

COPYRIGHT NOTICE:

**Alastair Iain Johnston: Social States**

is published by Princeton University Press and copyrighted, © 2007, by Princeton University Press. All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher, except for reading and browsing via the World Wide Web. Users are not permitted to mount this file on any network servers.

Follow links for Class Use and other Permissions. For more information send email to: [permissions@pupress.princeton.edu](mailto:permissions@pupress.princeton.edu)

## Socialization in International Relations Theory

### THE STATUS OF SOCIALIZATION IN INTERNATIONAL RELATIONS THEORY

Socialization is quite a vibrant area of inquiry in a range of social sciences. It is a core concept in studies in linguistics and the acquisition of language (Schieffelin and Ochs 1986), sociology and social psychology and theories of in-group identity formation and compliance with group norms (Turner 1987; Napier and Gershenfeld 1987; Cialdini 1987; Nisbett and Cohen 1996), political science and the acquisition of basic political orientations among young people or explanations of social movements (Beck and Jennings 1991), international law and the role of shaming and social opprobrium in eliciting treaty compliance (Chayes and Chayes 1996; Young 1992, Susskind 1994; Moravcsik 1995), and anthropology and the diffusion of cultural practices, among other fields and topics. It is gradually becoming a more vibrant area in world politics as well, since socialization would seem to be central to some of the major topics in international relations theory today: preference formation and change;<sup>1</sup> national identity formation; the creation and diffusion of, and compliance with, international norms; the effects of international institutions, among other topics.

It is curious, though, how undertheorized socialization has been in much of IR, despite the fact that most noncoercive diplomatic influence attempts by most actors most of the time are aimed at “changing the minds” of others, at persuading, cajoling, or shaming them to accept, and hopefully internalize, new facts, figures, arguments, norms, and causal understandings about particular issues. That is, the goal of diplomacy is often the socialization of others to accept in an axiomatic way certain novel understandings about world politics.<sup>2</sup> Especially in the second half

<sup>1</sup> This is particularly relevant when trying to explain how new states, “novices,” decide on the content and institutional structure of their foreign policies, not an unimportant topic when looking at the effects of decolonization or the collapse of the Soviet empire.

<sup>2</sup> As Nadelmann remarks in the context of prohibition regimes, “The compulsion to convert others to one’s own beliefs and to remake the world in one’s own image has long played an important role in international politics—witness the proselytizing efforts of states on behalf of religious faiths or secular faiths such as communism, fascism, capitalism, and democracy (1990:481).

of the Clinton administration, for example, the engagement of China was seen as a way of teaching Beijing about allegedly predominant norms and rules of international relations (free trade; nonuse of force in the resolution of disputes; nonproliferation; multilateralism, etc.). The engagers spoke of bringing China into the “international community” (defined normatively), an enculturation discourse if ever there was one. So even if, in the end, many attempts to use diplomacy to effect the internalization of new ways of thinking and behaving fail, it still makes sense to try to explain why actors (state and non-state) engage in this kind of activity in the first place. But of course, we do not really know how many of these attempts do fail because we have not really tried to define, isolate, and measure the effects of socialization processes in IR.

This is not to say that predominant IR theories ignore the concept of socialization entirely. Classical realism seems torn between its impulse to essentialize the drive for power in a self-help world on the one hand and its sensitivity to historical contingency on the other. Morgenthau, for example, does not rule out the possibility that actors internalize group norms of behavior such that action takes on a “taken-for-grantedness.” Indeed, he laments the disappearance of a time in European interstate relations when individual kings and absolute rulers heeded certain norms of behavior for fear of the social punishments from violation (e.g., shame, shunning, loss of prestige and status) (Morgenthau 1978:251–52). He even leaves open the possibility that definitions of power and interest are culturally contingent, implying at least that there is variation in how actors are socialized to conceptualize legitimate ways of pursuing legitimate interests. But if so, it is not clear how actors are socialized into or out of perceptions of the world as competition for power and influence in an anarchical system. In other words, by accepting the cultural contingency of power and interests, logic would suggest that Morgenthau would have to accept that the *realpolitik* impulses that characterize world politics are in fact not given, but learned. Yet for classical realists there is no obvious theory of socialization to explain radical variations in interpretations in the meaning of power and interest.

This is true as well for so-called neoclassical realism (Rose 1997). Rejecting the structural realist critique of reductionism, this scholarship has (re)discovered that subjective and intersubjective interpretations of power and interest matter in explaining the behavior of states and thus international outcomes. Yet it also persists in arguing that there are unchanging universal facts about international life that constrain state behavior, namely that international relations are a *realpolitik* struggle among self-interested, security-seeking, relative power-sensitive states operating in anarchy. I am not clear how you can have it both ways: once you allow for independent causal importance of subjective or intersubjective inter-

pretations of the external world, you open the door to the possibility that there can be vast disjunctures between estimates of this world and the “real” world of material power distribution and *realpolitik* pursuits of interest. That is, you open the door to the possibility that subjective and intersubjective interpretations of the world can change even as the realities of material power distribution remain constant. If this possibility can exist, then, in principle, the real world has a less independent, predictable effect on actor behavior. As such, the “realities” of anarchy and relative material power imbalances are no longer so determinative. If so, these realities are not likely to be important independent sources of actor preferences or beliefs about the external world. This is clearly not where the neoclassical realists want to end up. Moreover, this conclusion then begs the question of where these preferences and beliefs come from. Neoclassical realism has no answer, or at least none that flows logically from realist theor(ies). Thus, it has no theory of socialization.

Neorealism uses socialization to describe the homogenization of self-help balancing behavior among security-seeking states interacting under conditions of anarchy (Waltz 1979:127–28). But the use of the term is problematic. First, the process of homogenization is not really socialization in commonsense usage. While Waltz uses an example from crowd psychology to argue that interaction in groups can create a “collective mind” across individual members (1979:75), his discussion of interaction in IR essentially drops the collective mind image and replaces it with a “selection and competition” image. It is emulation and selection that leads to similarities in behavior of actors through interaction: states that do not emulate the self-help balancing behavior of the most successful actors in the system will be selected out of the system such that those remaining (assuming there are no new entrants into the system) will tend to share *realpolitik* behavioral traits.<sup>3</sup> It is unclear as to whether the theory assumes states will also share epiphenomenal *realpolitik* foreign policy ideologies, because the theory is unclear as to whether states are conscious agents pursuing balancing outcomes or simply unconscious participants in the creation of unintended systemic balances. That is, it is not clear whether social interaction in anarchy leads to emulation or mimicking.<sup>4</sup>

<sup>3</sup> For a sophisticated discussion of the neorealist concept of emulation, see Resende-Santos 1996. For an acknowledgment of structural realism’s tendency to describe competition, rather than socialization, see Thies 2001:2.

