There has been a growing interest in recent years in so-called socio-
logical approaches in international relations (IR) theory. Typically
these are juxtaposed with so-called economic approaches. In the study
of international institutions and cooperation, this usually means jux-
taposing arguments that stress socially constructed “ideas” versus ma-
terial “interests” as the primary sources of motivation for actors inside
institutions. More specifically, the fairly common claim is that socio-
logical approaches stress normative motivations (social obligation, for
instance), while economic approaches stress material motivations.

The reality is a little more complex than the stereotyping (on both
sides of this divide, I might add). Sociological approaches include
many kinds of arguments, some of which are clearly not consistent
with economic approaches but some of which are. The inconsistent
ones are by now well known. Typically the difference is characterized
as being between the logic of appropriateness and the logic of conse-
quences. Economic approaches reflect the latter. The dominant eco-
nomic approach in IR theory—contractual institutionalism—gener-
ally sees institutions as collections of rules and norms that constrain
state actors primarily through material sanctions but do not funda-
mentally alter actor preferences. At best, new information provided
by, or inside, institutions will alter “beliefs” about all facets of strate-
gic interaction but will not alter the underlying goals and desires of
the actors. There are reasons for this argument—some methodologi-
cal, some aesthetic—but it is fair to say that for contractual institu-
tionalism there is no theoretical reason to expect that strategic inter-
action among optimizing actors changes their underlying preferences.
The functionalist features of this approach mean that there are objec-
tively certain types of institutional designs and features that are opti-
mal for certain kinds of cooperation problems. The problem becomes
ensuring that actors have the information and resources to harmonize
cooperation problems and institutional design (Martin 1993b; Wallander 1999). By extension, designs that are not suited for the cooperation problem at hand are inefficient. Outcomes will be suboptimal (for states and for solving the cooperation problem).

Sociological approaches, on the other hand, generally treat institutions as environments of social interaction rather than as “boxes” of material constraints. This means the research focus shifts to the non-material (for example, psychological, affective, ideological) effects on progroup behavior that interaction with other human agents can generate. Concretely, it means that social interaction can change actor desires, wants, and preferences. Indeed, one would be surprised if social interaction didn’t have these effects, given the tenuousness with which preferences over ends are held by most people. Specifically, the ways in which preferences and/or indifference curves among preferences change will be structured by three variables or moving parts: (1) socialization of actor 1 prior to entry into a social interaction; (b) socialization of actors 2 . . . n prior to entry into a social interaction; (c) and the design of the institution within which this interaction takes place. (This latter variable refers to the features of the institution that either encourage or retard the changes in the desires, wants, and preferences established prior to interaction.) Once new preferences are internalized (or old ones reinforced) actor behavior is explained by role theoretic arguments. That is, actions that are consistent with the roles associated with a particular identity are enacted on normative grounds. Conversely, actions that are inconsistent with these roles are ruled out on normative grounds or are simply left unconsidered.

But the logics-of-consequences approach does not summarize the range of sociological arguments that one could make about institutions and cooperation. One can easily construct an optimization model of human behavior where the desire being maximized is non-material and social, requiring interaction with other agents who, essentially, bestow or withdraw what is valued by the actor. As is evident, this is not much different in principle from economic arguments, though in practice the latter tend not to focus on nonmaterial optimization. Even so, as I will argue, social optimizing arguments do provide alternative explanations for cooperation that economic arguments tend to miss.

Thus, to clear up a misconception, sociological approaches do not exclude optimizing behavior. They can add a range of different things to be optimized, things that by definition only exist when two or more human agents interact socially—prestige, self-esteem, honor,
status (all rooted in identity), for instance. They also suggest that when these goals are in conflict with material goals, the latter do not always win out. Or, even when material goals are the primary consideration, there are social processes that can lock in ways of doing things such that the long-term equilibrium behavior is not materially optimal. But they also suggest that there may be motivations that are far less focused on obvious, individually beneficial consequences and that are so taken for granted that the calculus of the consequences becomes perfunctory at best. Options and outcomes may simply not be conceivable for one set of actors compared to another set, because of variation in the range of socially produced categories of analysis and evidence. Sociological approaches make the study of international institutions, therefore, much messier but more realistic—less precise but more accurate. But this messiness may be compensated by the possibility, as I will suggest at the end of this chapter, that there may be implications for the conditions under which cooperation emerges that heretofore have been missed.

Thus sociological approaches do not entirely replace economic motivations at the heart of methodological individualist predilections of contractual institutionalism. Nor do they undermine “parsimonious” theory building about the impact of institutions on the behavior of states. Indeed, they validate such theory building by allowing genuinely alternative models of actor motivation. A parsimonious explanation may in the end be quite complicated and may require the incorporation of a range of values, including sociological ones. But we won’t know this until the impact of these other variables is explored.

The aim of this chapter is to examine in preliminary fashion two social processes that may mediate the impact of international institutions on state foreign policy processes. One is persuasion and the other has been termed “social influence.” This is not a rigorous test of these processes against alternative explanations, though I will address some of the possible alternatives during the discussion. Rather, it is merely a plausibility probe that extends arguments in the literature on persuasion and social influence to cooperation in international institutions.

The chapter’s mandate—the impact of international institutions on domestic (in this case foreign policy–related) agencies and actors—does mean, however, that we need to be explicit about what causes and what effects a sociological approach would look for. The dependent variable, strictly speaking, is not foreign policy outputs. Rather, it focuses on whether there are changes in the structure, identity, and
interests of these agencies that are attributable to interaction with the international institution. This is a less demanding task than showing whether these changes have, ultimately, any impact on foreign policy. The effect of any of these changes in the foreign policy agency, or agencies, on state foreign policy will, of course, depend on the structure of the foreign policy process. But the first step in exploring the impact on outputs obviously is to determine whether social interaction has any impact on the identity and interests of the agencies and actors themselves. The independent variable, at base, is the degree of involvement in international institutions. Practically speaking, this means, How do two key variables, namely institutional design and prior identities or interests of an agent or actor, interact to produce changes in the identity, interests, and arguments of actors inside domestic institutions? The microprocesses that may govern this interaction are, as I noted previously, persuasion and social influence. Implicit in the arguments examined here is that noninvolvement will mean these effects don’t appear. The “null” hypothesis, then, is that involvement should have none of the effects that socialization microprocesses suggest ought to appear. Since socialization, however, entails change over time in any particular agent (the effects of exposure to the socializing environment is, presumably, in some sense cumulative), it is hard to run a sophisticated quasi-experimental design where one can expose identical domestic agencies and actors to different starting conditions (involvement or noninvolvement in an international institution). So the best one can do at this stage is to ask, How do these features of a domestic institution change over time after exposure to an international institution (and its social environment), and do they change in ways that are consistent with persuasion and social influence? If the null hypothesis is rejected, we can then move on to a careful comparison with cases of noninvolvement. If the null hypothesis cannot be rejected, then there is no point in going on to such comparisons.

The unit of analysis here is the individual or small group (for example, the diplomats, analysts, strategists, and political leaders who are exposed to or participate in the social environments inside international institutions). This differs from many of the sociologically oriented studies to date. For the most part, when IR specialists or sociological institutionalists have looked for the effects of social interaction at the international level the unit of analysis has tended to be the state (or state elites in a fairly aggregated way). And the predominant method has been largely correlational, looking for the
presence of new local practices that are consistent with new global norms (Eyre and Suchman 1996; Meyer et al. 1997; Finneimore 1996b, 1996c; Price 1998). This presents obvious problems when examining particular institutions as social environments since states as unitary actors don’t participate in institutions; rather, state agents do—for example, diplomats, decision makers, analysts, policy specialists, and nongovernmental agents of state principals. Moreover, treating the state as a unitary actor presents problems when applying the most well-developed literature on socialization found typically in social psychology, sociology, communications theory, and even political socialization theory. Most of the literature examines the effects of socialization on individuals or small groups. And most of this literature acknowledges that there should be a fair degree of variation in the degree of progroup conformity due to local patterns of resistance and contestation. This variation tends to get lost in the extant sociological literature in IR. This neglect of microprocesses and local variation is surprising, given constructivists’ focus on reflective action by multiple agents: If this kind of agency exists in the diffusion of norms, what happens when it runs into reflective action by multiple agents at the receiving end of these “teaching” efforts? This question is left unexplored. The result is, however, that causal processes by which systemic normative structures (constitutive, regulatory and prescriptive) affect behavior are mostly assumed rather than shown.

Thus a sociological approach allows (even demands) that the unit of analysis be the individual or small group. As Cederman (1997) points out in reference specifically to constructivism, its ontology can best be captured by the notion of complex adaptive systems whereby social structures and agent characteristics are mutually constitutive, or locked in tight feedback loops, where small perturbations in the characteristics of agents interacting with each other can have large, nonlinear effects on social structures (see also Axelrod 1997; Hammond 1998). Thus it matters how individual agents or small groups are socialized because their impacts on larger emergent properties of the social environment can be quite dramatic. This focus on individuals and small groups also enables sociologically oriented analysts to deal with the legitimate critique from proponents of choice theoretic approaches that what is observed as the normatively motivated behavior of a group at one level may be the aggregation of the strategic behavior of many subactors at a lower level (Lake and Powell 1997, 33). There are good reasons, then, for sociological studies of international institutions to “go micro.”
The empirical data come mostly from China’s behavior in international security institutions in the 1990s. From a research design perspective, this is not a bad place to go to look for the effects of sociological processes. Sociological perspectives suggest that the impact of social environments will be greatest on new members, inductees, and novices. That is, interaction between a novice or an inductee and a social group leads to changes in the actor’s preferences over ends and actions, or both. Socialization is a process of creating members, inducting them into the preferred ways of thinking and acting. In Stryker and Stattham’s words, “Socialization is the generic term used to refer to the processes by which the newcomer—the infant, rookie, trainee, for example—becomes incorporated into organized patterns of interaction” (1985, 325). Berger and Luckman define the term as “the comprehensive and consistent induction of an individual into the objective world of a society or sector of it” (1996, 130). Ichilov refers to political socialization as “the universal processes of induction into any type of regime” (1990, 1).

Noviceness is, unfortunately, undertheorized in IR. Different states face similar exogenous choices with different “life” experiences. If the two major microprocesses of social interaction are to have a profound impact on state or substate actors, it should be most obvious in novices. Who are novices in IR? The obvious candidates are the agents of newly liberated or created states or of recently integrated states, such as the newly decolonized states starting in the 1950s and the newly independent states that emerged in the wake of the Soviet Union’s collapse, for instance. These are states that are most likely to experience the IR equivalence of “primacy effects,” where early experience and information will have out-of-proportion effects on inferences drawn from later experiences and information. New states literally have had to set up foreign policy institutions, to determine what their foreign policy interests are on a range of novel issue areas, and to decide in which of a myriad social environments in IR they should participate (for example, which institutions, which communities of states—middle power, major power, developed, or developing—and which competitive and cooperative relations to foster). It is precisely in the foreign policy communities of these kinds of states where one should expect social effects of involvement in international institutions to be greatest.

China is not exactly a novice in the same way as are, say, the newly independent states of the former Soviet Union. But in terms of its involvement in international institutional life, it clearly went
through a period of noviceness in the 1980s and into the 1990s, as it moved from virtual aloofness from international institutions to participation rates that are not all that different from those of the United States and other developed states (see fig. 1).

But China is also a hard realpolitik state where novice foreign policy actors will nonetheless approach institutions with fairly skeptical arguments about the value of multilateral cooperation for achieving state security. On top of this, Chinese leaders have had to calculate the costs and benefits of commitments to multilateral security institutions first in a bipolar era and then, after 1991, in a virtual unipolar era. These are material structural conditions under which involvement in, let alone agreement with, institutions that might constrain relative power should be viewed with a great deal of skepticism. In other words, whether one stresses domestic ideational constraints or systemic material constraints, one should not expect a whole lot of cooperation inside security institutions. Moreover, as I will discuss in more detail later, the cases from which I draw the evidence for these socialization microprocesses do not entail material sidepayments or sanctions as exogenous incentives to cooperate.

