Trust Building, Trust Breaking: The Dilemma of NATO Enlargement

Andrew Kydd

What determines the price of membership in an international institution? Barbara Koremenos, Charles Lipson, and Duncan Snidal hypothesize that uncertainty about the preferences of other states will increase that price, as stated in Rational Design conjecture M2, restrictive MEMBERSHIP will increase with UNCERTAINTY ABOUT PREFERENCES. When states are uncertain about the motivations of other states, they will demand costly signals of reassurance before being willing to cooperate fully.[1] In a multilateral context, this may take the form of an institution with a significant barrier to entry, a price of admission. The price of admission serves to separate states who are seriously interested in cooperation from those who have more exploitative motivations. More cooperative states will be willing to pay the price, and this will reveal their cooperative nature to others, facilitating cooperation. Less cooperative states will not be willing to pay the price, and this too will reveal their type, leading others to cooperate less with them.

The case of NATO enlargement is a perfect example of this logic at work. In the recent enlargement round, NATO established an extensive set of criteria to determine who would be admitted and who would not. The criteria included democratization, civilian control over the military, and the resolution of all border disputes and frictions with neighbors over ethnic minority issues. These hurdles served to separate the more cooperative states from the rest, enabling NATO to admit and cooperate more intensively with those states with proven cooperative credentials. At the same time, proponents of NATO enlargement argue that membership encouraged cooperation between the Eastern European states in spite of lingering mistrust.

I thank the participants in the First Annual Conference on EU–U.S. Relations, European Union Center, Georgia Tech, March 1999, where I presented an earlier version of this article. I also thank the participants in the Rational Design project, the editors of *IO,* and two anonymous reviewers for their feedback. I especially thank Stephan De Spiegeleire, Frank Schimmelfennig, Charles Glaser, Barbara Koremenos, Dan Lindley, Charles Lipson, James Morrow, Duncan Snidal, Robert Pahre, David Pervin, and Peter Rosendorff for their comments and suggestions.

1. Kydd 2000a,b.

International Organization 55, 4, Autumn 2001, pp. 801–828

Secretary of State Madeleine Albright, for instance, argued that enlargement would expand "the area of Europe where wars do not happen" thereby preventing conflicts that could draw Russia back into the region or necessitate NATO intervention. Proponents of enlargement see NATO as a benign institution representing the Western "security community" that serves to promote trust and foster cooperation among its members.[2]

Critics of NATO enlargement, including many academics, are more dubious of the merits of NATO expansion. John Lewis Gaddis found a near consensus among historians that NATO expansion was "ill-conceived, ill-timed, and above all ill-suited to the realities of the post–Cold War world."[3] They argued that NATO expansion would antagonize Russia, exacerbating its lingering distrust of the West and strengthening anti-Western elements in the Russian political system. This would in turn lead to lower levels of cooperation between Russia and the West.

Thus NATO enlargement poses an acute policy dilemma. NATO can be a benign security community that identifies more cooperative states and promotes cooperation among them and yet be perceived as an expanding alliance that Russia finds threatening. Although expanding the security community enlarges the zone of peace and mutual trust, it may generate fear among those still on the outside. This dilemma presents policymakers with a difficult choice. They can choose to expand the community and secure the benefits associated with greater cooperation among the members, paying the costs of a lower level of cooperation with the outside power. Or they can choose to forgo expansion in an effort to reassure the outside power, and suffer the consequences of greater instability among the excluded potential members.

I present a game theoretic analysis of the conditions that give rise to this dilemma and show how actors will choose to resolve it. To do so it is necessary to go beyond conventional models of alliances, which focus on public goods provision and deterrence. The most common models of alliances are the public goods provision games that have often been applied to the issue of NATO burden sharing.[4] Another important type of alliance model focuses on signaling and deterrence. Typically a defending power is interested in signaling its resolve to defend an alliance partner against a third party, in order to deter an attack.[5] Neither style of model adequately captures what went on during NATO enlargement, because they do not focus on trust. Trust and mistrust are at the core of the NATO enlargement dilemma—the goal of enlargement is to foster trust among the new allies, and the unwanted side effect is to lessen trust with Russia. Thus the model presented here focuses on trust, how it is built and how it is weakened.

2. For the origin of the security community concept, see Deutsch et al. 1957.
3. Gaddis 1998.
4. For the origin of this literature, see Olson and Zeckhauser 1966; and for a survey, see Sandler 1993.
5. See Morrow 1994a; and Smith 1995.

The model shows that enlargement poses a dilemma when the levels of trust are middling, and hence the level of uncertainty about preferences is maximized,[6] both between the potential new members and between the community and the outside power. If states are relatively certain about one another's preferences, there will be little reason to have a high entry price for an institution, because state motivations, benign or malign, will already be known. Likewise, if NATO and Russia are relatively certain about each other's preferences, NATO expansion will have no effect on NATO–Russian relations. It is where uncertainty over preferences is maximized that expansion with a high price of admission is valuable in sorting out the cooperative from the noncooperative states, and yet potentially damaging to NATO–Russian relations. Thus Rational Design conjecture M2, restrictive MEMBERSHIP increases with UNCERTAINTY ABOUT PREFERENCES, is supported by the model.

A further and possibly counterintuitive result of the model is that under certain conditions expansion will actually be *reassuring* to the outside power, not provocative as most analysts assume. This is also a function of the criteria by which allies are selected. If NATO were to expand unconditionally, admitting anyone who applied, it would be difficult to portray this to the Russians as an effort by a benign security community to foster cooperation, because membership would not be conditional on cooperation. Instead, it would look like an expansionist West attempting to encircle Russia. The more restrictive and demanding the membership criteria are, however, the more support the benign explanation of NATO behavior has, and the less convincing is the alternative explanation that NATO is out to get Russia and is assembling a large anti-Russian coalition. If the criteria are restrictive enough, conditional expansion may actually be reassuring, because it tells the Russians that NATO is not interested in unlimited expansion and that the stated explanation for expansion is probably correct. Thus with adequately restrictive membership criteria, NATO enlargement can be both beneficial in fostering cooperation among the allies and not too harmful or possibly even beneficial for NATO–Russian relations as well, eliminating the dilemma.

In what follows I will first discuss existing explanations of NATO enlargement. Conventional rationalist approaches have proven largely unsatisfactory; consequently, some analysts have turned to a constructivist alternative. I formulate an alternative rationalist approach to the problem, focusing on trust, reassurance, and the enlargement dilemma identified earlier. In the final section I present a game-theoretic model of NATO enlargement and examine equilibria in the model.

The Puzzle of NATO Enlargement

The enlargement of NATO is one of the most important developments in international affairs after the Cold War; it is also one of the most puzzling. Many factors

6. See below for the relationship between trust and uncertainty about preferences.

were at work in producing NATO enlargement, from domestic political issues, such as the existence of electorally significant East European émigré communities in the United States, to the personal rapport between U.S. president Bill Clinton and Czech president Vaclav Havel.[7] Yet certain aspects of the enlargement process seem difficult to explain with conventional theories of alliance formation.

The least puzzling part of NATO enlargement is the desire of the East European states to join the alliance. Most analysts interpret this simply as a desire for protection against Russia, which East Europeans still regard as a potential threat to their independence and autonomy. Given such fears, their desire to join NATO is perfectly understandable; NATO's obvious military superiority to Russia and its successful history of resistance to Russian expansion in the Cold War make it an appealing alliance partner. This desire of East Europeans to align with the stronger side, the West, is clearly at odds with Kenneth Waltz's balance-of-power theory, which predicts that states will join the weaker side.[8] However, it is consistent with Stephen Walt's balance-of-threat theory, which argues that states prefer to join the less threatening side, where perceived aggressive intentions is one component of threat.[9] Eastern European states, still feeling a potential threat from the East, turn to a less threatening alliance for shelter.

While the motivations of the new members seem readily comprehensible, the behavior of the existing NATO members seems less so. Why should current NATO members want new alliance partners? The central purpose of alliances is usually taken to be to increase the security of the members by deterring some external power or better preparing them to fight if deterrence fails. Yet three facts about the recent round of enlargement seem problematic in this light. First, the Russian threat is as low as it has been since the 1920s, and it does not seem to be increasing markedly. This diminished threat from the East leads some realists to predict that NATO will eventually cease to exist, at least as a genuine alliance.[10] Second, NATO enlargement will cost current members both in terms of money and in terms of potential involvement in defending the new Eastern European members.[11] Third, it is not clear what the new allies will contribute toward the common defense and deterrence. At a military level, their forces are far below NATO standards; indeed, bringing them up to Western levels is the primary expense involved in enlargement. One could argue that they bring additional strategic depth, yet NATO was able to hold the much more powerful Soviet Union at bay on the old inter-German border. Why spend money to acquire strategic depth that was not necessary when the threat was far greater than it could ever be again? The new allies might contribute to NATO's

7. For detailed accounts of the process leading up to enlargement, see Eyal 1997; and Goldgeier 1998.

8. Waltz 1979, 127.

9. Walt 1987, 25.

10. See Mearsheimer 1990, 5; and Walt 1997, 171.

11. On the issue of monetary costs, for the optimistic side, see Asmus, Kugler, and Larrabee 1996; and for the pessimists, see Perlmutter and Carpenter 1998; and Rubinstein 1998.

new mission of out-of-area peace enforcement,[12] but in the most recent case, Kosovo, the chief burden has been borne by the great power members, especially the United States, Britain, and France.