<sup>4</sup> I differentiate between emulation and mimicking in the following way: emulation involves the conscious, careful search for exemplars and success stories, a dissection of the reasons for their success, and the application of these lessons to the maximization of some specific expected utility. Thus, it involves internalizing, as well, the causal models of the way the world works that exemplars themselves use to maximize their utility. Thus, to emulate a successful balancer in an anarchical environment means also to share its *realpolitik* “worldview,” its cause-effect understandings. Mimicking involves copying what most other

In any event, it is simply not empirically obvious that this kind of selection even occurs. It is hard to pick exemplars in world politics due to the uncertainty about what constitutes success under the security dilemma. It seems odd to claim that uncertainty about relative power drives states to look for successful balancers, but that apparent uncertainty does not make it difficult to identify who in fact are the appropriate exemplars out there.<sup>5</sup> What lessons should a state draw from the collapse of the Soviet Union? That deterrence and containment work against threatening or rising power? Or that transnational arms control coalitions successfully socialized a group of influentials in the Soviet Union to adopt cooperative security strategies under the rubric of “new thinking,” despite US military pressure? Both the United States and the Soviet Union balanced against each other. One failed, one succeeded. How, then, do state actors decide whether or not balancing is a successful strategy in IR worthy of emulation? As Dan Reiter has argued, historical experience in alliances, rather than some search for obvious transhistorical exemplars, is often the criteria states use when deciding when and what type of balancing is appropriate (Reiter 1996).

Neorealism, then, exaggerates the structural pressures toward homogenization. Often different states do not sit in competition with each other over scarce resources; rather, some find “niches” where the requirements for survival are different, hence, their different traits can survive side by side without some selection pressures toward homogenization.<sup>6</sup>

For another, the death rates of states have declined dramatically in the twentieth century. Unsuccessful actors—those that eschew self-help, that fail to balance internally or externally—tend not to disappear anymore (Fazal 2001). New states have emerged in the latter half of the twentieth century in an era when failed or unsuccessful states are not routinely eliminated. These new states presumably retain heterogeneous traits and characteristics, supported in some respects by institutions and rules analogous to those that support socially weak and failed individuals in many domestic societies. That is, it is not obvious that the “fitness” of states has in-

---

actors in a social environment do in the absence of a conscious, calculated search for any one exemplar utility maximizer. Rather copying is done by a novice seeking a relatively efficient way of surviving in an uncertain and new environment prior to a sophisticated search for information about the most successful, exemplar utility maximizer.

<sup>5</sup> Farkas makes a similar point (1998:34).

<sup>6</sup> This is an important feature of evolution that neorealism’s rendition misses. As a somewhat useful comparison here, Darwin’s finches showed that two species can develop sufficiently different survival needs (e.g., the food for especially small finches is not substitutable for the food for especially large ones) and that these needs do not leave them in competition over similar resources. Thus, there is less competition over who is fittest, and hence one species does not tend to eliminate or crowd out the other. See Weiner 1994.

creased over time, given a constant anarchical environment—at least not fitness defined in terms of an ability to balance successfully. Somehow the international system has allowed “unfit” states to develop a range of strategies for surviving without self-help balancing—norms against aggression, arms control agreements, a concept of sovereignty that “equalizes” unequal actors, among others.<sup>7</sup> This being the case, the characteristics of the system structure must, by definition, be much more varied and complex than the simple tending-toward-balances anarchy of a neorealist world. Thus, the social environment in which these new states are socialized must be not only one that rewards or selects states that copy “successful” self-help balancers, but one that may also reward or support “deviant” heterodox behavior. If so, then so much for the homogenizing effects of social interaction—socialization—in anarchy (see Kocs 1994).

Second, most uses of socialization refer to a process of preference formation and/or change. Child socialization involves a child developing tastes, likes, dislikes—social and material—through social interaction first with the family and then broader social groups. Political socialization usually refers to the acquisition by young people of political orientations and preferences from parents or peers. For neorealism, however, socialization appears to have little to do with preferences and interests. Perhaps this stems from the microeconomic language and analogizing that Waltz uses—economics generally models preferences as stable, while different environments (institutions, price, supply, productivity, etc.) constrain the ability of actors to achieve preferred outcomes. In any event, for neorealism, material structure (what passes for a social environment for neorealism) is the key constraint on state behavior. Socialization simply results in a greater awareness by actors of the costs of pursuing preferences that neglect the structural imperative of balancing. That is, socialization means that states acquire a greater sensitivity to signals emanating from the material structure about who succeeds and who fails and why. The interpretation of this information should be relatively unproblematic for rational unitary actors—successful states balance, unsuccessful states do not. But the process by which an actor comes to read these signals correctly does not involve change in the nature of the actor—its identity or preferences or understandings of the nature of the international system. So it is hard to see why it should be called socialization.<sup>8</sup>

<sup>7</sup> History matters here. Many of these norms and practices that protect the survival of “unfit” states evolved in the twentieth century out of movements for self-determination and the diffusion of the principle of sovereign equality into the postcolonial world. My thinking here has been informed by Brenner’s helpful discussion of the distinction between evolutionary algorithms and learning processes in explaining social evolution (1998).

<sup>8</sup> As Wendt puts it, Waltz’s use of the term “socialization” is surprising, given there is no social content to neorealism’s concept of structure—it is the product of material power factors, *not* the product of the nonmaterial traits and characteristics (identities and prefer-

Finally, for neorealism socialization can go in only one direction—toward the convergence of behavior around realpolitik norms of behavior. This rules out the possibility of system-level socialization in non-realpolitik directions. Yet there are sufficient and substantively interesting deviations from neorealist claims—cases of norm-conforming behavior in the absence of obvious material threats or promises—to suggest that there are domestic and systemic normative structures that socialize actors (Kier 1997; Finnemore 1996b; Price and Tannenwald 1996; Price 1998). Indeed, one could legitimately question whether material structure plus anarchy does any socializing at all, given the empirical frequency of non-balancing behavior (Reiter 1996; Schweller 1994; Schroeder 1994; Johnston 1998b). Moreover, whatever realpolitik socializing that does go on is, arguably, not dependent on structural anarchy, but on prior realpolitik norms, the sources of which may reside at both the system and unit levels (Johnston 1996a, 1998a).

Contractual institutionalism generally does not focus on socialization processes per se in IR. For many contractual institutionalists, true to their microeconomic and game theoretic styles of analysis, the notion that social interaction can change preferences and interests or fundamental security philosophies and ideologies is not a central concern. Modeling is usually done assuming these things are fixed. Social interaction inside institutions is assumed to have little or no effect on the identities or interests of actors (or institutionalists are divided as to whether there are any effects).<sup>9</sup> That is, actors generally emerge from interaction inside institutions with the same attributes, traits, and characteristics with which they entered. These characteristics have no effect on the attributes, traits, or characteristics of the institution itself—an efficient institution in principle should reflect the nature of the cooperation problem, not the nature of the actors themselves—and these characteristics, in turn, have no impact

---

ences) of the interacting units. Thus, socialization becomes convergence of behavior around balancing and relative gains concerns, not the convergence of actor attributes (e.g., identities, preferences, and interests) (Wendt 1999). Although socialization in common usage does include the convergence of behavior around socially preferred models or exemplars, the process usually involves some degree of internalization of these exemplars such that they become normatively taken for granted, and thus elicit pro-social behavior in the absence of material constraints (rewards and punishments).