Thus China is an interesting place to start a plausibility probe,
since its status as both a hard realpolitik state and a novice means that the effects of socialization microprocesses (if there are any) should be unlikely, but if they occur it should be relatively easy to observe the potential contrast between a pre- and postsocialized China. This means, interestingly, that China is at one and the same time a most likely and a least likely case for socialization.

The evidence I use comes from acquired documents, interviews, and opportunities for participant observation while researching the development of China’s arms control community and policy over the past fifteen years. Certain ethnographic methods are unavoidable because of the opaqueness of China’s policy process, especially as it relates to security issues. But the methodological value of the China case, combined with the real-world implications of China’s participation in global and regional security institutions, justifies, I believe, this kind of evidence-challenged research.

Persuasion

As a microprocess of socialization, persuasion involves changing minds, opinions, and attitudes about causality and affect (identity) in the absence of overtly material or mental coercion. It can lead to common knowledge (which may or may not be cooperative), or it can lead to a homogenization of interests. That is, actors can be persuaded that they are indeed in competition with each other or that they share many cooperative interests. The point is, however, that the distance between actors’ basic causal understandings closes as a result of successful persuasion.

Persuasion is a prevalent tool of social influence. Social psychologists have shown, for instance, that in interpersonal relationships people tend to rank the changing of others’ opinions very high in a list of influence strategies, regardless of whether the other is considered a friend or an enemy (Rule and Bisanz 1987, 192). Some political scientists have called persuasion the “core” of politics, the “central aim of political interaction” (Mutz, Sniderman, and Brody 1996, 1). In Gibson’s view, politics is all about persuasion: “Real politics involves arguments; it involves people drawing a conclusion, being exposed to countervailing ideas, changing views, drawing new conclusions” (1998, 821; see also Berger 1995, 1).

The sociology-influenced literature in IR has tended to neglect how precisely persuasion works, despite its centrality to the constructivist argument. By usage persuasion usually means something akin
to the noncoercive communication of new normative understandings that are internalized by actors such that new courses of action are viewed as entirely reasonable and appropriate. Here much of the literature tends to refer in some form or another to Habermas’s theory of communicative action. The argument is that social interaction is not all strategic bargaining. Rather, it also involves arriving at shared basic assumptions about the deep structure of the actors’ interaction: Who are legitimate players, what are the legitimate interests to be pursued, and what is a legitimate value to be bargained over? Even more important, this agreement needs to be narrow enough so that a vast range of potential equilibria that could arise in their strategic interaction becomes off-limits. Thus bargaining is not simply a process of manipulating exogenous incentives to elicit desired behavior from the other side. Rather, it requires argument and deliberation in an effort to change the minds of others. As Hasenclever, Mayer, and Rittberger put it, “the parties to a conflict enter a discourse where they first try to bring about agreement concerning the relevant features of a social situation and then advance reasons why a certain behavior has to be avoided. These reasons—as far as they are convincing—internally motivate the parties to behave in accordance with the previously elaborated interpretation and the justified expectations of others” (1997, 176–77, emphasis mine; see also Knoke 1994, 3; James 1998, 7).

The problem here is that it is not obvious that communications theory provides any insights into reasons for the success or failure of persuasion. That communicative action has to be convincing is a huge requirement, and thus far constructivists have not really shown how debates over common knowledge, for example, convince actors to agree to a “mutually arrived at interpretation” of social facts. Under what social or material conditions is communicative action more likely to be successful? How would one know? The conditions seem to be quite demanding, involving a high degree of prior trust, empathy, honesty, and power equality.

There is a great deal of research in communications theory, social psychology, and sociology that focuses on these questions. Unfortunately, there is no obvious way of summarizing what is a disparate and complex literature (see Zimbardo and Leippe 1991, 127–67). But let me try.

Essentially, for the purposes of this chapter, there are two main ways in which an actor is persuaded. First, the actor can engage in a high-intensity process of cognition, reflection, and argument about the content of new information—what Bar-Tal and Saxe call “cognitive
capacity” (1990, 122). The actor weighs evidence, puzzles through counterattitudinal arguments, and comes to conclusions different from those with which he or she began. That is, the merits of the argument are persuasive, given internalized standards for evaluating truth claims. Arguments are more persuasive and more likely to affect behavior when they are considered systematically and, thus, when they are linked to other attitudes and schema in a complex network of causal connections and cognitive cues (Wu and Shaffer 1987, 687; Petty, Wegener, and Fabrigar 1997, 616; Zimbardo and Leippe 1991, 192–97). This process of cognition, linking one set of attitudes to another, is more likely to occur when the environment cues and allows for the actor to consider these connections. In other words, it is not likely to be spontaneous. As Gibson has shown with political intolerance among Russian voters, levels of intolerance and tolerance toward political opponents, and the overall balance between these two extremes, will change if counterattitudinal arguments are presented to respondents in ways that compel them to think harder about the implications of their initial attitudes. Thinking harder simply means people are cued, and have the time, to connect the implications of their initial attitude to outcomes that might affect their interests based on different sets of attitudes (1998, 826–31). Thus the probability of some change in attitudes through cognition increases in an iterated, cognition-rich environment (where there is plenty of new information that cues linkages to other attitudes and interests) (Sniderman, Brody, and Tetlock 1991, 210; Gibson 1998, 844).

Second, the actor can be persuaded because of her or his affect relationship to the persuader: Here the persuadee looks for cues about the nature of this relationship to judge the legitimacy of counterattitudinal arguments. Thus information from in-groups is more convincing than that from out-groups. Information from culturally recognized authorities (for example, scientists, doctors, religious leaders) is more convincing than that from less authoritative sources. This will be especially true for novices who have little information about an issue on which to rely for guidance (Zimbardo and Leippe 1991, 70; Wetherell 1987, 154). Information from sources that are liked is more convincing than that from sources that are disliked. Liking will increase with more exposure, contact, and familiarity. The desire for social proofing means that information accepted through consensus or supermajority in a valued group will be more convincing than if the group is divided about how to interpret the message (Petty, Wegener, and Fabrigar 1997, 612, 617, 623, 627, 629; Kuklinski and

Obviously persuasion in the end is a combination of all of these processes, and it is hard to run controls that might isolate the effects of any one.22 People are more likely to think hard and favorably about a proposition, for instance, when it comes from a high-affect source, in part because affect helps trigger resistances to information from other sources (Mohr 1996, 81–82). But one can still identify ideal combinations that could, in principle, be tested. Given, then, an effort by a persuader to provide information with a view to changing basic principled, causal, or factual understandings, there are certain kinds of social environments that ought to be especially conducive to persuasion. These conditions occur under the following circumstances:

- when the actor is highly cognitively motivated to analyze counterattitudinal information (for example, a very novel social environment);
- when the persuader is a highly authoritative member of a small, intimate, high-affect in-group to which the persuaded also belongs or wants to belong;
- when the actor has few prior, ingrained attitudes that are inconsistent with the counterattitudinal message, say, when the actor is a novice or an inductee in a new social environment or when the perceived threat from counterattitudinal groups is low;
- when the agent is relatively autonomous from the principal (for example, when the issue is technical or is ignored by the principal);
- when the actor is exposed to counterattitudinal information repeatedly over time.

In practice, as I will discuss in a moment, these conditions are more likely to hold in some kinds of institutions than in other kinds.

These conditions suggest some more specific hypotheses about the relationship between international institutional design and persuasion effects among actors at the level of national foreign policy agencies. These hypotheses depend on some systematic conceptualization of variation in institutional design, that is, a typology of institutional forms or institutional social environments. Unfortunately, none exists in IR at the moment. But one could imagine at least several dimensions for coding institutions as social environments. Here I
borrow and expand on the typology of domestic institutions developed by Rogowski (1999):

1. membership: for example, small and exclusive or large and inclusive
2. franchise: for example, where the authoritativeness of members is equally allocated or unevenly (though legitimately) allocated
3. decision rules: for example, unanimity, consensus, majority, supermajority
4. mandate: for example, to provide information, to deliberate and resolve, to negotiate and legislate
5. autonomy of agents from principals: low through high

Different institutional designs (combinations of measures of these five dimensions) would thus create different kinds of social environments, leading to differences in the likelihood and degree of persuasion and social influence. To take one extreme ideal, persuasion is likely to be the most prevalent and powerful socialization process when membership is small (social liking and in-group identity effects on the persuasiveness of the counterattitudinal message are strongest); when the franchise recognizes the special authoritativeness of a small number of actors (authoritativeness of the messenger is likely to be high); when decision rules are based on consensus (requires deliberation, with concomitant cognition effects); when the institution’s mandate is deliberative (requires active cognition, and agents may be more autonomous since there is no obvious distribution of benefits at stake, thus there is less pressure to represent the principal); and when the autonomy of agents is high, for example, when the issue is narrowly technical or when the principal just doesn’t care much (when the principal is less attentive or relevant). All these design-dependent effects will be enhanced for novices who are exposed to the environment over long periods of time (Zimbardo and Leippe 1991, chap. 5).

Assuming that an actor enters the institution and its particular social environment with preferences and causal and principled beliefs that are at odds with those of the group, if persuasion is at work one should expect to see the actor’s convergence with these preferences and beliefs after exposure to this environment. One would also expect to see conformist behavior later in the interaction with the group that would not have been expected earlier on.
Persuasion inside the ASEAN Regional Forum?

Broadly speaking, the empirical expectations from the literature on persuasion seem borne out by China’s participation in the ASEAN Regional Forum (ARF). The ARF was set up in 1994 as a forum for high-level discussion among states in the region on security issues. It is the only intergovernmental security institution as such in the Asia-Pacific. It is relatively small (around twenty states) and is very weakly institutionalized, operating without its own secretariat, for instance. The language of the ARF underscores this informality: Members are called “participants” because “members” sounds too permanent. Around two years into its operation it finally set up intersessional meetings to allow more ongoing detailed discussions between the annual meetings of foreign ministers. The decision rule is flexible consensus, whereby the chair of all ARF meetings (who must be from an ASEAN country) has a fair amount of leeway and legitimacy to call a consensus even when some states might not be completely supportive. This norm allows the ARF to gradually move issues onto its agenda even when some actors are not entirely on board. Thus, unlike many other consensus institutions (Lindell 1988), there is, in practice, no veto or unanimity rule in the ARF.24

There are also no negotiations over the distribution of “goods.” The ARF is not treaty institutions. The working philosophy of the institution—its ideology, if you will—is counterrealpolitik, drawing on notions of cooperative security, for example, the nonlegitimacy of military force for resolving disputes, security through reassurance rather than unilateral military superiority, nonprovocative defense, and transparency. Behavior that is reassuring rather than threatening should be the rule, such that the ARF can “develop a more predictable and constructive pattern of relations for the Asia-Pacific region.”25 This security philosophy implicitly assumes that states are essentially status quo (or can be socialized to accept the status quo), and as such ARF ideology takes it as both normatively and empirically true that reassurance behavior is a better route to security than traditional realpolitik strategies. Security is positive sum.26 To this end, the normatively appropriate and empirically effective means for achieving security involve the building of trust through confidence-building measures and the defusing of security problems through preventive diplomacy and conflict management. This theory of security is supposed to be the basis for developing habits of cooperation among ARF participants.
In short, the ARF is almost explicitly designed to accentuate persuasion through social interaction. What effect has this environment had on Chinese diplomats and analysts who have participated in the ARF or its related Track II activities, such as the Council for Security Cooperation in the Asia Pacific (CSCAP) and ARF-sponsored working groups on confidence-building measures (CBMs)?

It is quite clear that the public and internal discourse in China on multilateral security dialogues in the Asia-Pacific prior to China’s entry into the ARF in 1994 was highly skeptical of their value. Such institutions, Chinese leaders worried, could be dominated by the United States or Japan. Or China would be outnumbered, and sensitive bilateral disputes where China might have an advantage in bargaining power might be multilateralized (Garrett and Glaser 1994; Yuan 1996; Johnston 1990; Interview 1996d). The skepticism of multilateralism was rooted in even deeper realpolitik assumptions about international relations where structurally (and sometimes ideologically) induced zero-sum competition among sovereign states necessitated unilateral security strategies.