Considerations such as these have led some analysts to despair of explaining NATO enlargement in rationalist terms. In an insightful essay, Frank Schimmelfennig highlights these difficulties for rationalist approaches and then argues for an alternative, constructivist, explanation.[13] According to Schimmelfennig, "In the constructivist perspective, the enlargement of an international organization is primarily conceived of as a process of international socialization."[14] International organizations engage in socialization when they "teach" their set of constitutive norms and values to aspiring new members of the community. New members are graded on how well they have internalized the norms and values and are admitted when they have proven that they have sincerely adopted the new identity. NATO is "best understood" as an "organization of an international community of values and norms;" primarily democracy, liberty, and the rule of law.[15]

Schimmelfennig goes on to show how the process of NATO enlargement seems to conform to this logic. NATO's "Study on NATO Enlargement" outlines the goals that enlargement was to achieve and criteria for entry for potential new members.[16] The goals include not only the traditional aim of "collective defense" but also such things as spreading democracy and civilian control over the military, fostering cooperation, consultation, and consensus building, and increasing transparency in defense planning and military budgets. Membership criteria for potential members are also revealing. Heavily stressed are such attributes as democracy, civilian control over the military, and the resolution of all border disputes and ethnic conflicts.

These criteria might be dismissed as pleasant-sounding verbiage if not for the fact that the countries invited for membership in the first round—Poland, the Czech Republic, and Hungary—met them and the ones put on the slow track did not. Leaving aside the former Soviet Republics, Romania, Bulgaria, and Slovenia were at one time or another mentioned as possible members during the first wave and were all rejected in part because the political goals were not achieved. Hungary's inclusion is striking in this context. Hungary has indeed made much progress on democratization, building a liberal economy and, crucially, peacefully resolving post–Cold War frictions with Romania concerning the Hungarian minority in Transylvania. So Hungary scores well on the political variables of interest yet would seem to be a burden strategically. Landlocked and noncontiguous with any NATO country, Hungary would be difficult to defend without violating the territory of other states, notably Austria and Slovakia. South of the Carpathian Mountains, it is not on

12. Lepgold 1998.
13. Schimmelfennig 1998/99. See also his analysis of EU expansion in Schimmelfennig 2001.
14. Schimmelfennig 1998/99, 211.
15. Ibid., 213–14.
16. NATO 1995.

a central axis of advance to or from Western Europe, unlike Poland. Furthermore, it is contiguous with Yugoslavia and hence in a historically unstable neighborhood. Yet because Hungary meets the political criteria, it was admitted in the first round.[17]

The boon of NATO membership, then, seems to have been used to reward those East European states that took certain political steps, such as entrenching democracy and civilian control over the military and resolving ethnic and border disputes with each other, rather than in pursuance of any strategic logic related to defense or deterrence. This seems to accord with Schimmelfennig's constructivist account of NATO enlargement rather than with any received rationalist account. NATO is attempting to foster democracy because it is composed of democratic states, and such states simply have a preference that other states be democratic too. NATO expands to include states that are "like us" because we want other states to be "like us." The community of norms is extended through socialization.

Note that this explanation of the membership criteria differs sharply from Rational Design conjecture M2, restrictive MEMBERSHIP increases with UNCERTAINTY OVER PREFERENCES. In the constructivist account, the restrictive membership criteria are a product of the desire to have others be like us. They are a test of how socialized the new potential members are. The stringency and nature of the admission criteria are therefore determined not by uncertainty or instrumental calculations about who is likely to cooperate, but by how the identity of the institution gets defined. Who we are determines who we admit.

Trust, Mistrust, and NATO Enlargement

There are alternative rationalist accounts that can explain the same pattern of behavior, however, as Schimmelfennig himself acknowledges, and there are many questions about NATO enlargement that the constructivist account leaves unanswered. Most importantly, the constructivist account seems to lack a compelling explanation of why NATO enlargement was controversial. If this were a simple case of an international institution extending its norms by socializing new members, why did large sections of the Western policy community, individuals who presumably subscribe to those norms, object so vociferously to it? The debate over NATO enlargement was a battleground of competing arguments, to be sure, but it is difficult to interpret it as a battleground of competing norms. No one was arguing for a different set of norms, or that socialization of new members is bad. In fact, the primary arguments against NATO enlargement, and many of the ones in favor, were of a strategic nature. Costs and benefits were weighed, and the impact of actions on beliefs, and beliefs on actions, were central. Indeed, I argue that the essence of the NATO enlargement debate was an argument about benefits and costs having to do

17. For the beneficial effects of NATO enlargement on Hungarian democracy and Hungarian–Romanian relations, see Kramer 1999, 429–30.

with trust, mistrust, and cooperation, and that these issues are eminently suitable to strategic analysis. NATO enlargement, in this view, is primarily designed to foster trust and cooperation amongst the East European states, and its primary drawback is the increased distrust and potential noncooperation it might foster between NATO and Russia.

If we are to take the rhetoric surrounding enlargement seriously, the most important goal for the existing NATO members is to "enhance stability," that is, to foster cooperation and prevent conflict *between the East European states themselves.* This explains the strong emphasis on resolving territorial disputes and ethnic frictions that might lead to war. The role of democracy is also instrumental in this context. It is clear that members of the Clinton administration, particularly Anthony Lake, were influenced by the democratic peace literature and explicitly adopted the goal of fostering democracy. The main assertion of this literature is that democracies do not fight each other or are much less likely to do so than other regime types.[18] A democracy, then, is unlikely to have conflicts with other democracies and will be able to resolve those that it does have peacefully. To foster democracy, therefore, is to foster peace. Thus an alternative explanation of NATO's insistence on democracy and the resolution of disputes as criteria for membership is a desire to reduce the likelihood of conflict in Eastern Europe.

Furthermore, this need not be a purely altruistic preference on NATO's part. Conflict in Eastern Europe is bad for NATO even in the absence of any other-regarding desire to increase the welfare of East Europeans. Conflict in the region could generate refugee flows into the West, trigger increased criminal activity and smuggling, and reduce the gains from trade and economic integration with the region, as well as generate opportunities for Russia to reassert its influence in the area, possibly generating pressures for a NATO response. Secretary of State Madeleine Albright argued that the Eastern European states, to demonstrate their worthiness of admission, had "strengthened their democratic institutions, improved respect for minority rights, made sure soldiers take orders from civilians, and resolved virtually every old border and ethnic dispute in the region. This is the kind of progress that can ensure that outside powers are never again dragged into conflict in this region."[19] Thus, acting purely on the basis of self interest, NATO could reasonably insist on democratization and confidence building in Eastern Europe as a criteria for NATO membership.[20] NATO's goal in expansion, then, as many NATO officials have publicly stated, is to prevent conflict in the East by fostering mutual trust and cooperation.[21]

18. See Brown 1996; and Chan 1997.

19. Albright 1998.

20. Schimmelfennig acknowledges this point. Schimmelfennig 1998/99, 230.

21. For an interesting argument that NATO enlargement has not actually accomplished these goals, in particular, has not fostered democracy, see Reiter 2001. Reiter argues that the countries admitted were solid democracies with civilian control of the military before NATO enlargement became a possibility, and hence that NATO enlargement was irrelevant in promoting cooperation in Eastern Europe. Even if one agrees with this point, which I do not fully, my analysis still can explain both the enlargement criteria

Building trust between alliance members is not typically thought to be a central task of alliances, though it has often been mentioned as a salutary side effect. The reintegration of Germany into Western Europe was facilitated by its participation in NATO, as well as in the European Community. Trust building is more often thought of in the context of "security communities." As Karl Deutsch and his colleagues put it over forty years ago, "A security-community . . . is one in which there is real assurance that members of the community will not fight each other physically, but will settle their disputes in some other way."[22] For war to be unthinkable, it must be that members of these communities have reassured each other of their intentions to the extent that they no longer fear that other members might want to attack them. Secretary Albright's statement that enlarging the alliance would expand "the area of Europe where wars do not happen" reflects this security community logic.

If the chief benefit of NATO enlargement can be seen as building trust and fostering cooperation amongst the East European states, the chief cost of NATO enlargement is surely the lessening of trust and decline in cooperation between NATO and Russia. Many prominent opponents of enlargement have focused on this issue. In a remarkable open letter dated 27 June 1997 from a group of foreign policy experts to President Clinton opposing NATO enlargement, the Russian reaction was the first issue of concern. Signed by a broad spectrum of opinion leaders from Richard Pipes and Paul Nitze to Senator Bill Bradley and Arms Control Association president Spurgeon Keeny, the letter warned that, "In Russia, NATO expansion, . . . will strengthen the non-democratic opposition, undercut those who favor reform and cooperation with the West, bring the Russians to question the entire post–Cold War settlement, and galvanize resistance in the Duma to the Start II and III treaties."[23] Other opponents echoed this warning. Raymond Garthoff argued, "To have driven Russia from support of Desert Storm to support for the Saddam Husseins of the future by denying it a responsible role in the security architecture of the new world order would be a heavy burden to assume for expanding NATO."[24] John Lewis Gaddis lamented the fact that the Clinton administration appeared to be following the example of the harsh Versailles settlement after World War I, rather than that of the Vienna settlement after the Napoleonic wars or the post–World War II settlement, and thereby was violating a key principle of grand strategy: be magnanimous to defeated adversaries.[25] Other analysts of enlargement have also focused on this theme.[26]

---

and the enlargement dilemma, which are a function of policymakers' *perception* that NATO enlargement would promote democratization and trust building while harming NATO–Russian relations.