<sup>9</sup> I am grateful to Celeste Wallander for pointing out to me some of the divisions over institutions and preferences in the contractualist camp. Wallander allows for variation in interests but argues that institutions do not cause this variation (see Wallander 1999). Other contractualists claim to the contrary that interests can be changed through involvement in institutions, mainly via complex learning. Explicating this learning process ought to be high on the institutionalist research agenda (see Keohane 1984:132). But it not clear what the causal mechanisms would be, nor whether the process would be endogenous to the institution itself or a function of shifting domestic coalitions.

on actor identities. Iteration, the intensity of interaction, the provision of new information about the beliefs of other actors, and so on, do not seem to have any effect on the basic preferences of actors. Being enmeshed in an iterated but potentially finite PD does not make the D,C payoff less desirable, in principle. Whether social interaction is short run or long term, it has no effect on underlying preferences. All it does is change the costs and benefits of pursuing these preferences. The quality or quantity of prior social interaction among players should be irrelevant to the calculus of whether or not to defect (Frank 1988:143).

The undersocialized nature of institutions in contractualist arguments is highlighted by the motivations contractualists *do* focus on when explaining pro-group behavior. Cooperation is elicited in institutions in basically three major ways.

One is issue linkage. Take, for example, a suasion game, where one player has a dominant cooperation (C) strategy, leaving the other player to defect (D). The Nash equilibrium (C,D) is one that leaves the player with a dominant C strategy somewhat dissatisfied, while giving the player playing D its best payoffs. The dissatisfied player therefore has an incentive to use threats or promises (e.g., tactical issue linkage) to move the outcome to a more advantageous set of payoffs (Martin 1993). Persuasion here is nothing more and nothing less than an effort to change the cost-benefit calculations of the defecting player with exogenous positive or negative incentives so as to secure cooperation. Persuasion does not change that player's underlying desire to defect in a suasion game, nor does it change basic beliefs—or common knowledge—about what kind of game is being played.

A second way that institutions can elicit norm-conforming behavior is by providing reasons for actors to worry about their reputations. A prior reputation as a cooperator brought to a stag hunt game, for example, can reassure others that the actor genuinely prefers a C,C outcome. This can stabilize the Nash equilibrium. Thus, it is in the interest of actors with common interests to first acquire a cooperative reputation, particularly from situations in which cooperation can be quite costly. The desire to establish a trustworthy reputation thus can be an incentive to engage in norm-conforming, pro-social behavior (Kreps 1992).

The reputation argument has at least one major problem. As Frank (1988) points out, this kind of reputation should never be a reliable or credible one to a rational observer. Being able to observe reputation-building behavior means that such behavior is probably undertaken with the likelihood that it will be observed. Indeed, there is no point engaging in it for reputational purposes unless it is observable to others. But if a behavior or an action is designed to be observed, and both the observer and actor know this, then the observer should have doubts that it is indeed

high-cost behavior. The only way to reduce these doubts is for the actor to behave in such a way that an observer is convinced the behavior cannot be calculated. Thus, the paradox: the most credible reputation is one that is based on behavior that is automatic, emotional, and uncontrolled, not calculated (Frank 1988). This implies that in order to minimize in-group defection, distrust, and conflict, groups have an interest in instilling in their members deeply rooted, credible, “taken-for-granted” responses to social interaction. So, in the end, normative socialization becomes the basis for credible reputations.

A third feature of institutions that helps elicit cooperative behavior—perhaps the most important one for contractualists—is information. Interaction in institutions can provide new information that can affect beliefs about causality, about means-ends relationships, and/or about the preferences of others (Martin 1999:84). This information can reduce uncertainty about the credibility of others’ commitments, and thus help actors’ expectations converge around some cooperative outcome. Thus, information only affects beliefs about the strategic environment in which the actor is pursuing fixed preferences. It does not appear to feed back to a reassessment of the desirability of these preferences in the first place. The usual way that information affects preferences is through its effect on elite change (assuming the actor is an aggregate political entity such as a state). Information about the failure of some strategy, for instance, could lead to a loss of support for one set of elites pursuing their definitions of interests and their replacement by another set with different definitions of interests. There is a sort of infinite regress problem that much of the work on information runs into, however. That is, what makes this information about failure conclusive, unless there is prior agreement on the criteria for success and failure? What leads to prior agreement on these criteria? Information about the validity of these criteria that all actors find credible? What leads to this kind of agreement on the credibility of the criteria about credibility? Information about the credibility of the credibility of the credibility of these criteria, and so on? At any stage one could simply state, unproblematically, that actors received credible information about a phenomenon and leave it at that for the purposes of modeling interaction from that point on. But this does not escape the problem that at any given point, the criteria for establishing the credibility of new information are problematic.

Thus, there is something vaguely mystical about how contractualists treat information inside institutions. Information is rarely contested; it has an obviousness about it that unproblematically reduces uncertainty. The meaning of new information often seems to need no interpretation. Why do institutions reduce uncertainty? Because of new information about preferences, beliefs, and strategies. What makes such information

credible? Usually, contractualists point to the costliness to the provider of the information (costliness in terms of some loss of material welfare or power). But contractual institutionalists have no theoretical advantage here, no theory of the conditions under which new information will influence preferences, beliefs, or strategies and by how much. Moreover, the social context of information appears to be irrelevant, or at least the origins of common definitions of costliness, essential for information to be credible, are unexamined. The Bayesian updating of beliefs, as a practical empirical research matter, says nothing about how to determine how a myriad historical social relationships that constitute one subset of beliefs filters this new information.<sup>10</sup>

Yet 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). Economic transactions, for instance, bargaining over price where people exchange information relating to their preferences and their bottom line, vary dramatically depending on whether or not the parties are friends—friends offer higher payments and accept lower prices than strangers (Halpern 1997:835–68). Face-to-face bargaining is more likely to lead to mutually beneficial exchanges even under conditions of asymmetric information (e.g., where the seller, for instance, has private information about the value of the product) because it elicits norms of honesty that increase trust and information sharing (Valley, Moag, and Bazerman 1998).<sup>11</sup> Even in prisoners’ dilemma (PD) relationships, information about the other as an opportunist is not static. Hayward Alker reports on lab experiments involving two individuals playing an iterated PD: the two could not communicate directly with each other, but transcripts were made of their comments about

<sup>10</sup> I do not include here a review of the literature on signaling games and information in IR. I will discuss this in more detail in the context of persuasion. Suffice it to say, the more interesting work on how information is interpreted in the context of signaling games between players with asymmetrical private information overlaps considerably with the constructivist interest in the social context of information, since the credibility of signals in conveying information depends in many instances on prior social relationships and the identities and interests formed from these. Nonetheless, much of the signaling literature focuses on how the context of information alters actors’ beliefs about the credibility of information (mostly about the credibility of threats and promises in IR or about the policy attitudes of politicians in American politics), not the basic identities or interests of actors per se.