Since China has entered the ARF, however, there have been some noticeable changes in the discourse. Initial statements made to the ARF (for example, Foreign Minister Qian Qichen’s comments at the first ARF in 1994) stressed what can only be seen as traditional rules of the road for the management of relations among sovereign, autonomous states. These included the five principles of peaceful coexistence, economic ties on the basis of equality and mutual benefit, and the peaceful settlement of disputes (Yuan 1996, 11). Terms, concepts, and phrases associated with common or cooperative security were absent.

By late 1996 and early 1997, however, Chinese working-level officials directly involved in ARF-related affairs began to articulate concepts that were, to a degree, in tension with traditional realpolitik arguments. Most significant in this regard was that by early 1997 ARF-involved analysts and officials had unofficially floated a concept of “mutual security.” The term meant, according to one Chinese participant, that “for you to be secure, your neighbor had to be secure,” a common security concept based on the notion of “win-win” (see Chu 1997). It is possible the Chinese may have felt under pressure to develop an original Chinese contribution to the multilateral security discourse. So “mutual security” was preferred to “common or cooperative security,” as these were terms too closely identified with the Organization for Security and Cooperation in
Europe (OSCE) process and thus were too provocative inside the Chinese policy process.

Around this time, another analyst involved in ARF-related work in a think tank attached to the State Council wrote a paper on confidence building in the Asia-Pacific. The paper provided a sophisticated explanation of Western theories of CBMs, noting for example the military reassurance purposes of CBMs. The author also elaborated on mutual security, noting that the concept was embodied in the April 1996 Five Power (China, Russia, Kazakhstan, Kyrgyzstan, Tajikistan) Treaty on CBMs (Liu 1997).

The invocation of the Five Power Treaty is important. The treaty comes as close to a OSCE-type CBM agreement as anything in the Asia-Pacific region, with provision for limits on the size and type of military maneuvers allowed within certain distances of borders, provisions for military observers and military exercises, and so forth. In internal Chinese debates over multilateralism, whether one believed that the principles of the treaty had broader applicability to the region was an indicator of one’s skepticism toward multilateralism in general. That those articulating the concept of mutual security would also invoke the Five Power Treaty as an example, precedent, or exemplar suggests that the term signified an acceptance of more intrusive and formal security institutions. Publicly during most of 1997, however, the term “mutual security” was not fully integrated into policy discourse.

This situation changed by November 1997. The Chinese paper presented to the ARF Intersessional Support Group on CBMs in Brunei explicitly noted that the Five Power Treaty embodied the notion of mutual security and could be used as a source of ideas for the rest of the Asia-Pacific. Mutual security was defined as an environment where the “security interests of one side should not undermine those of the other side. . . . This kind of security is a win-win rather than zero-sum game” (China 1997, 3).

The concept of mutual security received the highest-level endorsement when it was included in remarks by Foreign Minister Qian Qichen of China at the Private Sector’s Salute to ASEAN’s Thirtieth Anniversary in December 1997. Since then the thinking on multilateralism has continued to evolve.

As expected, the thinking is closest to ARF ideology in the agency most directly involved in regional security dialogues, namely, the Ministry of Foreign Affairs’ (MOFA) Asia Department. In 1998 the Asia Department, realizing that it required more sophisticated
theoretical arguments to bolster and justify the mutual security discourse and policy inside the policy process, commissioned a study by a respected specialist in regional multilateralism from the Chinese Academy of Social Sciences (Interview 1999c). The report explicitly argued that military power and traditional territorial-based concepts of national security were no longer the most important issues in China’s future security in the region. Rather, the report argued, China faced an increasing array of nontraditional security problems that could not be solved through the augmentation of national military power alone and thus should focus more energy on developing multilateral cooperative solutions to security problems, including greater activism in the ARF. The report noted—in recognition of security dilemma dynamics—that China’s behavior on the ground was one reason for states’ worrying about China’s rising power. To deal with this, the report argued, China had to signal that it basically accepted extant rules of international and regional order while trying to moderate these rules and norms through existing international institutions and procedures. In other words, China’s rise was a potentially destabilizing element in international relations because of perceptions of Chinese power in the past, but China had to credibly signal that it was in essence a status quo power. The report explicitly borrowed arguments and concepts from Western, including Canadian, multilateralists and included an appendix that introduced some of the multilateralist lexicon to its audience (for example, integration theory, interdependence theory, and democratic peace theory) (Zhang 1998).

The Chinese discourse on multilateralism, then, has moved quite some distance—from public skepticism to informal articulations of mutual security and common security to public affirmation of the concepts. Moreover, the concepts have been explicitly linked to a real-world institutional exemplar of these principles, the Five Power Treaty, a document that is consistent with, indeed modeled in some ways after, OSCE-style concepts. This discourse is most prominently identified with people and organizations actively involved in ARF and ARF-related Track II activities.

My argument here rests, obviously, on the critical question of whether the multilateralism discourse has, in some sense, been internalized among those working most closely in the ARF environment. The evidence that this may be the case is indirect at the moment. But China’s involvement in the ARF and related processes seems to have led to the emergence of a small group of policymakers with an emerging, if tension-ridden, normative commitment to multilateral-
ism because it is “good” for Chinese and regional security. ARF policy in China was put in the hands of the Comprehensive Division of the Asia Department of the MOFA. This division did the preparatory work for ARF meetings and Track II activities. Initially, in ARF activities the Chinese representatives were unaccustomed to the give-and-take of corridor debate and negotiation. They also came to the discussions with a watchful eye for developments that might impinge on sensitive security or domestic political issues. Over time, however, with experience in informal discussion and familiarity with the ARF norms of interaction, these officers have become more engaged, relaxed, and flexible.

Most interesting has been their apparent endorsement, within limits, of multilateralism as being compatible with Chinese security interests. More than one foreign diplomat in Beijing, interacting extensively with these MOFA officers, has suggested that their agenda is to tie China gradually into regional security institutions so that some day China’s leaders will be bound by the institutions. They see ARF involvement as a process of educating their own government. Some have even remarked that involvement in the ARF has reduced the likelihood of China’s resort to force over disputes in the South China Sea because there are now more diplomatic (read, multilateral) tools at China’s disposal (Interview 1996b, 1998d; Acharya 1998). The main conduit for the infusion of these sorts of ideas, into this group at least, has tended to be experience in Track I and II activities, not so much the absorption of academic literature on multilateralism (Interview 1996b, 1996d). It seems that this group’s influence over Chinese ARF policy may have been helped by further institutional change in China. In January 1998, the Asia Department set up a separate division just to handle ARF and Track II diplomacy. It has been the Asia Department, and in particular the ARF division, that has most actively promoted internal research on how to flesh out concepts such as mutual security, cooperative security, and common security. The goal is to use these arguments internally to justify more activism in regional multilateral institutions (Interview 1998, 1999).

There is some intriguing concrete evidence of the commitment these individuals have in protecting China’s multilateral activities from domestic political critics—hence an indication of their growing normative stake in the ARF. A senior Canadian official involved in ARF diplomacy reported that the Chinese delegates to ARF discussions apparently did not report back to Beijing any references by other delegations to the CSCE as a possible model for the ARF. The
Council on Security Cooperation in Europe (CSCE) is not just a symbol of a more intrusive, constraining regime; it is also a regime that deals with human rights (Smith 1996, 22). Downplaying this information was probably deemed important to preserve support or acquiescence for further institutionalization of the ARF. Other Canadian diplomats have reported that sometimes the multilateralists in the MOFA will help other states frame proposals for ARF-related activities in ways that will make these more acceptable in Beijing (Interview 1998d). While only anecdotal, this evidence suggests that over time the character of Chinese obstruction or resistance in its ARF diplomacy on the ground has shifted from protecting given Chinese interests to protecting Chinese multilateral diplomacy from potential domestic opposition. Tentatively speaking, one could plausibly see this as diplomacy more empathetic with the institution and less empathetic with other People’s Republic of China (PRC) constituencies that may have different views of the value of multilateralism. In other words, it may reflect an emergent solidarity with the institution, its ideology, and its participants.35

Obviously, one would like to do a more systematic comparison of the attitudes toward multilateralism of those exposed to the ARF process and of those not exposed. Also, one would like to ascertain the routes through which persuasion has or hasn’t worked inside the ARF. Why might Australian or Canadian ideas about common security be more persuasive than U.S. skepticism inside the ARF, for instance? One would expect to see a hierarchy of persuasion whereby diplomats in the MOFA directly exposed to the ARF environment would rank higher on some index of multilateralism followed by PLA officers who have also been exposed, followed by MOFA officials and PLA officers without any direct experience in the ARF. I have no way of conducting such systematic research, but based on observation and conversations with regional security specialists in China, I think this ranking generally holds. The truest believers are in the regional security division of the Asia Department in MOFA. And PLA officers who handle regional security in the Comprehensive Office of the Foreign Affairs Bureau of the Ministry of Defense are probably more sympathetic to multilateralism than are other PLA officers.

There are, of course, other possible explanations for the emergence of a multilateralist discourse within this ARF policy community. First, it could all reflect a deceptive effort to exploit cooperation from other states. A realpolitik actor would have incentives to use
such a discourse deceptively: If one believed one were in a Prisoner’s Dilemma (PD) environment, cooperative discourses might be designed to encourage others to cooperate, creating opportunities to acquire the “temptation” payoff. This would, in principle, be especially attractive to an actor in an institution like the ARF, with little or no monitoring capacity and no ability to punish defection. Many in the U.S. government view the mutual security discourse precisely as just that: a deceptive effort to redirect attention from inconsistencies between Chinese security behavior (sharp increases in military expenditures, provocative military exercises, and so forth) and the ideology of the institution, while trying to underscore the inconsistencies between U.S. bilateral alliance strategies in the region and the ideology of the institution.36

I am not convinced of the pure deceptiveness of this discourse, however, for at least two reasons. First, there is a relatively easy test of this hypothesis. If it is right, then the strongest proponents of the mutual security discourse, and the Five Power Treaty as an exemplar agreement for the region, should be the strongest opponents of U.S. bilateral alliances in the region, for example. A careful tracking of the discourse, as I tried to do earlier suggests that the strongest proponents are precisely those who in private interactions with diplomats and scholars indicate a deeper commitment to multilateralism—functional specialists in the MOFA are somewhat more pro-American voices in the strategic analysis community. While these people are generally opposed to the expansion of U.S.-Japanese security cooperation and would like to use multilateral diplomacy to pressure the United States to limit the scope of its military cooperation with Japan, they also recognize that the alliance is a reality and may indeed constrain Japanese remilitarization.37 Those who are less enamored with the mutual security discourse are found mostly in the military, and it is in the PLA where some of the strongest skeptics of the U.S.-Japan alliance are found. Moreover, the Chinese CBM proposals that were clearly biased against U.S. military power in the region (for example, observers at jointly military exercises, reductions in military reconnaissance activities aimed at ARF members, and so forth) appeared first in 1995–96, well before the “mutual security” concept emerged, and were promoted by the PLA, not the Asia Department of the MOFA.

Second, if protomultilateralist arguments in China were all only deception, we would expect that, as the ARF handles increasingly intrusive and sensitive issues that may impinge on core interests or relative power issues, the PRC would balk at further change in the
institution and agenda. That the ARF is already discussing the South China Sea in multilateral terms, intrusive CBMs, and preventive diplomacy mechanisms is evidence to the contrary. Moreover, that there are Chinese (proto)multilateralists who are now holding up the Five Power Treaty as a potential model for East Asia suggests that PD preferences are no longer uniform across the actors in the Chinese policy process. Finally, if the multilateral discourse were deceptive one would not find it promoted in internal studies and documents that are not supposed to reach the light of day.

Another possible explanation is that the acceptability of multilateralist discourses was due to new information about the benign nature of the ARF. As contractualists might put it, beliefs about, and hence strategies toward, multilateralism have changed, but preferences and interests have not. At most, more favorable attitudes toward multilateralism might serve instrumental reputational purposes. The problem I have is with the concept of “new information.” IR theory has tended to underestimate the uncertain status of new information. Information is interpreted and the same information can be interpreted differently in the context of similar institutional rules and structures. Empirically, we know that the same information will be interpreted differently depending on whether it comes from “people like us” (the information is more authoritative and persuasive) or comes from a devalued “other” (Kuklinski and Hurley 1996, 127). Social context is an important variable in how well information reduces uncertainty in a transaction and in which direction this uncertainty is reduced (for example, clarifying the other as a friend or adversary).