22. Deutsch et al. 1957, 5. For a constructivist take on security communities, see Adler and Barnett 1998.

23. Available on the Web at ⟨http://www.cpss.org/nato/oplet.htmd.⟩

24. Garthoff 1997, 10.

25. Gaddis 1998, 145.

26. See Pierre and Trenin 1997; Asmus and Larrabee 1996; Brown 1995; and Mandelbaum 1995.

The idea that alliance formation can be provocative, or produce fear on the part of an excluded country, has also not been central to the alliance literature, but it has been explored. Glenn Snyder presents the most thorough analysis of the "security dilemma" aspects of alliance formation.[27] He argues that states that are basically security seekers will nonetheless feel a need to form alliances because they are not sure of the intentions of other states. Since possible adversaries may have aggressive intentions, it is necessary to build up one's own power against them, and acquiring allies is one way to do so. Forming alliances, however, and increasing one's level of commitment to one's allies, will be provocative, and increase the adversary's level of fear, causing the adversary to seek to strengthen its alliances in turn.[28] This generates a familiar "spiral" of increased fear and conflict, even though both sides have fundamentally defensive motivations.[29] It is this phenomenon that opponents of NATO enlargement see as the primary strategic cost to be paid as a result of admitting new allies.

The relationships among the criteria for membership, the rules governing expansion, and how provocative expansion is to the Russians are rarely discussed by proponents or opponents of enlargement. It is widely recognized that admitting certain specific countries will be especially provocative, particularly states that were once part of the Soviet Union, but the relationship between the general membership criteria and the beliefs of the outside power is underanalyzed. It would seem that unconditional expansion would be highly provocative, because an aggressive alliance might want to maximize the number of adherents, and hence encircle its potential victim.[30] Intuition also suggests that the more restrictive the criteria for membership, the less provocative the alliance would be to outsiders. The model I present later takes this logic further, however, and demonstrates that if the criteria are restrictive enough, conditional expansion may actually be reassuring, thus eliminating the dilemma of expansion altogether. Conditional expansion can reassure by demonstrating that the alliance is not interested in *unconditional* expansion. By not expanding to include any country that asks to join, NATO demonstrates that it is not attempting to encircle Russia with a ring of hostile allies. Not expanding at all would be even more reassuring, of course, but conditional expansion can still be at least somewhat reassuring.

This, then, is the dilemma of NATO enlargement. The chief benefit of enlargement is to extend the security community to new members, building trust and fostering cooperation. The chief downside is the increased distrust and weakening of cooperation between NATO and Russia. In this sense, NATO enlargement is all about trust. Trust, in turn, is a suitable subject for rationalist, strategic analysis. Beginning with the work of James Coleman, a rational choice literature on trust and

27. Snyder 1984.
28. Ibid., 477.
29. See Jervis 1976, 62; and Kydd 1997.
30. For a contrary argument that offensive alliances are smaller than defensive ones, see Schweller 1998, 61.

reassurance has taken root in sociology, as well as in economics, political science, and international relations.[31] What moves build trust, and what moves decrease it? How are the costs of trust breaking to be weighed against the benefits of trust building? These are questions that a strategic analysis can answer for us. Before getting to a specific model of NATO enlargement, however, I discuss the relationship between trust and uncertainty about preferences.

Trust and Uncertainty About Preferences

The idea that mistrust causes conflict is a basic element of the security dilemma, the spiral model, and in a sense, structural realism as a whole.[32] How should we understand trust and mistrust? I argue that trust is related to uncertainty about the underlying motivations or preferences of the other side, one of Koremenos, Lipson, and Snidal's key variables. Classical international relations theory makes a useful distinction between status quo and revisionist states.[33] Status quo states, or security seekers, are basically satisfied with the way things are and want to preserve the status quo. Revisionist states are not content with the status quo and want to modify it in some way or perhaps even overthrow it entirely. They would be interested in expansion even if all other states were too weak to threaten them and their security was assured. Status quo states might be interested in expansion as well, but only tactically, to fend off a perceived threat to their security. In game-theoretic terms, status quo states would cooperate if they thought the other side would too, whereas revisionist states would defect even if they thought the other would cooperate.

A central problem that states face is that others' motivations are not always apparent; there is uncertainty about preferences. Countries may claim to be status quo and yet harbor revisionist desires. Given uncertainty about the preferences of other states, even status quo states may feel the need to compete for power in the international arena by engaging in arms races, building spheres of influence, or even launching wars if preventive or preemptive windows of opportunity arise. If other states cannot be trusted, it may make sense to take advantage of temporary or wasting assets and subdue them while it is still possible, rather than waiting until the potential threat can no longer be defeated.[34]

Trust is therefore related to uncertainty about preferences. I define a state's level of trust for another as its estimate of how likely it is that the other is status quo oriented, rather than revisionist. To trust another state is to think it relatively likely that the state is status quo oriented, so that if it acts aggressively it is because the state fears some other state, not because it is intrinsically expansionist. To mistrust a state is to believe it relatively likely that the state is really revisionist, and that it

31. See Coleman 1990, 91; Güth and Kliemt 1994; Watson 1999; and Kydd 2000a.
32. See Jervis 1976 and 1978; and Glaser 1994/95 and 1997.
33. Schweller 1998, 15–38.
34. On preventive war, see Copeland 2000, 11–34; on preemptive war, see Van Evera 1999, 35–72.

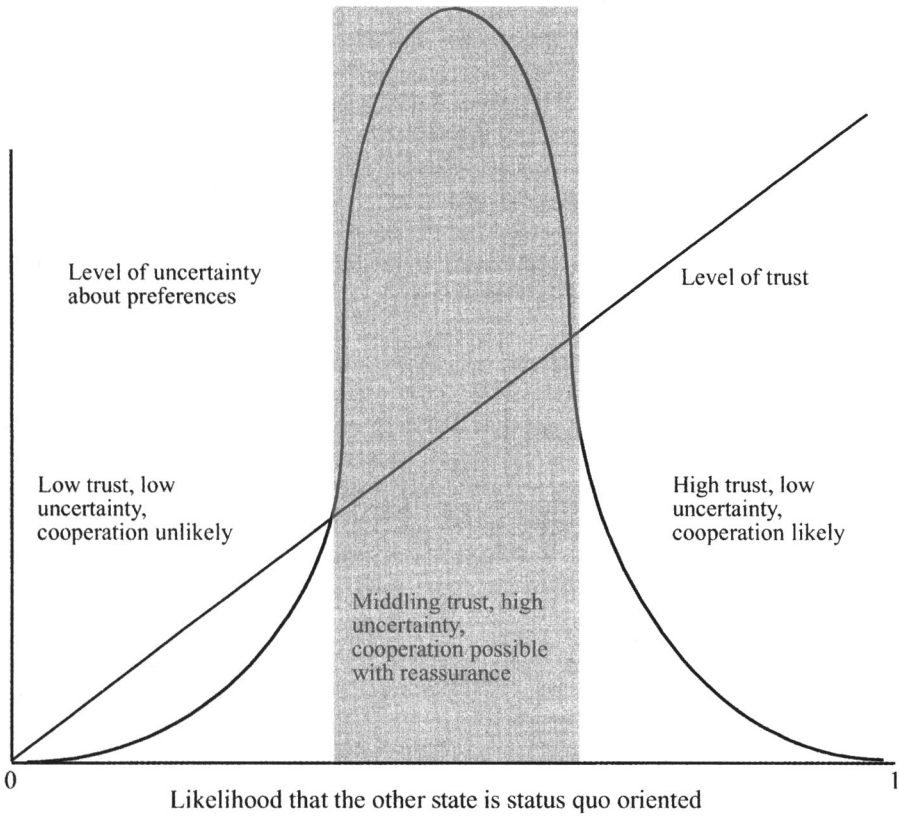

FIGURE 1. *Trust and uncertainty about preferences*

would be expansionist even if it thought its neighbors were status quo. With a state that one trusts, therefore, conflicts can be overcome through reassurance. If you can persuade them to trust you, to believe that you too are status quo oriented, then cooperation should be feasible since both sides would prefer to reciprocate cooperation. With a state that one mistrusts, however, no amount of reassurance can eliminate conflict, which is driven by the revisionist goals of the other state.