<sup>11</sup> As Valley, Moag, and Bazerman put it, face-to-face negotiations “nearly always involve a significant proportion of the early bargaining time taken up in getting to know more about one another . . . setting a positive tone for the entire negotiation” (Valley, Moag, and Bazerman 1998:230). That is, social interaction helps actors come to conclusions about the identity of the other, and hence how information from the other should be sent and interpreted.

what they thought the other was trying to do in their game. Judging from the transcripts, at the start of the interaction both assumed the other was defecting because it was in his nature, it was dispositional, and there were initially a lot of mutual efforts to punish. Once the interaction became more or less locked into a string of mutual cooperation, the occasional defection was dismissed as situational, or chalked up to random misperceptions (Alker 1996). That is, a string of mutual cooperative moves changed the actors' assessments of who exactly they were playing, such that noncooperative behavior was no longer seen as dispositional but was instead situational. In other words, descriptively similar information about a move, defection, was interpreted differently before and after actors entered into a social relationship. Social identity theorists have also noted that individuals playing a PD are more likely to cooperate than individuals representing groups. Group affiliation—the tendency for in-group identification to require extreme demonstrations of devaluing the out-group—becomes a critical contextual variable for understanding variation in cooperative behavior (Dawes, van der Kragt, and Orbell 1988). Thus, social context is an important variable in how well information reduces uncertainty in a transaction, and in which direction this uncertainty is reduced (e.g., clarifying the other as a friend or an adversary).

For contractualists, then, information rarely changes basic preferences or interests, and there is, as far as I know, no effort to see how it changes common knowledge such that actors jointly reevaluate the nature of the game being played.<sup>12</sup> Yet the significance of the iterated PD games that Alker reports on lies precisely in this point: a string of cooperative interactions led the players to take the same information—a defection by the other—and reinterpret it as a mistake or as a consequence of the situation rather than of the other player's disposition. In effect, the game was redefined, shifting from a PD where both players believed they and the other held a preference for the D,C outcome to an assurance game where both players believed they and the other held a preference for C,C.

To be fair, contractualist arguments do not a priori reject the possibility that information changes preferences instead of strategic environments. The advice is to test for both, but in practice the tendency is to discount the possibility of the former. There appear to be two reasons for this. First, modeling with fixed preferences is easier for contractualists influenced by choice-constraint, or choice-theoretic (game-theoretic) epistemology and

<sup>12</sup> Yet, it is plausible to argue that new causal information can lead to the creation of new interests: scientific information about the process of global warming, for example, has created an interest in stemming increases in greenhouse gas emissions where such an interest never existed before. On contractualism's neglect of explanations for common knowledge, see Risse 1997:1.

methodology.<sup>13</sup> Second, preferences are difficult to observe, as are changes in them. What often may appear to be a change in preferences is, instead, a change in strategies. Any likely source one might turn to to observe preferences (e.g., from statements through to actions) could well be itself a product of strategic interaction, hence unrepresentative of true preferences (e.g., a deceptive move, a second best action, a document that itself is a product of bargaining among actors in an interagency process). Thus, it is safer and easier to either assume preferences or to deduce them from some prior theoretical assumption about the nature of the actor.<sup>14</sup>

This seems to be a reasonable, cautionary argument for a sound methodological choice. It does reveal, however, an implicit disciplining move that constrains efforts to think about changing preferences through new information acquired via social interaction inside an institution. There are a number of possible responses. First, theoretically deduced or assumed preferences are not so easily juxtaposed with observed preferences. Theories do not appear *deus ex machina*. They are almost invariably based on some initial inductive observation about a phenomenon. It is often not logically obvious what the preferences of actors ought to be from observing their position in society or their organizational constitution as actors. Thus, for instance, it is unlikely there is an obvious *a priori* theory of the preferences of religious groups in the United States about humanitarian intervention without some information about the political ideologies of these groups. Nor can a useful theory about the preferences of militaries (e.g., that they organizationally favor offensive doctrines and capabilities) be easily derived from some deductive assumptions about the universal characteristics of militaries (as Kier [1997] shows). Theoretical assumptions about preferences can only come from moving back and forth between initial empirical observations and theoretical hunches. But for many choice-theoretic approaches, empirical observation of preferences is ruled out as too risky and the measurement error is considered too high. But from empirical observation it is clear too that the theoretical deduction of preferences is itself fraught with risks and uncertainties. It should not be *a priori* privileged as a “better” way of figuring out what actor preferences are. Why not pit hypotheses based on theoretically deduced preferences against inductively analyzed preferences to see which set predicts behavior that is more consistent with empirical evidence?

<sup>13</sup> As Keohane implied in his seminal work, *After Hegemony*, institutionalists use a structural-functional analysis of the constraints institutions place on actors, and it is to these exogenously given constraints that actors respond in ways influenced by their subjective or cognitive characteristics (1984:132).

<sup>14</sup> See Frieden 1999 for a sophisticated statement of these arguments.

Second, the bias against observing preferences seems to exaggerate the difficulty of observing them. To be sure, validity and reliability of measures for accessing the preferences of actors are problematic, since the only way to observe is to look at some phenomenon external to their cognition, e.g., a speech act, a gesture, a decision.<sup>15</sup> There is always the possibility that these external manifestations are in some sense strategic, and not direct representations of preferences. But given the theoretical importance of the question—whether preferences change through social interaction, and how stable the preferences are—it seems premature to give up trying to observe change. As Herrmann notes, human cognitions are the “input variables that make a human problem-solving paradigm go. They are essential whether the paradigm assumes rational maximizing behavior or not. It is surprising how little effort we have made to figure out ways to infer them” (Herrmann 1988:180). The response should be, how do we reduce the measurement error to the lowest possible level and maximize the internal validity and reliability of indicators of preferences and desires?

All of this is not to say that contractualists rule out the possibility of preference change. Some recognize that institutional life can change what people desire. As Shepsle and Bonchek note, “While debate and deliberation may often seem like window dressing, it is entirely possible that, from time to time, some persuasion, reconsideration, conceivably even coercion takes place, the result of which is that someone changes preferences” (Shepsle and Bonchek 1997:44). Contractual institutionalists just do not think it happens that often, or that it involves messy empirical work showing that it has happened. Game theorists, therefore, admit that they do not know how to model change in the likelihood function in Bayesian updating.<sup>16</sup>

Therefore, when behavior changes, persuasion is generally not the first answer for contractual institutionalists. When preference change *is* detected, the first impulse is to ask a liberal/pluralist question: how has the institution affected the distribution of power resources across domestic actors leading to new interests reflecting new political coalitions? Martin rightly notes that institutional theory no longer needs to show that institutions matter: that case has been made. Rather it needs to show how they matter: “[We] require a finer-grained understanding of the mechanisms through which institutions might exert their effects” (Martin 1999:86). But socialization is, apparently, not one of the mechanisms that interests contractual institutionalists.

<sup>15</sup> Developments in the theory and technology of brain imaging may change this, however.

<sup>16</sup> I thank Ken Shepsle for clarifying game theorists’ arguments about likelihood functions.

This disinterest in socialization is somewhat surprising, though. First of all, given the prominence of coordination games and focal points in institutionalist theorizing about social norms, habits, customs, and conventions that constrain rationally optimizing behavior, one might expect more curiosity about the social and historical origins of focal points. Institutionalists are ready to admit that focal points can be products of shared culture and experience.<sup>17</sup> Martin notes that bargaining inside institutions may allow states to establish focal points in a coordination game (Martin 1993:101). But the origins of these focal points in IR are not of central concern, and institutionalists do not explicate the microprocesses by which bargaining reveals or creates (or convinces actors to accept) a focal point inside an institution.<sup>18</sup> In essence, shared culture and social experience may help actors choose among equilibria, but institutionalist theory per se does not endogenize this choice. This is acceptable if one assumes stability in focal points and conventions. But this is an assumption, not an empirical claim. Moreover, a different ontology leads to a different assumption about focal point stability, and hence to greater interest in explaining the social origins of focal points. That is, a constructivist or complex adaptive systems ontology holds that continuous interaction between multiple agents over time leads to changing social structural contexts (emergent properties) that, in turn, affect how agents define their interests. In macro historical terms, this means that social conventions and focal points can evolve and change rather dramatically, nonlinearly, and in path-dependent ways. Indeed, this ontological difference between contractual institutionalists and constructivists is a major reason why the former are often not analytically interested in how or if preferences change. A constructivist ontology opens the door to relatively unstable preferences, giving more weight to preferences as determinative of behavior (and to normative social contexts), thus requiring researchers to pay special attention to observing rather than deducing preference change.

