accession procedures and has established an Accessions Division. The procedures laid out by the Secretariat are highly bureaucratic and “labyrinthine” (Kavass 2007: 456). All of these procedures have created substantial costs for potential entrants, in particular countries transitioning away from communism. In fact, the costs of accession are high enough to be a major concern to developing and transition countries, and have prompted organizations such as UNCTAD and the World Bank to publish extensive guides to the accession process in an effort to assist states to overcome these hurdles:

Countries seeking to become Members of the WTO must be prepared to perform a hefty volume of highly demanding work. Not only do they need to submit a voluminous amount of documents and attend meetings to answer questions; they also may need to make extensive and substantial changes to their tariffs and taxes, as well as revise many of their existing laws and regulations in order to bring them into conformity with the WTO norms and standards (Kavass 2007: 461).

Analysts express a consensus that the demands made of acceding countries have increased substantially over the GATT/WTO’s history (Ognivtsev, Jounela, and Tang 2001, 173). The process of escalating demands actually predates the creation of the WTO, going back to the 1980s. Mexico’s entry to the GATT provides a good example. It had reached an accession agreement during the Tokyo Round (1979), but decided in 1980 not to implement the agreement. In 1985 Mexico negotiated a new accession agreement that required a much higher entry cost than GATT members had imposed just six years earlier. The commitments to reduce trade barriers were far lengthier and more
precise, with fewer loopholes to protect Mexican industries. These enhanced demands came at the behest of the United States. VanGrasstek (2001, 127) finds the Mexican example typical: “Many of the countries that acceded to the GATT during the 1980s found the process to be more demanding, in large measure because of a change in policy on the part of the major trading countries.”

As the signaling model suggests, existing member states that worry about possible reneging of new members deliberately set barriers to entry high, even going beyond what the GATT/WTO agreements themselves specify (Butkeviciene et al. 2001, 230). The requirements that existing members impose on entrants, and the one-sided nature of the negotiating process, fit the assumptions of the model well. Analysts even argue that, beyond committing to comply with the WTO agreements, an acceding country must “pay a ‘membership fee’ in terms of specific concessions on tariff rates, commitments on agricultural subsidies and commitments on trade in services in return for its right to enjoy the benefits resulting from liberalization achieved in previous multilateral trade negotiations” (Ognivtsev, Jounela, and Tang 2001, 181). It is also worth noting that WTO procedures allow existing members a great deal of flexibility with respect to entry requirements. “Paradoxically for a rules-based organization, the WTO has no clear rules for the ‘price’ of membership” (Evenett and Primo Braga 2005, 2). The signaling model predicts that the entry barrier needs to be calibrated to each entrant’s costs and benefits, and so leads us to expect exactly this sort of flexibility in procedures.

The model predicts that we should see a higher entry cost imposed when a potential member could benefit substantially from joining and then reneging on its commitments, and we see widespread evidence of this dynamic in cases such as Mexico,
Russia, and China. The model also predicts that countries that would derive higher benefits from joining will be asked to pay a higher entry cost, and we also find widespread evidence supporting this implication in case studies. One analyst concludes that if existing members “know that their interlocutor is under strong political pressure back home to secure accession at any cost, the negotiators in Geneva will feel even more secure in setting a high price” (VanGrasstek 2001, 136).

This survey of evidence from case studies of GATT/WTO accession provides preliminary support for the signaling model. However, more systematic tests based on this experience should be possible. One research direction will involve looking more directly at negotiations over individual accessions, to determine whether the hypotheses summarized above about preferences over the terms of entry hold up. Another approach will include statistical analysis, attempting to explain variation in the entry fees demanded of different new members. The literature has identified a number of proxies for entry costs, ranging from the number of specific commitments made in particular sectors to the length of negotiations. While neither of these is a fully adequate indicator of \textit{ex ante} costs, looking for robust results across a number of indicators should allow for more precise testing of the signaling model.

\textit{Other applications: NATO and the EU.} Expansion of NATO and the European Union represent two other likely fruitful applications of the signaling model. In both cases, a substantial literature has developed regarding the process of expansion and accession of new members. In both cases, we observe variation in the stringency of entry requirements placed on new members, while having good access to internal discussion of
entry costs. In these cases, as in the WTO, we would expect to see existing members of
the institutions attempting to set entry barriers at an appropriate intermediate level, high
enough to screen out unreliable potential members but not so high that they deter reliable
members from joining. We would also expect to see that they could make greater
demands of potential members as the benefits of cooperation increase; and that when they
worry that a new member could gain substantially from reneging, they would set entry
barriers higher. NATO’s Partnership for Peace program, and comparison of the EU’s
successive waves of expansion, will allow for case-study and potentially statistical
analysis.

Conclusion

How do international institutions exert their effects on members? Typically,
scholars have looked at the ability of institutions to commit by imposing costs on
members if they renege on their commitments. However, institutions rarely if ever have
the capacity to fully commit all members to all of their commitments, all of the time.
Instead, they operate as weak commitment devices, leading to enhanced but inconsistent
cooperation. In this paper, I provide a model of institutions as weak commitment
devices, then add to the model the potential for institutions to also perform a signaling
function by requiring new members to pay an ex ante cost for joining the institution.