The relationship between trust and uncertainty about preferences is shown in Figure 1. The underlying variable is the likelihood that the other state is status quo oriented, or trustworthy. This probability ranges from zero to 1. Near zero, the state is relatively certain that the other is revisionist, not status quo oriented. Uncertainty and trust are both low, and cooperation is unlikely given the state's pessimistic beliefs. In the middle, the state is relatively unsure whether the other side is status quo oriented. Here uncertainty is maximized, and trust is at a middling level. In this zone of great uncertainty, whether cooperation takes place may hinge on costly

signals of reassurance or other incentives. Reassuring gestures, such as those posed by stringent admissions criteria to an international institution, can push the level of trust over the critical threshold and make cooperation possible. At the right in the figure, the state is relatively convinced that the other state is a trustworthy security seeker. Trust is maximized here, and uncertainty about preferences has declined to minimal levels again. Here, cooperation is quite likely because trust is high and costly signals of reassurance are less important.

In this context NATO was asking the Central European states to do two things, to cooperate with each other in the present in the face of mistrust in order to reveal their status quo nature, and to lock in domestic institutional structures that would provide assurance that they would cooperate in the future, that is, remain status quo states. NATO asked the Central Europeans to resolve outstanding territorial and ethnic disputes, and Hungary's eagerness to do so was both directly cooperative and a reassuring signal about its present underlying motivations. Furthermore, NATO asked them to lock in domestic structures, in particular, democracy and civilian control over the military, which are associated with status quo states, and are thus reassuring for the future. While the model I present is a single-shot game and thus focuses on present cooperation, NATO's insistence on democracy and civilian control over the military as institutional constraints on future behavior is in much the same spirit.

Modeling NATO Enlargement

The model of NATO enlargement I offer here is based on previous game-theoretic work on trust and cooperation but is closely tailored to the NATO enlargement question.[35] Consider a game involving $n + 2$ players, *west (W)*, *east (E)*, and a set of n potential allies, numbered 1, 2, 3, . . . n. As shown in Figure 2, *west* and *east* have a bilateral relationship, as well as relationships with each of the potential allies. The allies also have relationships with each other.

The game is divided into three stages. In the first stage *west* decides whether to offer a security guarantee to the potential allies. In the second stage the potential allies play a multilateral "trust game" with each other in which they may cooperate or defect. In the third stage *east* and *west* play a bilateral trust game. I first describe the structure of the game and then turn to the equilibria.

In the opening move of the game *west* can offer a security guarantee to the potential allies. I model this as a choice among three options. First, *west* could offer no security guarantees at all. The allies would then be left to fend for themselves, cooperating or not as they see fit. Second, *west* could offer conditional guarantees; that is, to encourage the allies to cooperate among themselves in the next stage, *west* can make membership in the "western bloc" contingent on cooperating with one's

35. Kydd 2000a,b.

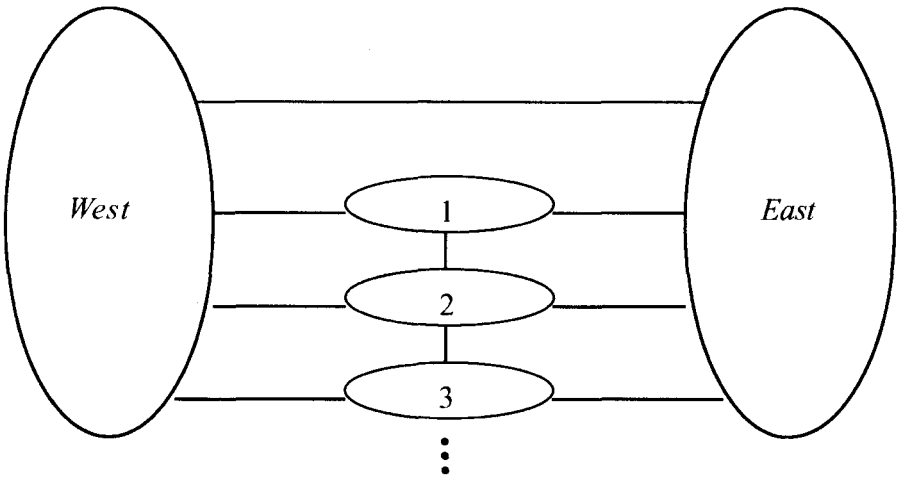

FIGURE 2. *The players and relationships in the NATO enlargement game*

neighbors in the subsequent multilateral trust game.[36] Third, *west* could offer unconditional guarantees to all the potential allies. In this case *west* offers a security guarantee regardless of the behavior of the potential allies. Receiving the security guarantee is worth g_i to country i.[37]

The second stage is a multilateral trust game between the potential allies. Each of the n players can be one of two types, "nice" or "mean." Nice types have Stag Hunt preferences, so they prefer to cooperate if they think the other players will. Nice types correspond to the concept of the status quo state. Using the traditional payoff notation in the analysis of the Prisoners' Dilemma, where T stands for temptation to defect while the other cooperates, R stands for reward for mutual cooperation, P stands for the punishment of mutual defection, and S for the sucker's payoff of unilateral cooperation, the payoff ordering for the nice type (player i) is $R_{iN} > T_{iN} > P_{iN} > S_{iN}$. Mean types have Prisoners' Dilemma preferences and thus prefer unilateral defection to mutual cooperation, corresponding to the revisionist state. Their payoff ordering is $T_{iM} > R_{iM} > P_{iM} > S_{iM}$.[38]

36. This raises a commitment problem. Given that NATO pays a cost (discussed later) to extend a security guarantee, it might be best for them to promise a security guarantee, and then renege on the promise after the allies have moved. I will assume that NATO faces reputational costs sufficient to render such a deceitful strategy unappealing.

37. In reality, of course, there is a much larger set of possible offers. Some states could be given guarantees even if they do not cooperate; others could be denied a guarantee even if they do. The three-part choice is the simplest framework in which we can examine how expansion could be threatening or reassuring, depending on whether it is conditional or unconditional.

38. For quasi-game-theoretic analyses of trust along these lines, see Bennet and Dando 1982; and Plous 1988. Glaser also suggests this strategy for modeling the security dilemma. Glaser 1997.

Nature chooses whether each player is nice or mean. The likelihood that player i is nice is denoted p_i. These probabilities are the game-theoretical representations of trust. A higher p_i corresponds to a higher level of trust or a lower level of fear. These exogenous levels of trust can come from past experience with another state, general experience with many states, or theoretical ideas about how international politics works. For instance, France was mistrustful of Germany in the aftermath of World War II because of the experience of invasion, whereas the United States may be more trusting of democracies out of a general experience that democracies keep their commitments more often than nondemocracies and because policymakers buy into the democratic peace theory. Each player knows its own type but not the type of the other players.

The players must choose to cooperate or defect in ignorance of what the other players have chosen, just as in a simple normal form game. When a player cooperates or defects, this act affects all other players. To allow for the fact that some countries are more important to a given country than others, however, I allow each country to weight the other countries individually. That is, country 1 can care very much if country 2 cooperates, but not so much if country 3 does. These weights are denoted w_{ij}, which represents how much country i cares about country j. For instance, countries would tend to weight countries close to them more highly than countries farther away, because the behavior of nearby countries has more of an impact than the behavior of more distant countries.

In the third stage of the game, *west* and *east* play a bilateral trust game of their own. For *east,* I assume that, as for the potential allies, there are simply two types, nice and mean, with a probability p_{EN} that *east* is nice and prefers to reciprocate cooperation, and consequent a probability of $p_{EM} = 1 - p_{EN}$ that *east* is mean and prefers to exploit cooperation.

For *west,* I posit four possible types, two nice types and two mean types. Instead of one nice type for *west,* there are two different versions of the nice type. Both have Stag Hunt preferences in the trust game and so would cooperate in it if they believed that *east* was likely enough to be nice. They are differentiated by the payoffs they receive from the behavior of the potential allies. The first nice type for *west* is "isolationist" (WISO). An isolationist *west* is not concerned with the behavior of the potential allies and finds that the costs of extending a security guarantee to potential allies, c_{WISOi}, outweighs the benefit to be derived from their cooperation, b_{WISOi}. The second nice type for *west* is "internationalist" (WINT). For the internationalist *west,* the benefit b_{WINTi} from each of the potential allies who cooperates in the multilateral trust game outweighs the cost of extending the security guarantee, c_{WINTi}. The internationalist *west* values the cooperation of the allies for its own sake and hence feels no need to expand the alliance if the potential allies will cooperate without a security guarantee. In spite of its willingness to acquire new allies, the internationalist *west* is not vindictive toward *east* and is not seeking to maximize power, and hence is willing to cooperate in the trust game with *east,* if *east* is. The prior probabilities are p_{WN} that *west* is nice, p_{WINT} that *west* is internationalist, p_{WISO} that *west* is isolationist, so that $p_{WN} = p_{WINT} + p_{WISO}$.