Second, contractualists, at least those who claim neorealist pedigrees, would have to admit that there are retardants in neorealist socialization

<sup>17</sup> The “conspicuousness” or “prominence” of some equilibrium outcome in a coordination game that turns it into a focal point can be a function of socialization in a shared “culture” (Morrow 1994:96).

<sup>18</sup> In more sociological terms, bargaining can entail a process of empathetic discovery of shared notions of prominence and conspicuousness about particular outcomes. Or it can lead actors to “discovering” shared focal points, in effect by creating feelings of group solidarity. In some cases, this could lead to discovering that the group shares a priori interests. More interesting, however, is the process by which this discovery is in fact the *creation* of new interests that actors believe are shared because they believe members of an in-group *should* share interests. In other words, the social and psychological rewards of membership precede the “discovery” of shared interests.

processes; hence, they would have to admit that there is at least a crude neorealist socialization process at work to begin with. The logic is as follows: given that some contractualists accept neorealist assumptions of anarchy and security maximizing, in order to explain why actors might develop absolute gains concerns and recognize common interests that might necessitate setting up institutions, they would have to admit the possibility that neorealist socialization can be imperfect, incomplete, obstructed. Yet, instead of asking whether incomplete socialization might be attributable to counter-socialization processes inside institutions, they would rather look to domestic political change, exogenous technological change, or the emergence of multiple issues areas to explain absolute gains concerns and common interests.

Finally, and most interesting perhaps, contractualists use a definition of institutions that seems to be very sociological: the intent and effect of an institution—if it operates efficiently—seems almost to be to socialize states. Martin, for example (following Mearsheimer, following Keohane), defines an institution as “a set of rules that stipulate ways in which states *should* cooperate and compete with one another. . . . [Institutions] prescribe *acceptable* forms of state behavior, and proscribe *unacceptable* kinds of behavior” (Martin 1999, emphasis added). This language implies that the most effective institutions are those that get actors to internalize or at least take for granted that which is prescribed and proscribed. States would over time drop calculations according to consequences, and replace them with invocations of appropriateness. Otherwise, so-called prescribed and proscribed behavior would simply be behavior that was rewarding or unrewarding rather than what was acceptable or unacceptable.<sup>19</sup>

Recent liberal theorizing in IR also seems to dance around the concept of socialization without incorporating it clearly into its ontology or epistemology. On the one hand, a leading liberal theorist,<sup>20</sup> Andrew Moravcsik, seems open to the possibility that social interaction changes interests and

<sup>19</sup> Abbott and Snidal (1998) try to meld rationalist and constructivist perspectives on the role and effect of formal international organizations, and touch on the possibility that IOs can help states “change their mutually constituted environment and, thus, themselves” (25). But the bulk of the article focuses on the uses of IOs by states or on the roles IO play in mediation, enforcement, and the provision of information. Despite the promising claim at the start to look at IOs as agents which “influence the interests, intersubjective understandings, and environments of states” (9), in the end the issue is essentially dropped.

<sup>20</sup> There is a more mainstream materialist branch of liberal theorizing about how domestic preferences interact with domestic political institutions to create national-level responses to external economic and political change. But scholars in this tradition—exemplified by IPE scholars such as Ron Rogowski and Jeffrey Frieden, among others—start from the assumption of fixed material preferences, and thus have nothing to say about socialization.

preferences through identity construction: he critiques constructivists, and rightly so, for neglecting agency at the state and substate levels when looking at the diffusion of norms. In particular, he takes them to task for neglecting how the formation and distribution of preferences at the domestic level mediates the diffusion of norms from the system structure (1997:539). This suggests that he accepts the possibility that this diffusion occurs, that it can lead to internalization (e.g., socialization), but that variation in internalization across states will be a function of prior state and substate identities. On this point he is, I think, essentially right.

Elsewhere he uses the language of sociology in a way that seems to keep the door open for theorizing about socialization. “Liberal theory seeks to generalize about the social conditions under which the behavior of self-interested actors converges toward cooperation or conflict” (1997:517). One social condition is the degree of convergence or divergence over values and beliefs. But it is not clear where this convergence comes from. He remarks that “for liberals, the form, substance and depth of cooperation depends directly on the nature of these patterns of preferences” (1997:521), that is, on the distribution of preferences created by social interaction among multiple actors. It is not clear, however, how particular distributions of preferences feed back to affect the degree of convergence or divergence over values that is predictive of cooperation or conflict. He then suggests that preferences may indeed vary in “response to a changing transnational social context,” and as an example, “the position of particular values in a transnational cultural discourse may help define their meaning in each society” (1997:522–23). This suggests another feedback loop, linking the degree of convergence or divergence of values to changes in preferences. Yet, he also appears agnostic as to whether social preferences or social identities inside states are socially constructed at the state or interstate level: “Liberals take no distinctive position on the origins of social identities, which may result from historical accretion or be constructed through conscious collective or state action, nor on the question of whether they ultimately reflect ideational or material factors” (1997:525). In the end, it seems that all liberal theory adds to the issue of socialization in IR is an awareness that the diffusion of norms will be filtered by domestic institutions that aggregate preferences and produce policy outcomes. The explanation for the social origins, content, and construction of these preferences is left outside of the theory. That is, liberal theory is agnostic about where preferences come from (Moravcsik 1997:525).

Needless to say, for social constructivists, socialization is a central concept. As Onuf puts it, “social relations *make* or *construct* people—*ourselves*—into the kinds of beings we are” (Onuf 1998:59). In their accounts of the creation and diffusion of international norms, constructivists

mostly focus on the “logics of appropriateness”—pro-norm behavior that is so deeply internalized as to be unquestioned, automatic, and taken for granted. This naturally raises questions about which norms are internalized by agents, how, and to what degree. Kratochwil and Ruggie imply that by treating institutions as social institutions “around which actor expectations converge,” the interesting question becomes the processes by which this intersubjective convergence takes place. The term “social institution” certainly implies that the degree of convergence before and after entering into an institution should be different for reasons that primarily have to do with interaction among agents inside the institution (Kratochwil and Ruggie 1986). So some process of socialization must be going on. Yet the empirical work in this regard has tended to follow the sociological institutionalists’ focus on macro historical diffusion of values and practices (such as rationalism, bureaucracy, and market economics), measured by correlations between the presence of a global norm and the presence of corresponding local practices (Eyre and Suchman 1996; Price 1998; Finnemore 1996b). Finnemore has gone beyond correlation to causation by focusing on how international agents (e.g., international organizations and ideas entrepreneurs) have actually gone about “teaching” values and constructing domestic institutions and procedures inside states that reflect emergent international norms and practices. States then adhere to these norms and practices even when these seem inconsistent with their material welfare or security interests.