The model demonstrates that a weak commitment institution, on its own, has a
number of undesirable properties. In particular, it allows unreliable new members under
some conditions to join the institution and fails to ensure that cooperation emerges among
reliable states. In contrast, when members must send costly signals to join the institution, and these signals are appropriately calibrated, reliable members can consistently distinguish themselves from unreliable. In this case, the fullest extent of mutually-beneficial cooperation emerges, and unreliable states are prevented from being able to sucker existing members of the institution.

This model, while simple, gives rise to a rich set of potentially testable empirical implications. If we consider international peacekeeping in civil wars, conceiving of peacekeeping as a weak commitment and signaling institution fits the pattern of peace and conflict well. More preliminary work on the WTO suggests that the model’s predictions about members’ preferences for the costliness of signals hold up. Future empirical work will focus on more systematically studying the WTO, and extending the model to expansion of NATO and the EU. While the model here treats states’ entry into an institution as a costly signal, extensions and alternative applications other institutional dynamics, such as approval of commitments, are also promising.
References

Baglioni, Angelo. 2008. “Corporate Governance Institutions as Signalling and Commitment Devices.” Ms, Università Cattolica, Milano, Italy


Figure 1

Sequence of Moves

Nature chooses whether A is reliable → A chooses whether to join the institution → B chooses whether to cooperate with A

A chooses to cooperate or renege → Payoffs realized
### Table 1
Equilibria of Signaling and Commitment Game

<table>
<thead>
<tr>
<th></th>
<th>Low reliability</th>
<th>High reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(p &lt; d/(b_r + d))</td>
<td>(p &gt; d/(b_r + d))</td>
</tr>
<tr>
<td><strong>Low signaling cost</strong></td>
<td><strong>Semi-separating</strong>: Reliable A chooses institution. Unreliable A chooses institution with probability pb/(d(1-p)). When B sees institution, cooperate with probability z/(a-c).**</td>
<td><strong>Pooling</strong>: All A’s choose institution. B cooperates.</td>
</tr>
<tr>
<td>(b_r &gt; a - c &gt; z)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Moderate signaling cost</strong></td>
<td><strong>Separating</strong>: Reliable A chooses institution. Unreliable A does not choose institution. B cooperates if it observes institution. **</td>
<td><strong>Separating</strong>: Reliable A chooses institution. Unreliable A does not choose institution. B cooperates if it observes institution.</td>
</tr>
<tr>
<td>(b &gt; z &gt; a - c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High signaling cost</strong></td>
<td><strong>Pooling</strong>: No A’s choose institution. B does not cooperate.</td>
<td><strong>Pooling</strong>: No A’s choose institution. B does not cooperate.</td>
</tr>
<tr>
<td>(z &gt; b_r &gt; a - c)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2
Unreliable A Payoffs, Comparing Signaling to Pure Weak Commitment

<table>
<thead>
<tr>
<th>Low reliability</th>
<th>High reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low cost to join</td>
<td>Indifferent: B cooperates with probability that leaves unreliable A indifferent; bears risk</td>
</tr>
<tr>
<td></td>
<td>Small loss: bears signaling cost z, still gets high reneging payoff</td>
</tr>
<tr>
<td>Moderate cost to join</td>
<td>Indifferent: no cooperation</td>
</tr>
<tr>
<td></td>
<td>High loss: does not induce B to cooperate</td>
</tr>
<tr>
<td>High cost to join</td>
<td>Indifferent: no cooperation</td>
</tr>
<tr>
<td></td>
<td>High loss: does not induce B to cooperate</td>
</tr>
</tbody>
</table>

Table 3
Reliable A Payoffs, Comparing Signaling to Pure Weak Commitment

<table>
<thead>
<tr>
<th>Low reliability</th>
<th>High reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low cost to join</td>
<td>Gain: bears small cost of joining institution, but gains probabilistic cooperation</td>
</tr>
<tr>
<td></td>
<td>Small loss: gets cooperation, bears small cost z</td>
</tr>
<tr>
<td>Moderate cost to join</td>
<td>Gain: gets cooperation, bears moderate cost z</td>
</tr>
<tr>
<td></td>
<td>Moderate loss: gets cooperation, bears moderate cost z</td>
</tr>
<tr>
<td>High cost to join</td>
<td>Indifferent: no cooperation</td>
</tr>
<tr>
<td></td>
<td>Large loss: no cooperation</td>
</tr>
</tbody>
</table>

Table 4
B Payoffs, Comparing Signaling to Pure Weak Commitment

<table>
<thead>
<tr>
<th>Low reliability</th>
<th>High reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low cost to A for institution</td>
<td>Small gain: gets some cooperation, but also bears chance of reneging</td>
</tr>
<tr>
<td></td>
<td>Indifferent (in expectation): cooperates, some reneging; bears risk</td>
</tr>
<tr>
<td>Moderate cost to A for institution</td>
<td>Large gain: gets cooperation with reliable A</td>
</tr>
<tr>
<td></td>
<td>Large gain: retains cooperation with reliable A, no risk of unreliable A reneging</td>
</tr>
<tr>
<td>High cost to A for institution</td>
<td>Indifferent: no cooperation</td>
</tr>
<tr>
<td></td>
<td>Loss: no cooperation</td>
</tr>
</tbody>
</table>
Appendix

To come.