Thus, if all of China’s ARF decision makers were realpolitik opportunists (that is, if they believed they were playing a PD game in some form in East Asia) and if this basic worldview were fixed, then new information would be interpreted through these lenses. It is probably true that the initial signals provided by an underinstitutionalized and nonintrusive ARF in 1994 could have been interpreted as nonthreatening by realpoliticians. But as the ARF agenda and institution evolved, the signals should have been interpreted with increasing alarm by realpoliticians, since the trend lines were toward issues and procedures that could place some limits on relative military power. Yet, for a small group of China’s ARF policymakers these signals were reinterpreted in less, not more, threatening ways. That this group of policymakers eventually believed this information was reassuring, while still expressing concern that others in the policy process (with more realpolitik views of multilateralism) might see this information
as less reassuring, suggests that the information provided by the ARF is often not unproblematically reassuring. Protomultilateralists did not enter the ARF with this more sanguine interpretation of this new information. Rather, this interpretation of the information came with exposure to the ARF and related Track II processes.

Social Influence

Social influence refers to a class of microprocesses that elicit pronormative behavior through the distribution of social rewards and punishments. Rewards might include status, a sense of belonging, a sense of well-being derived from conformity with role expectations, and so forth. Punishments might include shaming, shunning, exclusion and demeaning, or dissonance derived from actions inconsistent with role and identity. The effect of (successful) social influence is an actor’s conformity with the position advocated by a group as a result of “real or imagined group pressure” (Nemeth 1987, 237). The difference between social influence processes and persuasion is neatly summarized by the phrase Festinger used to describe compliance due to social pressure: “public conformity without private acceptance” (cited in Booster 1995, 96). Persuasion would entail public conformity with private acceptance (see also Betz, Skowronski, and Ostrom 1996, 116). The rewards and punishments are social because only groups can provide them, and only groups whose approval an actor values will have this influence. Thus social influence rests on the influenced actor’s having at least some degree of prior identification with a relevant reference group. Social influence involves connecting extant interests, attitudes, and beliefs in one attitude system to those in some other attitude system; for example, attitudes toward cooperation get connected to seemingly separate attitudes toward social standing, status, and self-esteem in ways that had not previously occurred to the actor (Zimbardo and Leippe 1991, 34). Put differently, social influence involves creating or exploiting new indifference curves between previously unconnected preferences.

Like persuasion, the microprocesses of social influence are multiple, complex, and still the subject of much debate. Generally, however, the literature has isolated a number of possibilities. The first cluster of arguments comes from social identity theory but is rooted in hypotheses about the cognitive discomfort associated with perceived divergence from group norms. Social identity theory is founded on powerful evidence that mere self-categorization as a member of a
particular group generates strong internal pressures to conform to the group’s norms and practices. Group members hang their self-esteem on appearing to be progroup. If a real or imagined disjuncture between group norms and those of any individual member appears (or is exposed and pointed out), the trauma to self-esteem can motivate an actor to reduce this discrepancy through greater conformity.\footnote{40}

A second possibility has to do with social liking. Liking typically means that an individual experiences a sense of comfort interacting with others with whom she or he is perceived to share traits. Thus people will be more likely to comply with requests from friends than from strangers (Cialdini 1984, 1987).

A third possibility comes from consistency theory. There is considerable experimental and field research that suggests people are loath to appear inconsistent with prior behavior or publicly affirmed beliefs. Cialdini found, for instance, that people were more likely to continue to conform to certain norms and behaviors after taking an on-the-record action that reflects these particular norms than if they were simply asked to conform (Cialdini 1984; 1987, 169; Petty, Wegener, and Fabrigar 1997, 620). This is, generally speaking, a positive evolutionary trait, since it means people will act more predictably, which can increase trust, the credibility of commitments, and the robustness of reciprocity (Cialdini 1984, 69). This does not mean that fear of losing a reputation for trust is the only motivation to behave consistently. Cognitive discomfort appears to be an important driver as well.

The desire to be consistent generates a powerful effect on individual conformity to the group. Membership in a group usually entails on-the-record statements or behaviors of commitment (for example, pledges of loyalty, participation in group activities, commitments to fulfill a membership requirement). These behaviors, even if relatively minor, establish a baseline or threshold identity such that behavior that diverges from these identity markers give rise to discomforting inconsistencies. That is, a behavioral commitment can generate a reevaluation of identity, such that a cluster of behaviors related to the new identity becomes appropriate and is reinforced by the desire to avoid defecting against this new identity (Jones 1985, 76; Petty, Wegener, and Fabrigar 1997, 612).

Finally, the desire to maximize status, honor, and prestige—diffuse reputation or image—can be another driver behind group-conforming behavior, as can their opposites, the desire to avoid a loss of status, shaming, or humiliation and other social sanctions. Status refers to “an
individual’s standing in the hierarchy of a group based on criteria such as prestige, honor and deference.” Typically, status is closely related to others’ “expectations of ability or competent performance” (Lovaglia 1995, 402; see also Choi 1993, 113). Competency or proficiency need not mean a mechanical ability to do some task but can mean a high ability to represent some normative ideal. A competent or proficient nonproliferator is, in the eyes of an antiproliferation community, a responsible actor, a consistent, effective proponent of nonproliferation norms. Image is the public manifestation of status. Image refers to the package of perceptions and impressions that one believes one creates through status-consistent behavior.

There are numerous motivations behind maximizing status. Often status brings with it power, wealth, and deference and vice versa (Gilpin 1981, 30–33). Often, however, status markers and immediate material gains are not correlated. For example, status markers such as citations, medals, or public recognition may have no obvious material reward. Moreover, the desire to maximize status need not entail efforts to defeat others to seize status: It can entail group-conforming behavior designed to “buy” status. The reward is psychological well-being from back-patting; the punishment is psychological anxiety from opprobrium.

A second possible motivation is to maximize reputational effects attributed to particular status markers. Here status is an instrument: A good image can encourage actors to deal with you in other arenas and can help build trust leading to reciprocity and decentralized (uninstitutionalized) cooperation (Kreps 1992). There are two problems with this conceptualization, however. The first is, as Frank (1988) points out, that if people know about this instrumentality then an actor’s image or reputation as a cooperator has no advantage. So it is in the actor’s interest to make cooperation automatic and deeply socialized to make the reputation for cooperation credible. But then no advantages can be accrued, since deception is abandoned. The second problem is that instrumentality assumes the actor is seeking some concrete, calculable benefit from having a good image, an image that can be translated into leverage in some explicit, linked, immediate issue area. Yet often there are no obvious concrete benefits, or they are quite diffuse and vague. Indeed, there are sometimes concrete material costs.

This is the third reason for a concern about status. A particular high-status image may be considered a good in and of itself. Frank, for one, argues that the desire to maximize prestige and status has psychological (even physiological) benefits (1985, 32). Harre attributes
the drive to people’s “deep sense of their own dignity, and a craving for recognition as beings of worth in the opinions of other of their kind,” a craving that depends on public affirmation of one’s worth, success, and status (1979, 3, 22). Hatch notes that “the underlying motivation is to achieve a sense of personal accomplishment or fulfillment, and the individual does so by engaging in activities exhibiting qualities that are defined by the society as meritorious” (1989, 349). Franck argues, in reference to the fact that most states abide by most institutional legal commitments most of the time in IR, that conformist behavior is due mainly to a desire to be a member of a club and to benefit from the status of membership (1990, 38; on the opposite effect, shaming, see Young 1992, 176–77; DiMaggio and Powell 1991, 4).

An individual will be sensitive to arguments that his or her behavior is consistent or inconsistent with his or her self-identity as a high-status actor. This sensitivity ought to depend as well on who is making these arguments. The more the audience or reference group is legitimate—that is, the more it consists of actors whose opinions matter—the greater the effect of back-patting and opprobrium (Dittmer and Kim 1993, 9, 14–15). The legitimacy of the audience is a function of self-identification. Actors more easily discount criticisms from enemies and adversaries than from friends and allies. Thus the strength of back-patting and opprobrium depends on two related factors: the nature of the actor’s self-categorization and which other actors, by virtue of this self-identification, become important, legitimate observers of behavior.

For there to be observers, of course, one needs some “technology” or mechanism for observation. Thus social influence processes require a forum or institution that makes conformity or nonconformity a public, observable act. The forum could be something as loose as a process by which voluntary reporting on some agreed commitment is scrutinized, where defectors would stand out by either not submitting a report or by submitting shoddy and incomplete ones. Or it could be something as strict as a multilateral negotiation process by which actors are required to state bargaining positions, justify them, and then vote in some form on the outcome. Thus constructivists and contractual institutionalists are both right. Constructivists are right that normatively induced cooperation requires shared understandings of what appropriate behavior looks like. But this may not be enough without an institutional structure that provides information about the degree to which actors are behaving in ways consistent
with this shared understanding. This information publicizes the distance between an actor’s behavior and the socially approved standard. It is this distance that generates back-patting and shaming effects. In principle, the larger the legitimate audience of cooperators, the more powerful are these effects. Legitimacy here refers to the degree to which it is the social norms of this particular audience that matter and that have an effect on a potential free rider. If, say, an actor completely rejected, or were unaware of, the social norms of a particular group, then no matter what the size of that group, it could not generate back-patting or shaming effects.

To summarize, the desire to maximize status and image—diffuse reputation—say, through the accumulation of back-patting benefits (symbolic status markers, praise, and so forth) can generate a number of incentives for prosocial behavior. Prosocial behavior motivated by status maximization is not directly altruistic (though the behavior required to acquire status might be). Rather, it reflects an actor’s egoistic pursuit of social rewards and avoidance of social sanctions. But it cannot exist without the prior existence of a group and without a common understanding of the value or meaning that the group places on putative status markers. This much, at least, must be shared by the actor and the group. Identification with a group, role, or category creates sensitivity to particular kinds of status markers. The accumulation of these becomes a motivation for cooperative behavior. A particular identity is a necessary but not sufficient condition—it established who the reference group will be, the kind of status markers to be accumulated, and the kinds of behavior that elicit these markers. But action is not purely appropriate in a normative sense, nor is it purely optimizing in a materialist instrumental sense.

If these are the reasons why actors might be sensitive to back-patting or opprobrium markers, how might this sensitivity affect the decision calculus of an actor who would prefer an outcome where she or he defects while others cooperate? Here we need to look at the effects of these social rewards and punishments on this actor’s calculation of the costs and benefits of cooperation.

Assume, for the moment, that the actor in question has internalized a hard realpolitik concern about shielding relative power (military and economic) from potentially constraining commitments to international regimes. As a first cut, one can model this type of actor’s diplomacy using a simple N-person’s PD model (fig. 2). The C line represents the payoffs to the actor who cooperates when exactly k other players cooperate, in the sense of limiting their military activities. The
Fig. 2. Free riding

\(D\) line represents the payoffs to the actor from defection when exactly \(k\) other players cooperate, in the sense of limiting their military activities. If the realpolitik actor were the only cooperator (for example, if it were the only one trying to reduce some global security “bad” while others continued to maximize their relative power) its payoff would be negative, since it would be constraining its relative power while having little effect on stabilizing the security environment. If it did not cooperate while others also defected, although it could not derive any benefits from the cooperation of others, it would be better off than had it unilaterally cooperated. Thus it would pay not to cooperate even if there were no other cooperators. This payoff from defection would hold even as the number of cooperators increased. As these players contributed to a public good the actor would benefit from the provision of this good, but by free riding it would not incur the cost of providing its share of the good. Thus the payoff line from defection will always be greater than the payoff line from cooperation. In an \(N\)-person PD game, this actor should never have an incentive to cooperate.

Suppose, however, that this actor is also sensitive to its diffuse image, to the desire to maximize status markers as well as relative power. Suppose, too, that all players have a capacity to praise and shame. Under these conditions, social interaction can induce caution in the pursuit of pure defection strategies that might have an adverse effect on status. Within international organizations and institutions the participating or cooperating audience is relatively large. While the opportunities to free ride are potentially greater — given the number of potential cooperators — the scrutiny of each player is more in-
tense and state behavior is often more transparent than in bilateral relations, due to the rules of these institutions. In this context, a concern about image has two very different effects on a realpolitik actor’s payoff structure, corresponding to the effects of back-patting and opprobrium.