The problems with constructivist approaches to socialization, however, are fairly basic. First, inheriting much of the epistemology of sociological institutionalism, constructivism has tended to leave the microprocesses of socialization underexplained.<sup>21</sup> It tends to assume that agents at the systemic level have relatively unobstructed access to states and substate actors from which to diffuse new normative understandings. That is, once actors are interacting inside institutions, the diffusion and homogenization of values in the “world polity” seem virtually automatic, even, and predictable. This leaves variation in the degree of socialization across units—the degree of contestation, normative “retardation,” and so on—unexplained. And it leaves the causal processes unexplicated.<sup>22</sup> Even Fin-

<sup>21</sup> For a similar critique, see Checkel 1998:335 and Risse 1997:2. As Wendt puts it: “In social (and IR) theory . . . it is thought to be enough to point to the existence of cultural norms and corresponding behavior without showing how norms get inside actors’ heads to motivate actions” (Wendt 1999:134).

<sup>22</sup> See for example, Meyer et al. 1997 and Haas 1998:26. Haas posits that “interpersonal persuasion, communication, exchange and reflection”—socialization—occurs in thick institutional environments where epistemic communities are active, but there is no discussion of microprocesses of persuasion nor conditions under which variation in the effectiveness of persuasion—hence the completeness of socialization—might be observed. Nadelmann identifies normative persuasion as a central process by which prohibition regimes emerge—for

nemore's detailed story of teaching often stops at the point where agents at the international level deliver norm-based lessons to rather passive students. There is less attention paid to the processes by which units or unit-level actors understand, process, interpret, resist, and/or act upon these "lessons." That is, it is unclear how exactly pro-normative behavior is elicited once the models of "appropriate behavior" are displayed or communicated to agents at the unit level. This neglect is surprising,<sup>23</sup> 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.<sup>24</sup> The result is, however, that causal processes by which systemic normative structures (constitutive, regulatory, and prescriptive) affect behavior are mostly assumed, rather than shown.<sup>25</sup>

Put visually, suppose for the moment one could conceive of the relationship between structures and agents as is done in figure 1.1.<sup>26</sup> At its simplest, social (normative) structures require institutional environments where the institution itself or actors within it try to transmit to new members the predominant norms of the structure. The agent processes these norms, filtered by prior features of identity, and they then mediate the development of foreign policies and practices that govern interaction with other agents and with the institution. The interaction of these policies and practices reproduces, while also perhaps modifying, the original social structures and institutional environments. Much of the constructivist empirical work focuses on the first step in this process and skips over the second, drawing correlations between the norms being taught by agent-

---

instance, antislavery norms in British diplomacy—but it is unclear why political leaders and government officials were persuaded by moral arguments (1990:494). Keck and Sikkink (1998) go a long way in looking at the microprocesses by which transnational activist networks "persuade," but international institutions as social environments per se are not the focus of their research. Adler (1998:133) also notes that the OSCE has an explicit mission to socialize members by trying to persuade them that they are, or ought to be, like "us"—liberal, cooperative, and sharing a European identity. But it is not clear why this persuasion ought to work on initial members who are somewhat illiberal and noncooperative. Finnemore and Sikkink (1998) address this general problem in their survey of constructivist work on norm diffusion, however.

<sup>23</sup> But this neglect is beginning to change with new studies on socialization in European institutions (see note 53), and on how normative localization or hybridization occurs when local agents with prior identities respond to norm diffusion (Acharya 2004).

<sup>24</sup> For similar critiques see Checkel 1998:332 and Moravcsik 1997:539.

<sup>25</sup> This may change as scholars pick up on Finnemore and Sikkink's summary of some plausible causal processes (1998).

<sup>26</sup> The figure describes the interaction with only one agent. The relationships described here would apply to all other agents involved in the interaction as well.

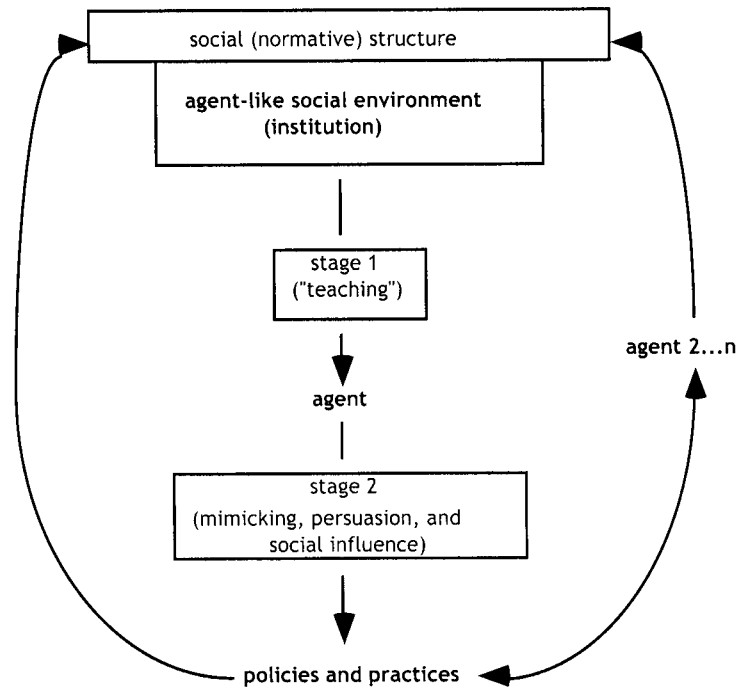


Figure 1.1. Structures and agents in a notional socialization process

like institutions on the one hand and the content of state practices on the other. Yet it is precisely the second stage that determines how effectively socialization takes place, if at all, and therefore how one might find variation in interests and behavior across actors who are nonetheless exposed to the same institutional social environment. In short, it is at this second stage where one needs to look in order to understand if, and how, socialization microprocesses are at work in world politics.

Second, when constructivists do begin to look at these microprocesses of socialization and the constitutive effects of social interaction, the focus is almost exclusively on persuasion. By usage, it 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 they tend to borrow in some form or another from Habermas's theory of communicative action. The argument is that social interaction is not all strategic bargaining. Rather, prior to strategic bargaining, actors have to arrive at "common knowledge"; that is, they must first come to share basic assumptions about the deep structure of their interaction: who are legitimate players 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, beyond the pale. In other words, for them to even interact strategically, they need to establish focal points that are so deeply accepted as to be stable.<sup>27</sup> Thus, right from the start, bargaining is not simply a process of manipulating exogenous incentives to elicit desired behavior from the other side. Rather it involves argument and deliberation all in an effort to change the minds of others.<sup>28</sup> As Hasenclever, Mayer, and Rittberger put it, “[T]he parties to a conflict enter a discourse where they try first 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” (Hasenclever, Mayer, and Rittberger 1997:176–77, emphasis mine; see also Knoke 1994:3 and James 1998:7).<sup>29</sup>

There are a couple of questions here. First, it is not obvious why, from the perspective of actually doing empirical research on socialization in IR, one should focus on Habermas to the neglect of a very rich research tradition on persuasion in communications theory, social psychology, and political socialization. It is certainly not a novel argument from the perspective of these traditions, and it is not clear how the application of the communicative theory of action would go about showing whether persuasion or coercion explained behavior that was increasingly pro-social over time. 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.<sup>30</sup> Constructivists seem to rely on an identity argument here: that is, persuasion is more likely to occur when two actors trust one another such that each accepts the “veracity of an enormous range of evidence, concepts and conclusions drawn by others” (Williams 1997:291). Put simply,

<sup>27</sup> See Johnson 1993:81 on this point.