There are also two mean types for *west,* both of which have Prisoners' Dilemma payoffs and so will defect in the final trust game with *east.* First, *west* could have "limited aims" (WLIM). The limited aims *west* is interested in expanding the alliance and not interested in cooperating with *east.* However, the limited aims *west* is not trying to maximize the size of the alliance and harm *east* at any cost. The limited aims *west* is therefore picky about who should be admitted to the alliance and tends to favor allies who cooperate, as does the internationalist *west.* For the limited aims *west,* however, the motivation is simply to have a well-regulated anti-*east* alliance, not to foster cooperation per se. Thus the limited aims *west* will offer security guarantees to cooperative allies even if the guarantees are unnecessary to get the allies to cooperate, that is, even if they would have cooperated without them. Hence, the limited aims *west* derives a benefit b_{WLIMi} from acquiring each new ally, provided that that ally cooperates. Noncooperative allies provide no net benefit. This benefit from acquiring cooperative allies outweighs the cost of extending the guarantee, c_{WLIMi}. For the limited aims type the benefit is only realized if the potential ally is brought into the alliance, not just by virtue of the country's cooperation.

The second mean type is the "expansionist" *west* (WEXP). The expansionist *west* is interested in expanding the alliance as far as possible, to maximize the size of the anti-*east* coalition. The expansionist *west* is *east*'s worst nightmare. I model this by positing that the expansionist *west* derives a payoff b_{WEXPi} from every potential ally to whom a security guarantee is offered, and this outweighs the costs, c_{WEXPi}, regardless of whether the ally cooperates. This net benefit outweighs any possible signaling effect; that is the expansionist *west* will prefer to extend unconditional guarantees even at the price of convincing *east* that *west* is mean. The expansionist *west* therefore has a dominant strategy to offer guarantees to all the potential allies, regardless of their type and likelihood of cooperation. The prior belief that *west* is mean is p_{WM}, the likelihood that *west* has limited aims is p_{WLIM}, and the likelihood that *west* is expansionist is p_{WEXP}, so that $p_{WM} = p_{WLIM} + p_{WEXP}$.

Play in the bilateral trust game between *east* and *west* is simultaneous, just like the multilateral trust game among the potential allies. The players must decide whether to cooperate or defect based on their payoffs and their beliefs about each other's type at that point in the game.

Equilibria in the Model

The model was created to analyze the dilemma between building trust and fostering cooperation among the members of a security community and breaking trust and damaging cooperation with an outside power. The two most basic questions to ask, then, are when does this dilemma arise, and how will the security community resolve it when it does. To answer these questions I now turn to the equilibria of the game.

TABLE 1. *Types of equilibria in the model*

Equilibrium name	*Unconditional guarantees*	*Conditional guarantees*	*No guarantees*
Semi-reassuring	WEXP		WLIM, WINT, WISO
Reassurance	WEXP	WLIM	WINT, WISO
Spiral	WEXP	WLIM, WINT	WISO

Note: WISO = *west* isolationist (nice), WINT = *west* internationalist (nice), WLIM = *west* limited aims (mean), WEXP = *west* expansionist (mean).

Equilibria in the game can be divided into three categories based on which type(s) of *west* extend conditional security guarantees, as shown in Table 1. In each of the equilibria, the expansionist *west* offers unconditional guarantees to all the potential allies, and the isolationist *west* offers no guarantees at all. In semi-reassuring equilibria, the internationalist *west* and the limited aims *west* offer no guarantees along with the isolationist *west.* In reassurance equilibria the internationalist *west* offers no guarantees, but the limited aims *west* offers conditional guarantees. Here the internationalist *west* reassures the *east* by not offering any guarantees. Finally, in the spiral equilibria, the internationalist *west* extends conditional guarantees, as does the limited aims *west.* This makes conditional expansion potentially provocative, as the nice internationalist *west* is behaving like the mean limited aims *west.* As we will see, however, this equilibrium is not always provocative; sometimes it can be reassuring as well.[39]

Mathematical details of the model are given in the appendix. Here I focus on a graphical representation (Figure 3) of the equilibria in the model. The vertical axis is the level of trust among the potential allies, or p_i. To keep the illustration two dimensional I focus on the symmetrical case in which $p_i = p_j$, so the level of trust among the potential allies can be considered as a single dimension. The horizontal axis is the level of trust between *east* and *west,* where again I consider the symmetrical case where $p_{WN} = p_{EN}$, so they can be represented as a single dimension. As foreshadowed in the introduction, the dilemma arises when the levels of trust are middling both between the potential new members and between the community and the outside power, hence in the center of Figure 3. This is the zone in which uncertainty about preferences is maximized.

At the top of the figure are the first and second reassurance equilibria, R1 and R2. Here the potential allies are so trusting of one another that they are willing to cooperate amongst themselves even without the added inducement of a security guarantee from a internationalist *west.* In this case the internationalist *west* can reap

39. I use the name *spiral* because this equilibrium is sometimes provocative. Perhaps "conditional spiral" would be a more accurate, if more cumbersome, name.

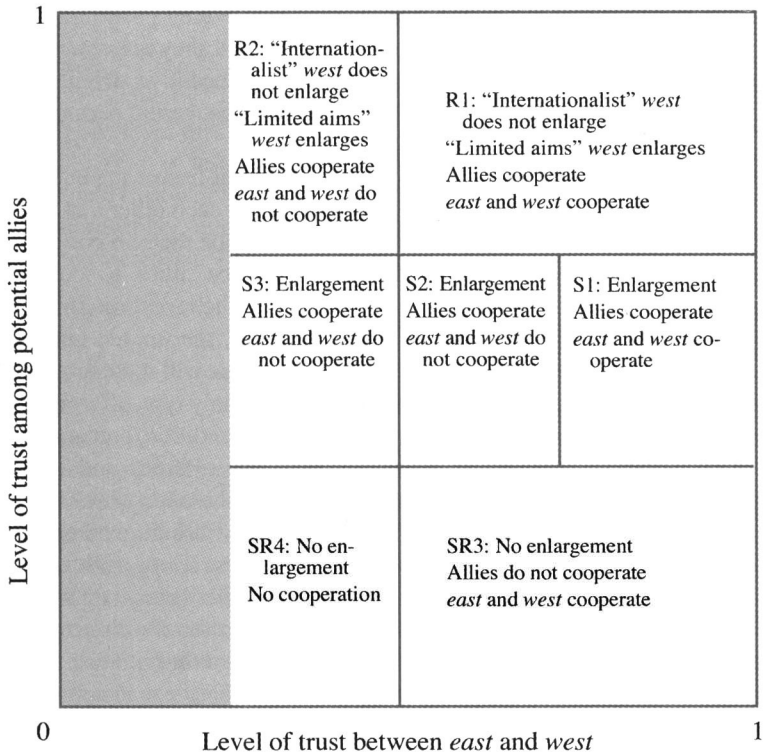

FIGURE 3. *Equilibria in the model*

the benefits of cooperation among the potential allies without paying the costs associated with the security guarantee and so has no incentive to extend the alliance to new members. Thus, the isolationist and the internationalist *west* both fail to expand the alliance, but the nice potential allies cooperate anyway. The limited aims *west* extends conditional guarantees, thereby revealing its type to the *east* and causing *east* to defect in the bilateral trust game between *east* and *west.* The limited aims *west* will be willing to do this provided the payoff from extending the alliance and the resulting mutual defection with *east* is greater than the payoff from refraining from expanding the alliance, imitating the nice types, and possibly exploiting a trusting nice *east.* The first and second reassurance equilibria are distinguished by the level of trust between *east* and *west.* In the first reassurance equilibrium, *west*'s prior level of trust for *east* exceeds a critical threshold, so the nice *east* and *west* will be able to cooperate. In the second reassurance equilibrium, the level of trust is low and this causes a failure to cooperate. That is, even though a nice *west* does not expand the alliance and this acts as a reassuring signal to *east, east* has done nothing to reassure *west,* so *west* will fail to cooperate. While some

might think that these equilibria are unrealistic because they posit high trust, there are plenty of cases where states have sufficient trust that they are able to cooperate in security affairs with each other without institutional incentives from third parties, the U.S.–Canada and U.S.–Great Britain relationships being perhaps the most salient examples.

At the bottom of Figure 3 are the third and fourth semi-reassuring equilibria, SR3 and SR4.[40] Here the potential allies are so suspicious of each other that not even the potential inducement of a security guarantee can persuade them to cooperate. Since a security guarantee would fail to persuade the nice allies to cooperate, the internationalist *west* does not bother to offer it and so behaves like the isolationist *west* by failing to expand the alliance. Here, however, the limited aims *west* also does not extend conditional guarantees, because the allies will not cooperate and the limited aims *west* wants cooperative allies as well. The only type offering guarantees is therefore the expansionist *west,* which offers unconditional guarantees. Thus some information is revealed if no guarantees are offered, and not offering guarantees serves as a signal that at least *west* is not the expansionist type. Some trust is built, but not as much as in the reassurance equilibrium, where not building serves as a perfectly reliable signal that *west* is nice. Cooperation is possible between *east* and *west* in the third semi-reassuring equilibrium where the likelihood that *east* is nice is high enough, and it is impossible in the fourth semi-reassuring equilibria where *east* and *west* are less trusting of each other.

In the reassurance and semi-reassuring equilibria, then, expansion produces no dilemmas. In the reassurance equilibria the internationalist *west* can refrain from offering guarantees because the allies will cooperate without them. In the semi-reassuring equilibria the internationalist *west* will refrain from offering guarantees because the allies will not cooperate even with the inducement of guarantees. In these regions uncertainty about preferences is low; the allies either trust each other or do not.