Back-patting is a benefit incurred from being seen as a cooperator or an active prosocial member of a group. An actor receives recognition, praise, and normative support for its involvement in the process. Back-patting can reaffirm an actor’s self-valuation, its self-categorization as a high-status actor, with concomitant payoffs for self- and public legitimation. Ceteris paribus, as the size of the cooperating audience grows, the actor accrues more back-patting benefits. Thus for every additional member of the institution, a potential defector receives a certain added payoff from back-patting as long as it cooperates. The benefits are cumulative. As figure 3 indicates this increases the slope of the payoffs from cooperation (from C to C').

Opprobrium, of course, carries social costs—a denial of the prior status and prestige of the actor—as well as psychological ones—a denial of the actor’s identity as one deserving of back-patting. Opprobrium can also be modeled as an accumulation of shaming markers that diminishes the value of free riding as the number of participants or cooperators in a regime increases. A certain social cost is incurred with each additional participant or observer in the reference group. As the group increases, the criticisms accumulate, which increases the
costs of defection. The effect, as shown in figure 4, is to depress the slope of the payoffs from defection (from $D$ to $D'$). At a certain point, an increase in the slope of the payoffs from cooperation and/or a decrease in the slope of the payoffs from defection may create a crossover point in the two lines. This is the point where the size of the audience ($k'$) is such that the back-patting benefits and opprobrium costs change the cost-benefit analysis. It is at this point that it begins to pay to cooperate, as the size of the audience increases.

When back-patting benefits and (implicit or threatened) opprobrium costs are combined, this can dramatically reduce the size of the audience needed to make it pay to cooperate (this is shown by $k''$ in figure 5). The policy objective for other players, at its simplest, is to make this $n$ as low as possible such that it does not take a large audience of cooperators to induce a defector's cooperation.

Note that the net effects of social influence on the cost-benefit calculus of cooperation in an institution are similar to the provision of material sidepayments and sanctions. It is important to point out, however, that back-patting and shaming change this cost-benefit calculus in a very different way than do sidepayments or sanctions. Side-payments or sanctions, whether provided by the institution or by a key player or players in the institution, typically have a constant effect on an actor's utility regardless of how many others cooperate, participate, backpat, or shame. Put graphically (fig. 6), the effect of

![Figure 4. Opprobrium](image-url)
The Social Effects of International Institutions

The sidepayments and sanctions is to raise the entire $C$ payoff line and/or depress the entire $D$ payoff line respectively, while not changing their slopes, such that the $C'$ payoff line ends up above the $D$ payoff line or that the $D'$ payoff line ends up below the $C$ payoff line. A sanction for defection (imposed by an enforcer or hegemon, for instance) is equally costly regardless of the size of the group cooperating. Cooperation brings higher utility than defection regardless of what other members of the group do. Put another way, the institution,
the forum in which the audience is active, is irrelevant for the initial
decision to cooperate. Back-patting or shaming, on the other hand,
have cumulative effects that depend precisely on how many others
cooperate, participate, backpat, or shame. That is, cooperation as a re-

result of material sidepayments or sanctions is not a social effect of the
institution. Cooperation due to social influence, however, is a social
effect only and would not exist without the existence of, and interac-
tion with, the group. Back-patting and opprobrium lose their impact
outside of a social group. These alternative forms of social influence
would not carry this weight if unilaterally directed by one actor to-
ward another in a bilateral, institutionless relationship.

The conditions under which one might expect social influence to
be greatest will depend in part on the type of environment to which
the national actors are exposed. Returning to the typology of institu-
tional designs I suggested earlier, given the microprocesses of social
influence, back-patting and opprobrium are more likely to be at work
than is persuasion when membership is large (maximizes the accu-
mulation of back-patting or shaming markers); when the franchise is
equally allocated (no obvious authoritative source of information);
when decision rules are majoritarian (behavior is on record and con-
sistency effects may be stronger); when the mandate involves negoti-
ations over the distribution of benefits; and when the autonomy of
agents is low (agents have to represent principals, thus reducing the
effects of persuasion on agents).

But how would one know if social influence had led to prosocial
or pronormative behavior in international institutions? Controlling
for effects on relative power and the presence or absence of material
sidepayments and punishments (discussed later), if social influence is
at work, one should expect to see the following effects:

- commitments to participate and join power-constraining in-
stitutions should take place in the absence of material side-
payments or threats of sanctions;
- arguments for joining or participating should stress back-
patting and image benefits, diffuse reputation benefits, and
opprobrium costs;
- initial bargaining positions, if stuck to, will put the state in a
distinct minority, isolating it from the cooperating audience
or reference group. Thus, commitments to prosocial behavior
will only be made when it is clear that noncommitment will
be highly isolating.
The Aggregation Problem

Before I get to the evidence for social influence, I need to be clear about the level of analysis here. The focus on sociological and social psychological microprocesses in IR compels one to look at individuals or small groups, to examine the effects of social influence on them, and then to trace the effects of these individuals on the policy process. This is a legitimate research process when one starts from a constructivist, complex adaptive systems ontology, as I do here. States as unitary actors are not persuaded, are not influenced by back-patting and opprobrium, and do not mimic.

The problem comes with observing social influence effects, in particular, on top-level decision makers. Who in the decision-making system is sensitive to back-patting or opprobrium? How do agents in institutions transmit international social influence signals back to political leaders? In the China case, why and how would people on the Foreign Affairs Leading Small Group—the institution in which the final decisions to sign onto regional security and arms control commitments were probably made—be sensitive to back-patting and opprobrium directed at the Chinese delegation in arms control negotiations? I don’t think there is an especially good theoretical or empirical answer to these sorts of questions. There is evidence that leaders tend to anthropomorphize the state themselves (indeed, the language of the state and its diplomacy is always highly anthropomorphic, for example, fatherlands, motherlands, prestige, dignity, unitary national interests, and so forth). They, and their attentive publics, tend to isomorphize criticism or praise of the state with criticisms and praise of the national in-group and of each individual in the in-group.44 In interviews with governmental and nongovernmental intellectuals and analysts in China, many of whom are critics of the regime’s repressiveness, a common argument was that leaders and the state are isomorphized when Chinese leaders interact with the external world. How Jiang Zemin is personally treated, for instance, is seen as a direct indicator of how China the collective is treated. There is no reason to believe that decision makers themselves don’t engage in this type of isomorphization and anthropomorphization. The process might go in the other direction as well: Criticism of the collective is personalized by decision makers. As one example from environmental diplomacy, Qu Geping, at the time the head of China’s National Environment Protection Agency, expressed personal frustration with the “pressure” he felt whenever he attended international conferences because of the
stories told about Chinese eating wild and endangered animals (Qu 1990, 195).

There is a growing sense among Chinese specialists of Chinese foreign policy that the current, post-Deng leadership in particular is much more sensitive both to China’s international image and, with lines between the two blurred, to their own. Jiang Zemin, in particular, is viewed as being especially sensitive to how China is portrayed and to the implications of this both for his own personal image and for his domestic legitimacy. That Jiang allowed himself to enter a very public debate with Clinton over human rights during both the U.S. and the China summits in 1997 and 1998 was seen as a sign of his strength, confidence, and stature internationally: He could debate the president of the United States on equal terms and, in Chinese views, hold his own. Jiang has emphasized that one of the key themes of Chinese foreign policy should be precisely the promotion of China’s image as a responsible major power.

China and the CTBT: A Case of Social Influence?

The Comprehensive Test Ban Treaty (CTBT) negotiations, conducted inside the UN Conference on Disarmament (CD), ran from January 1994 to August 1996. The goal of the negotiations was, essentially, to ban the testing of nuclear weapons in underground explosions. Proposals for a CTBT had floated around for decades but had run into opposition from one or more of the permanent members of the UN Security Council (P5) at different times. The original goal was to prevent states without nuclear weapons from acquiring them and to limit the modernization of weapons in the hands of states with extant weapons programs. Without an ability to test, states could not be confident in the reliability of their weapons designs, reducing the value of nuclear weapons in their arsenals. Over time the constraining effect of test bans on new and extant programs declined with advanced in computer simulation, nuclear power technology, and manufacturing skills. Nonetheless, there are certain kinds of advanced weapons designs that could still be severely limited by an inability to test through underground explosions (for example, the nuclear explosion–pumped X ray laser, enhanced electromagnetic pulse [EMP] weapons, microwave weapons, and superenhanced radiation weapons). The CTBT had also long been considered by developing states as an important pillar in the global nonproliferation regime because, in principle, it would restrain the P-5’s ability to proliferate vertically.
Right from the start, China signaled its discomfort with a CTBT. When the CTBT moved onto the CD agenda in 1993, Chinese arms control specialists and government officials let it be known that they considered the value of a treaty for reducing the demand for nuclear weapons to be quite secondary to things such as no-first-use declarations and security assurances (Wu 1994, 11–12; Liu 1993; Chen 1993). Their worry was that a treaty might freeze China’s nuclear capabilities in perpetual inferiority vis-à-vis the two major nuclear states. In particular it might inhibit the development of a survivable second-generation nuclear missile capability.\(^{48}\) Many in the nuclear weapons technical community and within the PLA have serious doubts about the credibility of China’s second strike capability. And there are some in the PLA who support the development of a limited nuclear war-fighting capability to control escalation in the event of a conventional and/or nuclear crisis. This would require a more flexible operational capacity that might include, among other requirements, multiple independently targetable reentry vehicles (MIRVs), theater and tactical nuclear weapons, a capacity to penetrate space-, air-, and ground-based ballistic missile defenses, and an ability to hit smaller, hardened military targets. Most of these systems involve smaller weight-to-yield ratios and hence new warhead designs.\(^{49}\)

Once negotiations got under way Chinese negotiators began to present bargaining positions that most observers believed were designed to slow down negotiations, at least until China had completed a test series on its second-generation warhead. Initially, negotiations proceeded under the assumption that a treaty would be ready for submission to the General Assembly (as is required before a CD-negotiated treaty becomes open for signature) by September 1996. The Chinese indicated too that they hoped a treaty would be complete by the end of 1996. It is likely that, given a spring and fall testing schedule, approximately, with two tests per year (generally four to six tests are needed for a new design), China’s decision makers would have needed at least until the fall of 1996 before they knew whether a new design was possible.\(^{50}\) Thus when negotiations began in 1994, it is safe to assume that China’s weapons testing community argued that delaying tactics might be needed for at least two years. This was a high-risk strategy, since there was no guarantee that the testing series would be successful, or successful enough, for a robust design.

Thus China’s delegation led off in 1994 with three positions that were likely to delay the treaty in some way. The first was a demand that the CTBT include a statement that committed states to unconditional
The first two positions were nonstarters (or “treaty killers”) for some or all of the other P-5. U.S. nuclear doctrine was adamantly opposed to NFU and SAs. And PNEs were opposed by almost all other states, including most of the G-21 (the developing states group in the CD), mainly because it was too difficult to differentiate between a PNE and a test explosion for military purposes, making verification of a CTBT impossible. Nonetheless, the Chinese apparently pushed for PNEs to delay the negotiation process if necessary. The third position was considered by many delegations to be frivolous, redundant, and excessively expensive.

These three positions were the core of China’s bargaining in the CTBT from 1994 to mid-1995. By the late summer of 1995, however, two additional positions emerged, both having to do with verification. One stated China’s opposition to the use of national technical means (NTM—information gathered from spy satellite and other technical means of espionage) as a source of information for triggering on-site inspection (OSI) of a suspected CTBT violator. China’s negotiators insisted that under no circumstances could NTM be used, given the asymmetries in NTM capabilities around the world. The other position concerned the decision rules leading to OSI. Essentially the Chinese supported a green light system; that is, OSI would not start automatically when a state requested OSI of a suspected violator. Rather, it would start after a decision by the treaty’s Executive Council to do so. The Chinese proposed a voting rule of two-thirds. Western states wanted a red light system where OSI started automatically unless stopped by a decision of the Executive Council.