<sup>28</sup> For an excellent exegesis of Habermas’s theory of communicative action, see Risse 1997.

<sup>29</sup> Sometimes persuasion can mean both something akin to communicative action *and* something more normatively coercive, entailing shaming or opprobrium. Here compliance with a norm need not be a function of internalization but is, rather, a function of state elites’ aversion to public criticism (Risse and Sikkink 1999:13–14; Keck and Sikkink 1998:16). This confuses persuasion with social influence, in my view.

<sup>30</sup> See the conditions explicated by James (1998:7–11, 15–17).

identification leads to affect and affect leads to a greater probability that the arguments and interpretations of the other will be accepted as valid, and internalized. How the initial level of identification necessary for communicative action comes to be is unclear.

But the more important problem is a second one. While it is understandable why constructivists would want to focus on persuasion—this is their trump card in disputes with neorealists and contractualists over whether social interaction can change actor preferences and interests in pro-social ways, and it is the purest type of socialization—at least two other effects of social interaction can lead to pro-normative behavior in the absence of exogenous material threats or promises. These are social influence and mimicking. The first is a term that encompasses a number of sub-processes—backpatting, opprobrium or shaming, social liking, status maximization, and so on—where pro-normative behavior is rewarded with social and psychological markers from a reference group with which the actor believes it shares some level of identification. The latter refers to copying pro-normative behavior as satisficing means of adapting to an uncertain environment prior to any detailed ends-means calculation of the benefits of doing so. With their almost exclusive focus on persuasion—or communicative action—constructivists have a hard time distinguishing between the range of microprocesses that help explain pro-normative behavior.

Socialization is too important to ignore in world politics. It is what actors in world politics often try to do to each other. It already appears to varying degrees and with varying importance in the main clusters of international relations theory. Treating international institutional life as an environment where social interaction, independent of material rewards and punishments, may change everything from actor preferences, to beliefs, to behavior could provide new insights into the conditions of international cooperation.

Given the fact that much human behavior is inherently overdetermined, that one act of cooperation is often the product of the simultaneous considerations of multiple interests, testing for the effects of socialization depends on careful research designs that create conditions helpful for observing the effects of socialization and for determining its substantive importance. It is to this issue that I now turn.

#### DEFINITIONS, MICROPROCESSES, AND METHODS

There is general agreement across the social sciences that socialization is a process by which social interaction leads novices to endorse “expected ways of thinking, feeling, and acting.” As Ochs puts it in relation to childhood socialization: “[T]hrough their participation in social interactions

children come to internalize and gain performance competence in . . . sociocultural defined contexts” (Ochs 1986:2). In Stryker and Statham’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” (Stryker and Statham 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.” It gives people identities; they are “assigned a specific place in the world.” Socialization, then, involves internalization: “the immediate apprehension or interpretation of an event as expressing meaning, that is, as a manifestation of another’s subjective processes, which thereby becomes subjectively meaningful to myself.” And internalization means the development of shared identification such that people come to believe “[w]e not only live in the same world, we participate in each other’s being” (Berger and Luckman 1966:129–30). Thus, socialization is aimed at creating membership in a society where the intersubjective understandings of the society become “objective facticities” that are taken for granted (Berger and Luckman 1966:44).

Political scientists have not wandered far from these basic themes in their definitions of socialization. Ichilov refers to political socialization as “the universal processes of induction into any type of regime.” These processes focus on “how citizenship orientations emerge” (Ichilov 1990:1). Sigel refers to political socialization as the “process by which people learn to adopt the norms, values, attitudes and behaviors accepted and practiced by the ongoing system” (cited in Freedman and Freedman 1981:258). Beck and Jennings essentially see political socialization as a process by which adolescents acquire political orientations from their families and from major sociopolitical forces and events in formative periods (Beck and Jennings 1991:743). IR theorists have generally simplified socialization to processes “resulting in the internalization of norms so that they assume their “taken for granted” nature” (Risse 1997:16). Ikenberry and Kupchan borrow from Sigel and define socialization as “a process of learning in which norms and ideals are transmitted by one party to another.” In IR they limit this to a Gramscian-like process whereby state elites “internalize the norms and value orientations espoused by the hegemon and, as a consequence, become socialized into the community formed by the hegemon and other nations accepting its leadership position.” This hegemonic order “comes to possess a ‘quality of oughtness’ ” (Ikenberry and Kupchan 1990:289–90).

There are a couple of common themes here: the first is that socialization is most evidently directed at, or experienced by, novices, newcomers, whether they be children, inductees into a military, immigrants, or “new” states.

The second theme is the internalization of the values, roles, and understandings held by a group that constitutes the society of which the actor becomes a member. Internalization implies, further, that these values, roles, and understandings take on a character of “taken-for-grantedness” such that they are not only hard to change, but that the benefits of behavior are calculated in abstract social terms rather than concrete consequentialist terms. Why should one do X? “Because . . . ,” or “Because X is the right thing to do,” or “Because X is consistent with who I am,” rather than “Because it will lead to Y, and Y benefits me.”

One should assume, however, that there can be degrees of internalization, given that not all actors are always exposed to exactly the same configuration of social pressures, nor do they enter into a social interaction with exactly the same prior identifications. Thus, while pro-social behavior because of its “appropriateness” may be the ideal, at the opposite end of the spectrum should be pro-social behavior because of its material consequences (positive and negative). At this point, pro-social behavior cannot be attributed to internalization or socialization in pro-social norms of the group.

But if internalization of pro-social values is the hallmark of socialization, and if the other end of the spectrum is behavior motivated by the calculation of material costs and benefits,<sup>31</sup> this leaves a vast amount of pro-social behavior produced by neither process.

This leads to a final and important point. The focus on internalization tends to lead constructivists to focus on persuasion. This is, as noted, what really distinguishes them from neorealists and contractual institutionalists: the internalization of group norms and values is largely a cognitive process of argumentation, reflection, and acceptance of the oughtness of particular norms. But beyond persuasion, the literature on socialization (outside of IR theory) identifies a range of reasons for why one might see pro-normative behavior in the absence of exogenous (dis-)incentives. Axelrod, for instance, lists identification (the degree to which an actor identifies with a group), authority (the degree to which “the norm and its sponsor are seen as legitimate”), social proof (essentially mimicking of a valued in-group’s behavior), and voluntary membership (where defection from group norms carries costs in self-esteem) as critical mechanisms for reinforcing pro-normative behavior (Axelrod 1997a:58–59). All of these depend on the acquisition of some kind of

<sup>31</sup> Putting aside, for the moment, the fact that constructivists would have to argue that at a certain level, the rational and conscious maximization of individual material benefit is due to the internalization of socialized values placed on material benefits, and is mediated by socially determined processes of rational calculation (e.g., long term versus short term; relative weight of trade-offs, etc.).

identification with or affective attachment to a group—that is, socialization. Ikenberry and Kupchan list three: exogenous shocks that lead to elite transformation in a state; exogenous material inducements that lead, over time (and somewhat mysteriously) to the internalization of norms that were once adopted for instrumental reasons; and normative persuasion or transnational learning through direct inter-elite contact (Ikenberry and Kupchan 1990:290–92). Beck and Jennings refer to three possible, somewhat overlapping, socialization processes whereby adolescents acquire the political orientations of their parents: parents provide social identities that bring with them political interests; power and affect relationships establish certain communication patterns in the family such that parents influence the political personalities of younger members; or the political traits of parents are transmitted through a process of inheritance or mimicking (Beck and Jennings 1991:744). Constructivism has tended to neglect many of these microprocesses.