In the middle band of the figure are the spiral equilibria. Here trust is at a middling level among the allies, and uncertainty over preferences is maximized. The key feature of the spiral equilibria is that the internationalist *west* extends conditional guarantees to the allies and the limited aims *west* does the same, so the internationalist and the limited aims *west* behave identically. The isolationist *west* does not extend guarantees, and the expansionist *west* extends unconditional guarantees. This means that beliefs about *west*'s type change after the first round. If *west* does not expand the alliance, it is identified as isolationist and hence nice for sure. If *west* extends unconditional guarantees, *west* is identified as expansionist and hence mean for sure. If *west* extends conditional guarantees, it is identified as either internationalist or limited aims, and there will be lingering uncertainty over whether *west* is nice or mean. The likelihood that *west* is nice will be equal to

40. The first and second are not possible for the parameter values illustrated in Figure 2. See the appendix for details.

$$P'_{WN} = \frac{P_{WINT}}{P_{WINT} + P_{WLIM}}$$

This posterior level of trust p'_{WN} may be greater than or less than the prior level of trust, p_{WN}. If it is less than the prior, conditional expansion will have been provocative, and trust will have been weakened. If the posterior belief is greater than the prior, interestingly, conditional expansion will have been *reassuring,* and trust will have been increased. The posterior will be smaller, and hence expansion will be provocative if

$$\frac{P_{WINT}}{P_{WN}} < \frac{P_{WLIM}}{P_{WM}}$$

and expansion will be reassuring otherwise.

It may seem paradoxical that conditional expansion of the alliance could be reassuring to *east,* so let us examine this condition more carefully. The key is that expansion is conditional on cooperation among the allies; that is, there are limits on expansion. Conditional expansion does two things. It proves that *west* is not the isolationist type (who would not have expanded at all), and that is provocative because the isolationist *west* is nice and would cooperate in the second round. That is, by expanding conditionally, *west* has shown that it is not *east*'s ideal partner, someone who will not expand at all in order to reassure *east. West* has some interests that override its concern for *east.* However, conditional expansion also proves that *west* is not the extreme expansionist type (who would have expanded unconditionally), which is reassuring, because the expansionist type is mean and would defect in the second round. Establishing significant restrictions on who may join signals that *west* is not *east*'s worst nightmare, the hostile power bent on encircling *east* with a ring of offensively capable military bases. Thus conditional expansion has both provocative and reassuring effects. Whether conditional expansion is provocative or reassuring on balance depends on the relative weight of these two factors. As the preceding equation indicates, if the proportion of mean types that have limited aims is large (the right side of the equation), conditional expansion is likely to be provocative, because then the likelihood that *west* is expansionist will be small, so eliminating this possibility will not be very reassuring. Conversely, if the proportion of nice types that are internationalist is large (the left side of the equation), conditional expansion may be reassuring, because *west* is unlikely to be isolationist, and eliminating this possibility is not very provocative.

The more restrictive the criteria for entry, and hence the smaller the expansion, the more reassuring expansion is likely to be. The more restrictive the criteria, the harder it is to imagine the mean type choosing such criteria, that is, having limited aims that correspond to the allies selected. If NATO expansion criteria ended up selecting only the Czech Republic, among all the possible entrants, it would be difficult for Russia to interpret this as the act of an aggressive *west,* because it would

be such a strange choice from an anti-Russian point of view. In this case the proportion of mean types that would have limited aims leading them to select only this ally would be small, and hence conditional expansion could be reassuring. However, if the criteria are fairly inclusive, so that almost anyone can join, it would be easy to interpret this as an act of a mean *west* that simply wants to expand the alliance to all but a few troublesome potential allies. If the rules for admission granted entry to all the former Warsaw Pact states that applied and all the former Soviet Republics but Tajikistan, such rules would be easily interpretable as a fig leaf for a mean *west.* Here the proportion of mean types who would like to expand to include this set of allies is potentially large, making expansion provocative. Thus highly conditional expansion may be reassuring, whereas the less restrictive the conditions on membership, the more likely expansion is to be provocative.

There are three varieties of spiral equilibrium, depending on what happens in the second-round trust game between *east* and *west.* In Figure 3 I illustrate the case in which conditional expansion is provocative. On the right side is the first spiral equilibrium, S1, where even though the internationalist *west* expands the alliance, the diminution in trust between *east* and *west* is not sufficient to make cooperation impossible between them in the trust game. This can occur when the posterior level of trust between *east* and *west* is high enough to begin with to compensate for the lessening in trust caused by expansion. Here, there is no downside to enlargement.

On the left side is the third spiral equilibrium, S3. This equilibrium holds when the level of trust falls below a certain threshold, such that *east* and *west* would not have cooperated even if *west* had refrained from enlarging. Here, *west* does not trust *east,* so it does not bother to refrain from enlarging because there would be no *east–west* cooperation anyway. Once again, there is an incentive to enlarge and no cost to be paid.

In the middle is the second spiral equilibrium, and, here, enlargement really does pose a dilemma. The prior level of trust between *east* and *west* is sufficient that cooperation, absent enlargement, would take place. Enlargement, however, lessens *east*'s trust for *west* to an extent where cooperation is no longer possible. Enlargement comes at a cost; therefore, cooperation between the allies is secured by expansion, but expansion hinders cooperation between *east* and *west.*

If securing cooperation in the *east–west* relationship was more important than getting the allies to cooperate, a reassurance equilibrium would be possible in part or all of this central box. In such an equilibrium, the internationalist *west* forgoes expanding the alliance to reassure *east.* The potential allies therefore fail to cooperate, but *east* and *west* do, provided that they are nice. In this reassurance equilibrium (R3) the tradeoff of the second spiral equilibrium between cooperation among the allies and cooperation between *east* and *west* is resolved in the opposite way, in favor of establishing cooperation between *east* and *west.*

How the community resolves the dilemma in this central region will depend on the payoffs involved in the two relationships. The greater the importance of achieving cooperation among the new members, compared with maintaining cooperation with the outside power, the more likely the community is to expand. For this

reason, analysts who opposed NATO expansion tended to stress the importance of the NATO–Russian relationship and the possible harm that would result if Russia stopped cooperating. They pointed to the still unratified START II treaty in the Russian Duma, the problems of loose fissile material and the potential for smuggling, and the other issues on which the West sought Russian cooperation. Proponents of NATO enlargement tended to minimize the possible extent of Russian noncooperation, arguing that they would at worst delay action on arms control treaties such as START II, which is of lesser importance in the post–Cold War world in any event.

Conclusion

States can use restrictive membership criteria as filters that enable potential members to signal their strong interest in cooperation and keep out problematic members who would be less cooperative. NATO enlargement is a case in point. The membership criteria NATO adopted—democratization, civilian control over the military, and the resolution of border and ethnic conflicts with neighbors—are a response to uncertainty over preferences and constitute signals that identify certain states as status quo oriented, and hence as good potential alliance members. However, NATO enlargement came at a price. Expansion deepened Russian suspicions of the *west* and strengthened nationalist sentiment. In combination with the NATO conflict with Serbia over Kosovo, NATO expansion helped worsen Western–Russian relations in the second half of the 1990s. This provocative effect of expansion, however, may have been mitigated by the restrictiveness of the criteria employed. That NATO did not expand to include all countries who desired membership signaled Russia that NATO was not an unlimited expansionist alliance, bent on minimizing Russian security regardless of the cost. Excluding several potential members helped mitigate the damage done by the inclusion of others.

Russian suspicions can be further assuaged by more reassurance from the West, but they will be greatly inflated if NATO continues its expansion into the territory of the former Soviet Union. Given the political and economic status of Belarus and Ukraine, it will be many years before they can meet NATO criteria for membership, even if they were to want it. The Baltic states are another story. These countries are making rapid strides, consolidating democratic political systems, free market economies, and resolving ethnic and territorial disputes with each other and with Russia. The West's principled stand against their incorporation into the Soviet Union by Stalin gives a historical and moral legitimacy to arguments that they should be defended against potential future Russian revanchism, even as their long history before World War I as part of Russia leaves Russians feeling that they are not really

foreign.[41] In considering the next round of expansion NATO should take care not to dilute the membership requirements already set down. If anything, the criteria should be made more stringent rather than less, to maximize the potentially reassuring effect of restrictive membership criteria on those left outside.

Appendix

I consider Perfect Bayesian Equilibria of the model.[42] Off the equilibrium path, I assume that conditional guarantees convince *east* that *west* is mean with limited aims, while no guarantees convince *east* that *west* is nice and isolationist. I restrict attention to equilibria in which the nice types cooperate in the trust games if cooperation is sustainable given their beliefs and payoffs, and hence never coordinate on mutual defection when mutual cooperation is possible. I also assume that when a security guarantee does not improve the payoff, it is not offered; that is, the *west* breaks ties in favor of not offering the security guarantee. Finally, I assume that the limited aims type of *west* is at least minimally interested in expansion, that is, would find it worthwhile to expand if it had no adverse impact on the prospect of *east–west* cooperation.