As negotiations entered the first period in January 1996, China continued to stick with its treaty-killing positions on NFU and PNEs, as well as its positions on NTM and OSI. Observers and members of other delegations were unsure whether, with these stances, China would even sign a treaty. A major break in the Chinese bargaining occurred with the introduction in late May 1996 of a “clean” chair’s draft treaty. The draft left out language proposed by China on NFU/SAs and PNEs. This was, in a sense, a face-saving way of dropping these two positions: It required only that China agree to use the
clean draft as the basis of negotiations instead of having to publicly retreat from its NFU and PNE positions.\textsuperscript{54} Sha Zukang, the head of the delegation and the chief decision maker on arms control in the MOFA, had announced China’s flexibility on PNEs after he returned from Beijing in early May. This suggests that it was during the break between the first and second negotiating periods (April) that the Chinese interagency process made a final decision to abandon the PNE position. The reason, apparently, was that China did not want to be seen as the major obstacle to the treaty.\textsuperscript{55}

The Chinese position on OSI and NTM did not change, however, with the introduction of the clean draft. The draft treaty accepted a green light system for OSI but proposed a simple majority rule for triggering inspections rather than China’s preferred two-thirds rule. It also allowed for some use of NTM as a source of evidence for requesting OSI. Both of these were resolved after intensive bilateral negotiations with the United States in late July, around the time of China’s last nuclear test. The United States, intent on getting a treaty before the United Nations General Assembly (UNGA) opened in September 1996, was willing to compromise with China on OSI: Both sides agreed to a three-fifths voting rule for triggering OSI, a compromise between the U.S. preference for simple majority and the Chinese preference for two-thirds.\textsuperscript{56} The Chinese essentially caved on NTM, however, as most states were supportive of some role for NTM. With these compromises made, China signed the treaty and declared a unilateral moratorium on testing, abandoning at the last moment its long-standing position that it would stop testing with entry into force.

How precisely did China get to this point when it had initially considered the treaty to be potentially detrimental to China’s relative power capabilities? Let me address the alternatives to a back-patting/opprobrium explanation first. The simplest explanation—drawing from certain realist premises—is that the treaty was, in the end, costless. This argument could come in two forms. The first is that China had, by July 1996, finished its testing series for a second-generation warhead design; the testing and weapons communities were essentially satisfied with the information gathered from the tests.

There are at least two problems with this argument, however. First, many in the testing and weapons communities have since expressed dissatisfaction with the treaty. Many consider the treaty to be a sacrifice for China.\textsuperscript{57} While it is possible that sufficient information came from the last series of tests to design a newer, lighter, and
smaller warhead and to increase confidence in the safety of warheads, it is not clear that the weapons designers believe this does much to reduce the technical asymmetries (for example, relative power capabilities gaps) between China’s nuclear weapons and those of the superpowers. Nor can China test for the reliability and safety of the 1996 design. Even U.S. government specialists in Chinese nuclear weapons believe that the moratorium on testing will create difficulties for the Chinese, with stockpile aging after 2000. There are continuing complaints that U.S. possession of simulation technology and China’s lack of such technology mean that the United States can develop new generations of weapons designs while China is essentially frozen with its second-generation warheads. In other words, the CTBT may worsen China’s relative capabilities (Zou 1998, 43–44).

Second, there is some evidence that China’s leaders essentially decided to sign the treaty in the fall of 1995, prior to the negotiation session in January 1996. A comparison of the composition of China’s CTBT delegation over time is revealing on this score. It wasn’t until the January session that China sent the most senior official in charge of the nuclear weapons and nuclear arms control research in the military-industrial complex, General Qian Shaojun. Qian’s presence probably indicated that the Chinese believed negotiations had entered the endgame, that a treaty was inevitable, and that now the task was to ensure that the provisions that really mattered to China were in the treaty. Given the quickened pace of negotiations in an endgame, Qian’s presence also reduced the amount of time it would take for China’s nuclear weapons community to respond to changes in the treaty. Qian was also, roughly speaking, the military’s equivalent to Sha Zukang. This meant that decisions on how to respond to negotiations could be made more quickly and between relative equals, both probably with a mandate to decide on China’s final bargaining positions. To have this mandate, however, they would have needed instructions basically approving China’s signature in the first place, thus the likelihood that the basic decision to sign was made by China’s top political leaders in late 1995.

The timing of China’s positions on OSI and NTM would roughly fit a decision to sign in 1995. From the third negotiation session in the late summer of 1995 through the first negotiation session of 1996 China focused its bargaining on verification and dropped the treaty-killing provisions on PNEs and NFU. That China introduced two key positions related to verification in late 1995 suggests that its leaders had deduced that a treaty was inevitable, that China’s treaty-killing
positions would run into tremendous flack in the CD, and that China should start to focus on those parts of the treaty that it could really change to its advantage.

Thus the decision to join almost certainly came prior to at least the last two nuclear tests, both conducted in 1996. This would suggest that the decision came before there was conclusive information about the success of the second-generation warhead design.

A second general realist explanation for why the treaty was probably costless relates to India’s refusal to sign unless the CTBT explicitly included a time-bound commitment from the P-5 for nuclear disarmament. India’s refusal meant that the treaty would not enter into force and thus China would not be legally restricted from testing as long as India was outside the treaty. There are some problems with this argument too. While the timing is unclear, if China did indeed decide to sign in 1995, this came before India began to link its signature to a timetable for disarmament, that is, before China could have known that India’s position would prevent entry into force. The Indians had been hinting at such linkages starting in January 1996, especially leading up to, and after, the elections in May, when the Hindu nationalist BJP formed a coalition government. But it wasn’t until June that India stated firmly it would not sign unless there was a time-bound commitment to disarmament. China had insisted on India’s inclusion in the treaty as a condition for entry into force, but this insistence had come well before India came out in opposition to the treaty (Zou 1998, 36).

In a sense, the more interesting Chinese decision was the one to enter the negotiations in the first place in 1994. At that time India was a strong supporter of the CTBT (the Hindu nationalist influence on foreign policy didn’t affect India’s bargaining until late 1995 and into 1996). Thus India’s later opposition could not have factored into China’s calculations. At the time, though, the British and the French appeared unhappy with the idea of a treaty. China could have probably blocked the treaty before its negotiations had started, linking with the other two medium-sized nuclear powers to prevent a P-5 agreement to negotiate a treaty in the CD. It could have done this behind the scenes, with little fear of superpower retribution. Indeed, before the CTBT or any of its major provisions could be brought before the CD, the P-5 in general had to reach a consensus on doing so (Zou 1998, 37). Yet China chose to accept the CD’s mandate to negotiate a treaty, even though in 1994 Chinese decision makers were even more uncertain than in 1996 about the technical benefits a test series might produce.
In short, there is little evidence that China’s decision makers believed the treaty process was relatively inconsequential for its relative military power capabilities. China’s decisions to join the negotiations and accept a treaty are still puzzles from mainstream realist perspectives.

In contrast to neorealist explanations, contractual institutionalism would look for positive or negative incentives that mediated the military costs of the treaty, for example, sidepayments and sanctions or reputational benefits for immediate exchange relationships. Interestingly, the United States concluded early on that sidepayments might be necessary to ensure Chinese signature. In November 1994 it apparently offered China help in developing computer simulation technology that would allow the Chinese to maintain confidence in the safety and reliability of their weapons. According to one highly placed Department of Defense official the offer came to nothing because the Chinese didn’t have sufficient data from their tests to use the software effectively. Apparently another approach was made in 1996, but nothing came of it either. The Chinese military has continued to float suggestions that the United States provide this kind of technology, implying that China alone among the P-5 still doesn’t have it.62 Some nuclear warhead designers in the United States also doubt that China could develop sophisticated new warhead designs using simulation technology alone.63 So, on balance, the Chinese do not appear to have received any useful technological sidepayments to join.

There were also no threats of sanctions—at no point was there any implicit or explicit threat of economic or technological sanctions should China not join CTBT negotiations, should it delay, or should it not sign. Indeed, since the election of a large number of conservative republicans in the 1994 congressional elections, the U.S. Congress has been very leery of the treaty.

What about short-term reputational incentives? In a sense this is the hardest to deal with precisely because the behavioral effects one would expect to see are essentially identical to diffuse image or status concerns. But I have two problems with the immediate reputational argument. The first is that institutionalists typically haven’t developed scope conditions that might differentiate between immediate reputational concerns and very diffuse ones. This makes the claim that a decision maker was worried in an instrumental fashion about his or her reputation as a cooperator almost unfalsifiable. In this specific instance, there was no particular exchange in the Sino-U.S. relations to which one can point where Chinese leaders believed that signing onto
the CTBT would lead to a better deal for Beijing. Sino-U.S. relations had deteriorated in the wake of the March 1996 Taiwan Strait crisis, but both sides moved quickly at the highest political levels to improve relations, independent of the CTBT. Moreover, since the CTBT was opposed by the U.S. Senate, it is unclear what political leverage China could acquire with the making of U.S.-China policy by signing onto the treaty. It certainly bought no credit with Congress. Second, in none of my interviews did my Chinese interlocutors mention this short-term instrumental consideration. Indeed, one could plausibly expect that such an admission would be easier to make than an admission of diffuse image concerns. The former implies that any Chinese concessions are at least justifiable, given the need to maintain good relations with the most powerful state in the system. This is easier to sell publicly than to admit that concessions were made because of a diffuse sense of pressure from a large number of states. This is something, presumably, decision makers would prefer others not to know. Thus, admitting that diffuse image concerns were a critical motivation would seem to be a more credible statement than admitting that instrumental reputational calculations were key.

But what, then, is the positive evidence for social influence effects? The consistent refrain in interviews I conducted with people involved in the interagency process and with observers of Chinese bargaining from other countries was that China could not stay out of, or in the end sabotage, the CTBT because of the costs to China’s international image and because of the status benefits from participating in one of the pillar treaties of the nonproliferation regime, a regime supported by an overwhelming majority of states in the system. The language used by Chinese interlocutors to discuss joining and then signing was status oriented. The CTBT was a “great international trend”; there was a nebulous “psychological pressure” to join once the United States, Russia, United Kingdom, and France had committed and there was clear strong support in G-21; China’s signing was consistent with it being a responsible world power, and joining the treaty was part of a “global atmosphere,” such that China would have been isolated had it ignored this atmosphere. One of the members of the Chinese CTBT delegation argued publicly (the first statement of this kind as far as I am aware) that one of the key reasons why China ended up supporting the CTBT was opinion among developing states: “Taking into account its historical friendly relations with them, China had to maintain its image in third world countries. China’s image as a responsible major power is reportedly moving to
the fore. The necessity of maintaining its international image was a reason for China’s decision to adjust its position on the CTBT negotiations” (Zou 1998, 15, emphasis mine). These are unusually direct admissions of the impact of this form of international pressure from a regime that has traditionally publicly claimed that diplomatic pressure on China is counterproductive. That the Chinese bargained hard over verification issues—in particular OSI—even in the face of considerable dismay among delegations, does not undermine the argument about social influence. Bargaining to dilute the verification and punishment elements of the treaty in the last months of negotiations was premised on the existence of a basic acceptance of the core distributional features of the treaty.

There is, as one might expect, given the nature of the institutional environment in which China was operating, little evidence that normative persuasion affected the decision making behind CTBT accession. In this case one would look for two arguments in particular that would reflect an internalization of the theory behind the CTBT’s value as a treaty. The first would be arguments in the policy process to the effect that the CTBT would have a critical stabilizing effect on vertical and horizontal proliferation. This was after all the primary argument made in favor of test bans well back in the 1950s. Even though technological change through to the 1990s meant that non-testing would not have as dramatic a constraint on the development of some new weapons designs (particularly for the United States with its advanced simulation technologies), the general argument behind the CTBT still held—a ban on testing would reduce the reliability of, and hence the value of, nuclear weapons, putting some constraints on vertical and horizontal proliferation. One did not find this argument in the Chinese policy process. Indeed, Chinese arms controllers involved in the process continued to express doubts about the value of the CTBT in this regard.

A second argument, somewhat less obvious and more controversial among the proponents of a CTBT, would be that the CTBT does not undermine the deterrents of the nuclear states because, above a certain threshold of capabilities, more is redundant and does not buy security. In other words, for many advocates the CTBT has been implicitly premised on the notion of minimum deterrence: Once a nuclear power has enough to inflict unacceptable damage, there should be no need for continued testing. Minimum deterrence would be a transitional strategy toward a nonnuclear world. The question of whether Chinese strategists accept the notion of minimum deterrence
is controversial. There is substantial evidence that many believe minimum deterrence makes a virtue out of necessity and that China needs a somewhat larger, more flexible limited war-fighting capability to control escalation from conventional to nuclear conflicts and to control nuclear escalation itself. While the evidence as to the impact of these ideas on force posture and operations is unclear, there is no evidence that the concept of minimum deterrence was used in the Chinese policy process to support signing the CTBT. Thus persuasion—in this case the internalization of new cause-and-effect arguments about how to achieve security embodied in the philosophy of a CTBT as an institution—appears not to have been at work in this case.

In sum, it appears that Chinese decision makers’ sensitivity to China’s image and status as a responsible major power was a critical driver in their decision to sign onto the CTBT. This sensitivity—a function, I would argue, of an emerging identity in which its key status markers come from participation in international institutions that regulate interstate affairs—was especially acute because of the high profile and legitimacy of the nonproliferation regime, of which the CTBT was one important pillar. This identity created new trade-offs between status and relative power that had not existed in the Maoist (or early Dengist) period in Chinese foreign policy. The pressure to choose status over relative power was, as one would expect, especially great in such an institution. Can one develop a metric for determining a distribution of states (and state identities) in terms of sensitivity to status and security? Possibly. Indeed, it would be essential for comparing across identity types and for further testing social influence hypotheses about the relationship between institutional design and back-patting or opprobrium effects. I think that the CTBT case shows, at least, the plausibility of these hypotheses.

Conclusion

This chapter has examined some broadly sociological arguments about the effect of participation in international institutions on domestic foreign policy processes. To summarize the arguments and evidence, let me simply outline some of the implications for understanding how cooperation between domestic actors (in this case, foreign policy actors) and international institutions might work.

The first set of implications has to do with persuasion and institutional design. Typically, contractual institutionalists argue that efficient institutional design depends on the type of cooperation problem; for
example, a PD-type problem requires information (monitoring) and sanctions, and an assurance problem primarily requires reassurance information (Martin 1993b). The flip side is that one can identify inefficient institutional designs for a particular cooperation problem as well (for example, an institution that is designed only to provide assurance information but has no monitoring or sanctioning capacity would be inefficient for resolving PD-type problems). In addition, Downs, Rocke, and Barsoom (1997) argue that so-called transformational institutions (inclusive institutions that bring genuine cooperators and potential defectors together in an effort to instill norms and obligations in the latter) are less likely to provide efficient solutions than a strategic construction approach. This latter approach to institutional design stresses exclusive memberships of true believers where decisions are made on the basis of supermajority rules. The gradual inclusion of potential defectors under these conditions ensures that the preferences of the true believers predominate as the institution evolves. Downs, Rocke, and Barsoom’s critique of the transformational approach rests explicitly on skepticism that the preferences of potential defectors can change through social interaction. The problem is that they do not develop any operational indicators for determining whether interests, desires, and goals undergo endogenous change as state actors move into and become more involved in institutions.

In any event, if one considers the possibility of preference change among individuals and small groups of actors inside institutions, then one is forced to revisit these common notions of efficient institutional design. An institution that appears inefficient to contractual institutionalists (for example, an assurance institution for a PD problem), may actually be efficient for the cooperation problem at hand. If, say, a player (or subactors in a policy process) with PD preferences can be socialized (persuaded) to internalize stag hunt preferences through interaction in a social environment with no material sanctioning or sidepayments, then assurance institutions may work in PD-like cooperation problems. An efficient institution might then be reconceived as the design and process most likely to produce the most efficient environments for socializing actors in alternative definitions of interest. As I have argued, the ARF case, and the sociological literature on which it rests, suggests that such an institution may have to be informal, weakly institutionalized, and consensus based—the opposite of an institutional design appropriate for dealing with PD problems.
A second broadly theoretical implication of social influence is for collective action. Social influence effects may provide insights into how groups resolve the collective action problem that hinders resolving collective action problems. That is, traditionally scholars have argued that a critical solution to free riding is to offer material sidepayments (and sanctions) to make collective action pay for the individual. The conundrum has been, however, that the offering of sidepayments is itself a collective action problem. Who will take up the burden of offering sidepayments, given the resources required to do so? Hegemons and activists are usually part of the answer to this puzzle (though why activists should exist in the first place is hard for collective action theorists to specify a priori). Social rewards and punishments, however, are a particularly interesting kind of incentive for overcoming collective inaction. They are relatively cheap to create but are often infused with a great deal of value. This means that new status markers can be manufactured and distributed without necessarily diminishing their value. In principle any member of a group therefore can provide social sidepayments at relatively low cost, indeed at zero cost if the member can also receive these kinds of sidepayments for providing them to others. This is, after all, what back-patting entails—a mutual, virtuous circle of bestowing and receiving social rewards. Cheap, but social, talk, then, can indeed be cheap to produce but nonetheless can still be considered credible precisely because of its social value. Thus, because status markers are so highly valued, it doesn’t take much of a costly commitment by providers of these markers to establish the credibility of promises to bestow, or of threats to retract, these markers. All this suggests that one reason why collective action problems are often less frequent and debilitating than theorists expect (Green and Shapiro 1994, 72–97) may have to do with the fact that actors are also motivated by the desire to maximize social rewards and that these are relatively easy for groups to produce and distribute.

Following from this argument about collective action, social influence arguments also suggest that the conventional wisdom about optimal size of institutions and groups may need rethinking. Drawing from Mancur Olson’s work, the contractual institutionalist argument is that ceteris paribus more actors make cooperation more difficult (collective action problems, problems of monitoring and punishing defection, and so forth). Transaction costs increase with more actors. Decentralized institutions are therefore handicapped in dealing with “problems of transaction costs and opportunism” (Abbott and Snidal 1998, 15). From a social influence perspective, however, more may be
better. Status back-patting and opprobrium effects are likely to be stronger when the audience or reference group is larger.

In sum, sociological approaches introduce at least two new and different processes by which actors come to cooperate inside institutions—persuasion and social influence. These, added to the more traditional focus on material sidepayments and sanctions and on instrumental reputational concerns, present a more complete range of possible causal processes at work inside international institutions. Indeed, they bring to IR theory a set of motivations for behavior that have long been recognized as central to the study of politics. They suggest broad scope conditions for examining the relationship between international institutional design and state action—in large, high-profile, formal negotiation institutions one should expect that social influence, material sidepayments or sanctions, and reputational concerns will be the most likely candidates for explaining cooperation. In smaller, informal, consensus-based deliberation institutions, the key processes will likely be those, or some combination, of persuasion, instrumental reputation, and sidepayments or sanctions. In bilateral relationships material sidepayments or sanctions, instrumental reputation, and possibly persuasion would be processes for which to look.

These plausibility probes suggest, therefore, that with further refinement some important and genuinely competitive empirical arguments could be set up to test the variable effects of institutional designs on state behavior, tests that could examine the relative importance of social processes and motivations and economic processes and motivations. This is the first step to truly understanding the scope conditions under which these kinds of motivations are or are not at work. Surely this will make for a more accurate and precise understanding of the effects of international institutions than if one were to start from the assumption that only one or the other is relevant.

Notes

I would like to thank Jean-Marc Blanchard, Dan Drezner, Kathryn Hochstetler, Steve Krasner, Helen Milner, participants at the conference “The Interaction of Domestic and International Institutions,” and the anonymous reviewers for their comments and criticisms.

1. This juxtaposition is increasingly under attack from two very different sources. Critical constructivists have pointed out the illogic of assuming that there is an arena of human activities where “ideas” are not present in the in-
interpretation of interests (Laffey and Weldes 1997). Some of those coming out of the contractual institutionalist community have also argued that there is no inherent contradiction between a thin rational choice epistemology and an interest in explaining where preferences come from. See Katzenstein, Keohane, and Krasner 1998.

2. Contractualists will admit that sanctions could be social or that non-material considerations such as reputation could provide incentives to accept the constraints of the institution. As I will note this points to an area of convergence with sociological approaches. As a practical matter, however, almost always the empirical work focuses on material incentives. Moreover, the goal of improving one’s reputation from one set of interactions is usually to apply this reputation to acquire, most commonly, material benefits from another set of interactions.

3. See Frieden 1999 for a sophisticated statement of the reasons for assuming fixed preferences.

4. I am referring here to mimicking. I have not discussed mimicking—copying the descriptive norms of a group as opposed to internalizing its prescriptive norms—and its path-dependent effects on organizations because one could debate whether it clearly falls within a sociological perspective.

5. For similar critiques see Checkel 1998, 332; Moravcsik 1997, 539.

6. This may change as scholars pick up on Finnemore and Sikkink’s summary of some plausible causal processes (1998).

7. This is, after all, the point of much of the work on how transnational networks affect state behavior (Keck and Sikkink 1998; Evangelista 1999), “teaching” and the diffusion of norms, and the creation of national interests (Finnemore 1996b).

8. Ruggie calls this a focus on “innovative micro-practices,” a hallmark of constructivist research (1998, 27).


10. See for instance the argument in Chafetz, Abramson, and Grillot 1996. The authors argue that the debate in the Ukraine in the 1990s over whether to sign the Non-Proliferation Treaty (NPT) was in large measure a debate over whether Ukraine’s identity was that of a great power (hence it could legitimately keep and develop nuclear weapons) or a middle European power (in which case it should denuclearize and join the NPT).


12. I do not use socialization here in a normative sense. Nor do I believe that Chinese actors prior to entry into these institutions were in some sense unsocialized. They were merely differently socialized.

13. The documents come from the Conference on Disarmament record, a collection of over one hundred conference papers presented by Chinese arms controllers at international conferences over the last ten years, and articles from major internal and open sources journals. In addition I have conducted around ninety interviews with Chinese, American, Canadian, Japanese, and
Singaporean arms control officials over the past approximately six years on the topic of China’s involvement in international security institutions. Finally, I have been a participant observer in a number of Track II conferences on security, where I could interact intensively with Chinese arms control experts and officials, including two sessions of the Summer Study in Science and World Affairs, organized by the Union of Concerned Scientists (1993, 1996), the ISODARCO Beijing Arms Control Seminar (1994), meetings between the Ford Foundation and several arms control research institutions in China (1998), meetings between the MIT arms control group and Chinese nuclear weapons and missile design institutions (1996), and the Monterey Institute of International Studies Second U.S.-China Arms Control Seminar (1999).

14. This section draws heavily from Johnston 1999b.
15. See Johnson 1993, 81, on this point.
16. For an excellent exegesis of Habermas’s theory of communicative action, see Risse 1997.
18. Despite the volume of this literature, “To date there is precious little evidence specifying who can be talked out of what beliefs, and under what conditions” (Berger 1995, 8).
19. Sometimes scholars identify a third condition for persuasion, namely, the characteristics of the persuadee herself or himself, for example, innate cognitive processing abilities (the greater, the strength of existing attitudes, the greater the degree to which an individual is sensitive to appearing consistent).
20. Or as Gibson puts it, “Especially when people do not have much experience with political institutions and processes, it is easy to imagine that their initial viewpoints are poorly anchored in a highly articulated and constrained belief system, and that considerable potential for effective persuasion exists” (1998, 821).
21. There is a signaling literature that is beginning to address the scope conditions for persuasion as well. Lupia and McCubbins (1998) argue that there are two basic conditions, namely, that the persuadee believes the persuader to be knowledgeable about an issue and that the persuadee believes that the persuader’s intentions are trustworthy. Any other factors, such as ideology, identity, culture, and so forth, are only predictors of persuasion to the extent that they reveal information to the persuadee about the persuader’s knowledge and trustworthiness. The problem here is that the argument, in my view, avoids the more interesting question about the empirical frequency with which social variables such as perceived ideology, identity, and/or cultural values are in fact the primary cues that people use to determine the degree of knowledge and trustworthiness of a persuader and thus come prior to beliefs about knowledge and trustworthiness. The answer has important implications for how social interactions lead to socialization and how different institutional designs might lead to different socialization paths. Lupia and McCubbins tend to focus on the role of external forces in clarify-
ing beliefs about knowledge and trustworthiness of persuaders. Since, they argue, social and political environments are rarely ones in which persuader and persuadee interact face to face over long periods of time, the familiarity or personal interaction route to beliefs about the persuader’s knowledge and trustworthiness tends to be less common. This may be true at the national level of persuasion (for example, political messages from politicians aimed at masses of voters), but it is not necessarily true at the level of social interaction in international institutions among diplomats, specialists, and analysts. Here the first route—familiarity, iterated face-to-face social interaction—may be more common, hence affect based on identity, culture, and ideology may be more critical for persuasion than external forces and costly signals.