Cooperation in the Multilateral Trust Game

The column vector containing the likelihoods that each of the potential allies is nice is denoted $\mathbf{p} = (p_1, p_2, p_3 \ldots p_n)'$. I assume that the w_{ij} sum to 1 for each country and that $w_{ii} = 0$. The row vector, $\mathbf{w}_i = (w_{i1}, w_{i2}, w_{i3} \ldots w_{in})$ contains the weights that player i assigns to the other players. If nice types cooperate and mean types defect, the expected payoff for the nice type of player i for cooperating can be derived as follows. If no other country cooperates, player i gets the sucker's payoff, S_{iN}. For each other country j, there is a p_j chance that they cooperate, yielding a benefit of $w_{ij}(R_{iN} - S_{iN})$, and a $1 - p_j$ chance that they will defect, yielding nothing. Thus the overall expected value of cooperating for player i is:

$$S_{iN} + w_{i1}p_1(R_{iN} - S_{iN}) + w_{i2}p_2(R_{iN} - S_{iN}) + w_{i3}p_3(R_{iN} - S_{iN}) + \ldots w_{in}p_n(R_{iN} - S_{iN}).$$

Using vector notation, this expression can be more simply expressed as

$$S_{iN} + \mathbf{w}_i\mathbf{p}(R_{iN} - S_{iN}).$$

If player i defects, the payoff is

$$P_{iN} + w_{i1}p_1(T_{iN} - P_{iN}) + w_{i2}p_2(T_{iN} - P_{iN}) + w_{i3}p_3(T_{iN} - P_{iN}) + \ldots w_{in}p_n(T_{iN} - P_{iN}),$$

41. For the debate on NATO and the Baltic states, see Asmus and Nurick 1996; Kamp 1998; and Blank 1998.
42. Morrow 1994b, 170.

which can be re-expressed as

$$P_N + \mathbf{w}_i\mathbf{p}(T_{iN} - P_{iN}).$$

The payoff for cooperation beats that for defection if

$$\mathbf{w}_i\mathbf{p} > p^{*i} \equiv \frac{1}{1 + \dfrac{R_{iN} - T_{iN}}{P_{iN} - S_{iN}}}.$$

If a guarantee has been offered, the expected payoff for the nice type of player i for cooperating is

$$g_i + S_{iN} + \mathbf{w}_i\mathbf{p}(R_{iN} - S_{iN}).$$

If player i defects, it does not get the security guarantee. The payoff is the same as before so that cooperation beats defection if

$$\mathbf{w}_i\mathbf{p} > p^{*ig} \equiv \frac{1 - g_i}{1 + \dfrac{R_{iN} - T_{iN}}{P_{iN} - S_{iN}}}.$$

Note that since $p^{*ig} < p^{*i}$, cooperation is possible for lower levels of trust if a security guarantee is offered than if it is not.

For simplicity, I assume that all of the nice types are willing to cooperate at the same level of risk, though their payoffs and weightings may vary, so that we can restrict attention to symmetric equilibria in which either all nice types are willing to cooperate or none of them are. This enables us to consider three zones of trust between the potential allies. In the low trust zone, $\mathbf{w}_i\mathbf{p} < p^{*ig}$ for all i, so that the nice allies would not be willing to cooperate even with a guarantee. In the medium trust zone, $p^{*ig} < w_i p < p^{*i}$ for all i, the potential allies would cooperate if and only if they got a security guarantee. In the high trust zone, $p^{*i} < \mathbf{w}_i\mathbf{p}$ for all i, all allies will cooperate even without a security guarantee.

Cooperation in the East–West *Trust Game*

Analogously to the previous case, one can show that cooperation is possible in the *east–west* trust game if

$$p_{EN} > p^{*W} \equiv \frac{1}{1 + \dfrac{R_{WN} - T_{WN}}{P_{WN} - S_{WN}}}$$

and

$$p'_{WN} > p^{*E} \equiv \frac{1}{1 + \dfrac{R_{EN} - T_{EN}}{P_{EN} - S_{EN}}},$$

where the prime denotes a posterior belief, after *west*'s first move.

West's *Decision on Enlargement*

The expansionist *west* has a dominant strategy to offer unconditional guarantees, and this strategy is dominated for the other three types. Given three other types of *west*—internationalist, isolationist, and limited aims—and two remaining options—offer no guarantees or conditional guarantees—there are eight conceivable patterns of behavior for the first decision, three of which are possible in equilibrium, as indicated in Table 1. In the five patterns not shown, either the isolationist *west* is offering guarantees when it could switch to not offering guarantees and convince *east* that it is nice and thereby save on the potential costs of expansion with no adverse signaling effects, or the limited aims *west* is not offering guarantees when it could defect to offering conditional guarantees, which is preferred and would have no adverse signaling effects.

Semi-Reassuring Equilibria

In the semi-reassuring equilibria the expansionist *west* offers unconditional guarantees, and the other types offer none. Therefore, if *east* observes unconditional guarantees, it is convinced that *west* is expansionist, and therefore mean. If *east* observes no guarantees, its belief that *west* is nice shifts to

$$p'_{WN} = \frac{p_{WN}}{p_{WN} + p_{WIJM}}.$$

This is greater than the prior belief, but not equal to 1, hence the equilibrium is called semi-reassuring.

If *west* deviates to offering conditional guarantees, this convinces *east* that *west* is limited aims, therefore mean. The isolationist *west* is always happy with this equilibrium, since the isolationist *west* prefers not to extend guarantees for its own sake, and extending them will have adverse signaling effects.

The internationalist *west* is happy to refrain from offering guarantees in the low and high trust zones, where guarantees would have no impact on the behavior of the potential allies. In the medium trust zone, guarantees would cause the allies to cooperate but would produce noncooperation for sure with *east*. If $p_{EN} < p^{*W}$ or $p'_{WN} < p^{*E}$, cooperation with *east* is impossible anyway, so this does not act as a disincentive; consequently, the equilibrium is impossible. If $p_{EN} > p^{*W}$ and $p'_{WN} > p^{*E}$, this is a sacrifice, so the payoffs must be compared. If we gather the benefits and costs of extending security guarantees into row vectors $\mathbf{b}_{WINT} = (b_{WINT1}, b_{WINT2}, b_{WINT3}, \ldots b_{WINTn})$ and $\mathbf{c}_{WINT} = (c_{WINT1}, c_{WINT2}, c_{WINT3}, \ldots c_{WINTn})$, we can write the payoff for enlarging the alliance for the internationalist type as: $(\mathbf{b}_{WINT} - \mathbf{c}_{WINT})\mathbf{p} + P_{WN}$. The payoff for not enlarging, which makes

cooperation with *east* possible but not certain, is $p_{EN}R_{WN} + (1 - p_{EN})S_{WN}$. Failing to enlarge beats enlarging if

$$p_{EN} > p^{*WINT} \equiv \frac{(\mathbf{b}_{WINT} - \mathbf{c}_{WINT})\mathbf{p} + P_{WN} - S_{WN}}{R_{WN} - S_{WN}}.$$

For the limited aims *west*, in the low trust zone this equilibrium works, since no allies will cooperate anyway. In the medium trust zone, allies will cooperate conditional on getting the guarantee, so the mean type will get $(\mathbf{b}_{WLIM} - \mathbf{c}_{WLIM})\mathbf{p} + P_{WM}$ if it expands, and $p_{EN}T_{WM} + (1 - p_{EN})P_{WM}$ if it does not. So not expanding is best if

$$p_{EN} > p^{*WLIM} \equiv \frac{(\mathbf{b}_{WLIM} - \mathbf{c}_{WLIM})\mathbf{p}}{T_{WM} - P_{WM}}.$$

In the high trust zone, for the limited aims type the calculation is identical because this type does not reap the benefits of cooperation without expansion, unlike the internationalist type. So the same constraint holds.

Summing up, the semi-reassuring equilibrium is possible in the low trust zone; in the middle trust zone if $p_{EN} > \max(p^{*W}, p^{*WINT}, p^{*WLIM})$, and $p'_{WN} > p^{*E}$; and in the high trust zone if $p_{EN} > \max(p^{*W}, p^{*WLIM})$, and $p'_{WN} > p^{*E}$. There are four versions of the semi-reassuring equilibrium. In the high trust zone is SR1 (see Figure 3), in which both the allies and *east* and *west* cooperate if no expansion takes place. In the medium trust zone is SR2, in which the allies fail to cooperate and *east* and *west* do if no expansion takes place. In the low trust zone are SR3, in which the allies do not cooperate and *east* and *west* do (without expansion), and SR4, in which neither the allies nor *east* and *west* cooperate. The strategies in the trust games and the boundary conditions of the equilibria are shown in Table 2.

Reassurance Equilibria

In the reassurance equilibria the limited aims *west* extends conditional guarantees, but the internationalist *west* does not. Extending guarantees therefore convinces *east* that *west* is mean for sure, $p'_{WN} = 0$, whereas not extending them persuades *east* that *west* is nice for sure, $p'_{WN} = 1$.

The isolationist *west* is again happy with this equilibrium under all conditions. Expanding would be costly and provocative.