22. Jorgensen, Kock, and Rorbech (1998) found in a study of televised political debates in Denmark, for example, that the most persuasive debaters were those who used a small number of extended, weighty discussions of specific qualitative examples. The use of these specific, straightforward, and logical examples seemed to accentuate the authoritativeness of the debater, and they were easier for viewers to assess and adjudicate.

23. Risse argues, for instance, that nonhierarchical and networklike international institutions “characterized by a high density of mostly informal interactions should allow for discursive and argumentative processes” (1997, 18). Martin and Simmons, coming at the question from a contractual institutionalist perspective, also imply that institutions where participants are reliant on “expert” sources of information should be “most influential in promoting cooperation” (1998, 742). See also MacLaren 1980, 67–68, for similar hypotheses.

24. Efforts to buck or shirk flexible consensus decisions will generate more negative peer pressure than if clear opposition is registered through a vote (Chigas, McClintock, and Kamp 1996, 42–43). Obstinance in a consensus system threatens to undermine the effectiveness of the entire institution because this is premised on consensus. Obstinance is judged to be fundamentally at odds with the purposes of the institution. “Principled stands” against efforts to declare consensus are viewed as less principled than had they been expressed in a losing vote (Steiner 1974, 269–71).

25. ARF (1995). See also the comments by Defense Minister Hajib Tun Rajak of Malaysia, cited in Dewitt 1994, 12–13, and Lee Kwan-yew’s comments about the ARF as a channel for China’s reassuring Southeast Asia about its status quo intentions in Makabenta 1994.

26. These are the security principles of the OSCE as well. The primary difference between OSCE and ARF definitions of common and cooperative security is that the former includes human rights and liberal domestic governance as a component of interstate security. The ARF, sensitive to the postcolonial sovereign-centric ideologies in ASEAN and China, excludes this element.

27. Track I refers to official meetings and institutions: Track II refers to nonofficial meetings that involve officials, academics, and other nongovernmental
personnel. CSCAP has emerged as the primary Track II institution for the ARF, where ideas and arguments that might be too controversial at the official level can be floated and discussed less formally, without requiring national policy commitments.

28. See the paper presented by Shi Chunlai (a former ambassador to India and a key figure in China’s CSCAP committee) and Xu Jian at the ARF-sponsored Paris workshop on preventive diplomacy (November 1996). Shi appears to have been China’s first authoritative participant in ARF-related activities to have used the term “common security.” The paper argues that common security was central to the post–Cold War need for a “renewal” of old security concepts based on the “dangerous game of balance of power” (see Shi and Xu 1997).

29. They did this at the first Canada-China Multilateral Training Seminar in January 1997. The seminar brought together a small number of key officials handling the ARF in the MOFA Asia Department and a couple of analysts from China’s civilian intelligence institution—China Institute of Contemporary International Relations (CICIR)—who were also in the ARF interagency process.


31. This was the distinct impression I received when interviewing military and civilian specialists on the ARF in 1996. Interviews in 1998 and 1999 with analysts involved in discussions of multilateralism also suggested that officials in the MOFA Asia Department were more likely to see the potential applications than, say, PLA officers and traditional security analysts.

32. While an internal study, the document itself is not classified or stamped with the words “internal circulation only.” Technically the report was commissioned by the MOFA Policy Research Office, but the primary consumer was the Asia Department.


34. The MOFA-commissioned study on comprehensive security, however, clearly lists sources and notes from Canadian and American academic specialists in multilateralism.

35. One key issue is how this change in attitudes toward multilateralism has affected China’s diplomacy inside the ARF. Since this doesn’t fall within the mandate of this chapter, strictly speaking, I will not address it here. The short answer is that this growing acceptance of multilateralism has enabled China to accept changes in the institutionalization and agenda of the ARF that it would not have accepted at the start in 1994—namely, a greater level of institutionalization (for example, the creation of functionally specialized intersessional meetings) and more intrusive agendas (for example, certain CBMs, discussions of preventive diplomacy, and so forth). I expand on these changes in the ARF in Johnston 1999b.
36. The argument is outlined in Garrett and Glaser 1997. My own conversations with Pentagon officials involved in Asia policy confirms this particular interpretation.

37. This argument was made by some of the civilian analysts and military officials involved in ARF policy-making I interviewed in 1998 and 1999. Indeed, the MOFA-commissioned report on comprehensive security is explicit in stating that China should not and need not replace U.S. military superiority in the region and that China needs to balance militarily against U.S. power, especially if effective, practical multilateral security institutions can be set up in the region (see Zhang 1998, 20–21, 26–27).

38. Even this is problematic from an institutionalist perspective. As realpoliticians, the Chinese should have been especially suspicious of an institution that activist states such as Canada, Australia, and to some extent the United States supported. The information that their involvement should have supplied, for realpolitician skeptics, was precisely that the ARF was a potentially constraining institution.

39. These interpretations were consistent across public and private internal discussions.

40. On social identity theory (SIT) and the psychological discomforts of nonconformity, see Turner 1987; Gerard and Orive 1987; Stryker and Statham 1985; Barnum 1997; Axelrod 1997.

41. Keohane notes, for instance, that one of the things international institutions do is provide a forum in which an actor’s conformity with group standards can be evaluated. He links this to a more instrumental notion of reputation than I do here, however (1984, 94).

42. This distinguishes status concerns from obligation and from material interest as a source of cooperative behavior. It blends the economic language of optimizing with the sociological language of group processes or social interactions and thus helps to underscore why (small r) rationalist and sociological or constructivist approaches are not necessarily mutually exclusive. (I use small r rationalism to refer to the assumption that agents are optimizers or maximizers in the expected utility sense. Big r Rationalism refers to the ontological and epistemological assumptions of positivism and modernity [that there is an observable and knowledge external world independent of the normative preferences of the observers]).

43. The reasons behind using this type of game for illustrative purposes are twofold: First, while security dilemmas can be modeled using PD or stag hunt assumptions, the former is a tougher cooperation problem because of its assumption about opportunism. So, in a sense, using an N-person PD presents a hard case for the promotion of cooperation through social influence mechanisms. Second, empirically, it is my view a valid simplification to assume that the realpolitik of Chinese decision makers leads them to view their own security environment more in PD terms than in stag hunt terms.

44. On the question of aggregation see Finnemore and Sikkink 1998, 904.

46. This is based on conversations with scholars from the Foreign Affairs College and the Institute of World Economics and Politics at the Chinese Academy of Social Sciences, October 1998.

47. Many of the details of the negotiations come from Rebecca Johnson’s excellent reportage on behalf of the NGO community. See her “Geneva Updates” and “Acronym Reports” on the DFAX web page at <http://csf.colorado.edu/dfax/index.htm>.

48. See the summary in “Guowai dui women he shiyan de fanying” [Foreign Reaction to Our Tests], in Guowai hewuqi dongtai [Trends in Foreign Nuclear Weapons], 12, no. 6 (1994.12.2): 1. See similar concerns expressed in other internal circulation materials such as Fu 1994, 7; Wu Zheng 1996, 10; and Zou 1998, 12.

49. See Johnston 1995–96 for details on the concept of limited deterrence. Adding to the uncertainty of the effects of a test ban on China’s deterrence is that apparently there has been little discussion inside China over how many penetrated warheads against what targets is enough to constitute a deterrent. This dearth of discussion has also hindered China’s ability to think through the pros and cons of a fissile material production cutoff. The weapons community was also worried that a ban on testing would bring economic hardship to its large testing infrastructure, including the city of Malan, which serviced the Xinjiang test sites.

50. China had lobbied behind the scenes at the UNGA First Committee (disarmament) in the fall of 1995 to prevent a vote calling for a treaty by the fall of 1996. The committee in fact voted overwhelmingly for the date, and in the end China abstained rather than having to stand publicly in opposition to the vote.


52. I interviewed a member of the U.S. delegation in the fall of 1995, and he was not at all certain that China would sign.

53. Up until this point delegates were negotiating a rolling text with as many as 1,200 brackets, indicating disagreements over wording, grammar, and punctuation. A clean text was introduced by the ad hoc nuclear test ban (NTB) committee chair, Rammaker of the Netherlands, presented as a flexible fait accompli for more focused negotiation. The presentation of the clean text, and its acceptance as a basis for bargaining, signaled the endgame of negotiations.

54. The cover for China’s concession on PNEs was “an assurance that such explosions could be reviewed at periodic review conferences, although Chinese officials acknowledged that such reviews were unlikely to lead to revisions of the prohibitions” (Zou 1998, 38).

55. This is implied by Zou’s summary that “When people gradually came to the conclusion that PNE was one of the obstacles to signing the treaty, China finally made a major move on this issue in early June 1996” (1998, 23).
56. The United States wanted the triggering of OSI to be as easy as possible, hence its preference for a simple majority. China, determined to minimize the intrusiveness of the verification system, initially wanted OSI to be as difficult as possible. For a fascinating description of negotiations on this question, see Zou 1998, 31–35.

57. Recent allegations that the Chinese had “stolen” information about U.S. nuclear weapons designs have been remarkably ambiguous about what information was acquired, how, and when. The intelligence community consensus, in contrast to the more hyperbolic Cox Report, is that it is unclear what information was acquired, when, how, and with what significance for the flexibility of China’s nuclear options. Actually, the Cox Report implies in places that the CTBT, if enforced, does negatively affect the reliability of China’s second-generation warheads. Until these issues are clarified, I believe that the concerns that had been relayed in public and private by China’s testing community about the constraints a CTBT will place on warhead modernization are still credible ones. For a discussion of these issues see May 1999.

58. Interviews with analysts in the Institute of Applied Physics and Computational Mathematics (IAPCM), China Academy of Engineering Physics (CAEP), PLA, and U.S. Department of Defense. See also Li and Zhou 1997, 241. For a critique of the CTBT for its irrelevance in reducing incentives to proliferate, see Chen and Wang 1996.


60. One of the Department of Defense members of the U.S. CTBT delegation also speculated that this timeline is probably correct (Interview, January 1998). The January 1996 Chinese delegation also included a deputy director and arms control expert from the IAPCM. This too would suggest that the issue for China by January 1996 was not whether to sign but which specific provisions at the margins in the treaty that might affect the nuclear weapons community were acceptable.

61. One Chinese informant, a well-respected arms control specialist outside of the formal CTBT interagency process, claimed that the French had approached him early on about possible cooperation to prevent a CTBT. I have no way of verifying this claim independently, however.

62. The issue came up in my discussions with specialists connected to the Chinese military in January 1998.


64. These arguments come from interviews with arms control officials and experts in MOFA, the Chinese military, the Chinese nuclear weapons research community, and the U.S. Department of Defense. The interviews were conducted in 1996 and 1998.

65. The Chinese were made very aware of the specific pressure from developing states after its May 1995 nuclear test, which took place forty-eight hours after the end of the NPT extension conference. In Zou’s words, “The May 15, 1995 test inflicted the most political damage to China. . . . This test
made the Non-Aligned Movement very angry” (1998, 13). This is an interesting and exceedingly rare public admission of sharp differences between China and other developing states on nuclear disarmament issues.

66. In some cases, when I followed up with questions specifically on the role of image concerns, my interlocutors would quickly downplay what they had just said, since this implied that China was indeed susceptible to external pressure.

67. This is not dissimilar to Johnson’s (1993) argument that cheap talk, in the context of persuasion whereby interests and identities converge inside a social relationship, establishes focal points that are necessary to reduce the strategic indeterminacy of bargaining games.

68. These motivations parallel the three that James Q. Wilson (1973) identifies: material incentives, purposive incentives (persuasion), and solidarity incentives (social influence).