The internationalist *west* is happy with this equilibrium in the low trust zone, where the allies will not respond to incentives anyway. In the high trust zone, the internationalist *west* is also happy; the allies will cooperate without guarantees. In the medium trust zone, expanding will cause the allies to cooperate and *east* and *west* to fail to cooperate. Not expanding makes it possible for *east* and *west* to cooperate. The conditions are therefore the same as in the semi-reassuring equilibrium, and the internationalist type will refrain from expanding if $p_{EN} > \max(p^{*W}, p^{*WINT})$. The condition on *east*'s level of trust for *west*, p_{WN}, is not binding here, because not expanding will reassure *east* completely no matter its prior beliefs.

https://doi.org/10.1162/002081801317193600 Published online by Cambridge University Press

TABLE 2. *Boundary conditions for equilibria in the model*

Equilibrium	*Trust game behavior of nice* east *and* west	*Trust game behavior of nice allies*	*Boundary condition on trust between allies*	*Boundary conditions on trust between* east *and* west
SR1	Cooperate only if no guarantees	Cooperate	$p_i > p^{*i}$	$p_{EN} > \max(p^{*W}, p^{*WLIM})$ $p'_{WN} > p^{*E}$
SR2	Cooperate only if no guarantees	Cooperate only with conditional guarantees	$p^{*i} > p_i > p^{*ig}$	$p_{EN} > \max(p^{*W}, p^{*WINT}, p^{*WLIM})$ $p'_{WN} > p^{*E}$
SR3	Cooperate only if no guarantees	Defect	$p^{*ig} > p_i$	$p_{EN} > p^{*W}$ $p'_{WN} > p^{*E}$
SR4	Defect	Defect	$p^{*ig} > p_i$	$p_{EN} < p^{*W}$ or $p'_{WN} < p^{*E}$
R1	Cooperate only if no guarantees	Cooperate	$p_i > p^{*i}$	$p^{*WLIM} > p_{EN} > p^{*W}$
R2	Cooperate only if no guarantees	Cooperate	$p_i > p^{*i}$	$p_{EN} < \min(p^{*W}, p^{*WLIM})$
R3	Cooperate only if no guarantees	Cooperate only with conditional guarantees	$p^{*i} > p_i > p^{*ig}$	$\max(p^{*W}, p^{*WLIM}) > p_{EN} > \max(p^{*W}, p^{*WINT})$
S1	Cooperate only if no or conditional guarantees	Cooperate only with conditional guarantees	$p^{*i} > p_i > p^{*ig}$	$p_{WN} > p^{**E}$ $p_{EN} > p^{*W}$
S2	Cooperate only if no guarantees	Cooperate only with conditional guarantees	$p^{*i} > p_i > p^{*ig}$	$p_{WN} < p^{**E}$ $p^{*W} < p_{EN} < p^{*WC}$
S3	Cooperate only if no guarantees	Cooperate only with conditional guarantees	$p^{*i} > p_i > p^{*ig}$	$p_{WN} < p^{**E}$ $p_{EN} < p^{*W}$

The limited aims type expands in equilibrium; not expanding would convince *east* that *west* is nice. The equilibrium is therefore impossible in the low trust zone, because expanding is pointless and ties are broken in favor of not expanding. In the medium trust zone the calculation is the same as in the semi-reassuring equilibrium, but the sign is reversed, so that the mean type will expand if $p_{EN} < \max(p^{*WLIM}, p^{*W})$. The constraint on p_{WN} is not present, because not expanding is reassuring. The same is true in the high trust zone.

Summing up, the equilibrium is impossible in the low trust zone, possible in the medium trust zone if $\max(p^{*W}, p^{*WLIM}) > p_{EN} > \max(p^{*W}, p^{*WINT})$, and possible in the high trust zone if $p_{EN} < \max(p^{*W}, p^{*WLIM})$. Note that the equilibrium is possible in the middle trust zone only if $p^{*WLIM} > \max(p^{*W}, p^{*WINT})$. There are three versions of the reassurance equilibrium. In the high trust zone is R1, in which allies cooperate and *east* and *west* cooperate provided there is no enlargement, and R2, in which the allies cooperate, but *east* and *west* fail to cooperate because of *west*'s low trust for *east*. In the medium trust zone is R3, in which allies cooperate only if *west* is mean and extends guarantees.

Spiral Equilibria

In a spiral equilibrium, not extending guarantees persuades *east* that *west* is isolationist and hence nice for sure, $p'_{WN} = 1$. Extending guarantees causes beliefs to be updated according to Bayes's rule, such that

$$p'_{WN} = \frac{p_{WINT}}{p_{WINT} + p_{WLIM}}.$$

The posterior belief p'_{WN} may be greater than or less than the prior, p_{WN}. The posterior will be smaller, and hence expansion will be provocative if

$$\frac{p_{WINT}}{p_{WLIM}} < \frac{p_{WN}}{p_{WM}},$$

and the posterior will be larger, and hence expansion will be reassuring, if the reverse holds.

In the low trust zone the internationalist *west* will deviate to not offering guarantees, because they would be pointless. In the high trust zone, the same holds. Therefore the equilibrium is only possible in the middle trust zone. The isolationist *west* is happy with the equilibrium always, as before.

The internationalist *west* must be willing to offer guarantees, in spite of the possibly adverse signaling consequences. For a high enough level of p_{WN}, *east* will still be willing to cooperate. This level can be found by equating p'_{WN} to p^{*E} and solving for the prior p_{WN}, which indicates that *east* will still be willing to cooperate if

$$p_{WN} > p^{**E} \equiv p^{*E} + (1 - p^{*E})p_{WISO} - p^{*E}p_{WEXP}.$$

Thus if $p_{WN} > p^{**E}$, the internationalist *west* will be willing to enlarge, because there will be no downside. For p_{WN} below this cutoff, the consequences of enlarging will be noncooperation between *east* and *west*. If $p_{EN} > p^{*W}$, cooperation would be possible without enlarging, so the internationalist *west* will enlarge if $p_{EN} < \max(p^{*W}, p^{*WINT})$. Otherwise,

cooperation is not possible between *east* and *west*, so the internationalist *west* is happy to expand.

The limited aims *west* must be willing to offer guarantees. For p_{WN} above p^{**E}, there is no downside, so the mean *west* will be willing to do so. Below p^{**E}, the limited aims *west* will be willing to offer guarantees if $p_{EN} < \max(p^{*W}, p^{*WLIM})$.

Summing up, there are three versions of the spiral equilibrium. In S1, $p_{EN} > p^{*W}$, and $p_{WN} > p^{**E}$, so that there is no downside from enlarging and the mean and internationalist *west* enlarge but *east* and *west* cooperate anyway. In S2, $p_{WN} < p^{**E}$, and $p^{*W} < p_{EN} < p^{*WINT}$, and the internationalist and mean *west* expand, causing the allies to cooperate and *east* and *west* to fail to cooperate, but if *west* does not expand, *east* and *west* cooperate. Finally, in S3, $p_{EN} < p^{*W}$ so that *east* and *west* would not cooperate even if *west* refrained from expanding. (*West* would reassure *east*, but *east* would not reassure *west*, so cooperation would still be impossible.)

Numerical Example

The numerical example illustrated in Figure 3 has the following parameter values. For the payoffs, I use the typical 4, 3, 2, 1 values. For the potential allies, $R_{iN} = T_{iM} = 4$, $T_{iN} = R_{iM} = 3$, $P_{iN} = P_{iM} = 2$, $S_{iN} = S_{iM} = 1$. For *east* and *west*, $R_{EN} = T_{EM} = 4$, $T_{EN} = R_{EM} = 3$, $P_{EN} = P_{EM} = 2$, $S_{EN} = S_{EM} = 1$, and $R_{WN} = T_{WM} = 4$, $T_{WN} = R_{WM} = 3$, $P_{WN} = P_{WM} = 2$, $S_{WN} = S_{WM} = 1$.

I assume five potential allies, and that $w_{ij} = 0.25$. For the benefits from getting the allies to cooperate; for the internationalist *west* I assume $b_{WINTi} = 3$, $c_{WINTi} = 1$; for the isolationist *west*, $b_{WISOi} = 1$, $c_{WISOi} = 2$; and for the limited aims *west*, $b_{WLIMi} = 3$, $c_{WLIMi} = 1$. I let the benefit from having a security guarantee be $g_i = 0.5$. For the probabilities, I assume that for the allies, $p_i = p_j$, and for *east* and *west* that $p_{EN} = p_{WN}$ so that I can illustrate the equilibria in two dimensions. I also assume that the likelihood that *west* is isolationist is 1/3, ($p_{WISO} = 0.33$) so that p_{WN} varies between 1/3 (if $p_{WINT} = 0$) and 1, hence the gray region along the left axis of Figure 3 where $p_{WN} < 1/3$.

Plugging the numbers into the formulas given earlier, we get $p^{*i} = 0.5$, $p^{*ig} = 0.25$, $p^{*E} = 0.5$, and $p^{*W} = 0.5$. The boundary conditions for *east* and *west* are $p^{*WINT} = 10p/3 + 1/3$, $p^{*WLIM} = 5p$, and $p^{**E} = 2/3 - 1/2p_{WEXP}$.