Arguably, these multiple processes boil down to three: mimicking, social influence, and persuasion. A critical question, then, is when and to what degree do these separate processes help explain why actors change their behavior in pro-normative or pro-social ways? In practice, these processes are likely to be interactive. But separating out their key differences is important because this will help us examine how durable pro-social conformity is over time and what kinds of institutional designs are most conducive to this durability. Thus, broadly speaking, the speed, uniformity, and effectiveness of norm diffusion ought to depend a great deal on what kind of institutional social environment leads to what kind of socialization microprocess.

### *Mimicking*

Mimicking is a microprocess whereby a novice initially copies the behavioral norms of the group in order to navigate through an uncertain environment. It is an efficient means of adapting to uncertainty prior to any detailed ends-means calculation of the benefits of doing so. To be sure, this microprocess stretches the concept of socialization somewhat, since pro-group behavior is only indirectly an effect generated by the nature of the social environment. Rather, pro-group behavior is a function of the desire to survive in a novel social environment. That is, while mimicking is distinct from exogenously induced threats or punishments, and is not characterized by individual efforts to optimize long-run material well-being, it is not the same class of causes as persuasion and social influence. The former two are mechanisms that motivate. In mimicking, the mechanism that motivates copying can be survival under uncertainty. But mim-

icking can also stem from persuasion, or from social incentives to accept another actor as a behavioral exemplar, in which case the motivation is covered by the other two microprocesses. So mimicking could be one behavioral outcome of persuasion and social influence, a microprocess that shows how these two other microprocesses can lock actors into path-dependent behavior.

On the other hand, choosing which groups to mimic involves a degree of prior identification. Moreover, mimicking pro-social behavior can lead to internalization of norms through repetition. Or, alternatively, by mimicking an actor goes on record as behaving in some particular way. It may then be loathe to deviate from this precedent for status and image reasons. Many of the procedural constraints and work habits and standard operating procedures that actors develop in order to minimally function inside an institution come from mimicking the behavior of others in the group. These can then limit the legitimate forms of participation in the institution (Frank 1985:18; Biddle 1985:162; Ochs 1986:2–3; Cialdini, Kallgren, and Reno 1991:203–4). Mimicking can also lead to acceptance of the intersubjective norms of the group governing basic communication. In other words, mimicking can achieve basic agreement on legitimate ways to resolve conflicts even though no Habermasian communicative action has taken place.<sup>32</sup> In short, mimicking can be both a condition for and effect of more direct forms of socialization such as social influence and persuasion.

### *Social influence*

Social influence is a microprocess whereby a novice's behavior is judged by the in-group and rewarded with backpatting or status markers or punished by opprobrium and status devaluation. The appropriate reference in-group and the degree to which certain backpatting and opprobrium signals are valued depend on prior identity construction. Social rewards and punishments can elicit pro-normative/social behavior in the absence of persuasion or direct internalization. Rewards might include psychological well-being, a sense of belonging, a sense of well-being derived from

<sup>32</sup> James notes that communicative action involves agreement on the medium through which discussion of conflict takes place. This agreement—like learning a common foreign language—comes prior to all other communication (James 1998:7–11); hence, socialization, whether via persuasion or social influence, requires a basic level of communicative cooperation to begin with. I have no problem with this argument. But there is no reason why this basic agreement need be achieved through discussion and argumentation that two or more equal, empathetic actors eventually find convincing. Mimicking the language, procedures, and habits—the parameters of this medium of communication—can lock an actor seeking to survive in uncertainty into such agreement as well.

conformity with role expectation, status, and so on. 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). Conformity can be either with the descriptively normative behavior of a group (e.g., what most people in the group do) or with its prescriptively normative behavior (what most people in the group believe should be done). 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" (Booster 1995:96). Persuasion would entail public conformity with private acceptance. Persuasion, at least of the kind where the authoritativeness of the persuader is what convinces, has been called "mediated informational influence, e.g., "I thought the answer was X . . . but everybody else said Y, so it really must be Y." Social influence, instead, comes in the form of "mediated normative influence," e.g., "I believe the answer is X, but others said Y, and I don't want to rock the boat, so I'll say Y."<sup>33</sup> 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 having at least some 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, e.g., 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). Thus, one could call this a second-order socialization microprocess—while actors' preferences and interests may not change, these interests are linked together in ways in which they were not in the past. For instance, preferences in the security field become linked to preferences in social status and prestige.

### *Persuasion*

Persuasion is a microprocess whereby novices are convinced through a process of cognition that particular norms, values, and causal understandings are correct and ought to be operative in their own behavior. Persuasion has to do with cognition and the active assessment of the content of a particular message. As a microprocess of socialization, it involves

<sup>33</sup> M. Deutsch and B. Gerrard, cited in Betz, Skowronski, and Ostrom 1996:116.

changing minds, opinions, and attitudes about causality and affect in the absence of overtly material or mental coercion.<sup>34</sup> It can lead to common knowledge, or “epistemic conventions” (that 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 conversely that they in fact share an interest in cooperation. The point is, however, that the gap or distance between actors’ basic causal understandings closes as a result of successful persuasion.

Persuasion can itself result from one or all of the three microprocesses. The first pertains to the nature of the message. Persuasion involves a process of cognition where counter-attitudinal messages are compared with preexisting arguments. The latter change because of the superiority of new evidence weighed by a priori internalized truth standards. The second pertains to characteristics of the persuader, in particular the affective relationship created by the social or intellectual attractiveness of the persuader. This attractiveness heightens the persuader’s authoritativeness, and hence the persuasiveness of his/her message. The third has to do with the cognitive and social characteristics of the persuadee that mediate the evaluation of the content of the message and/or the authoritativeness of the messenger. Since persuasion involves the internalization of new attitudes, values, and norms, this type of microprocess ought to lead to the most durable and self-reinforcing pro-social behavior.<sup>35</sup>

#### RESEARCH DESIGN ISSUES: WHY INSTITUTIONS?

We need to know at least three things in order to test for the presence and effects of socialization. First, what are the characteristics of the social environment in which agents are interacting at time  $t$ ? If this environment has agent-like or “teaching” properties, what are the norms and associ-

<sup>34</sup> I just want to underscore that I am talking about changes in fairly fundamental beliefs, not relatively shallow, transient, or low-level attitudes about the efficacy of certain political choices and strategies. The difference is not always obvious, but as I will argue in chapter 4, persuasion as it pertains to socialization is interesting precisely because it involves basic reevaluations of collective “thought styles” (Farkas 1998:43) that can include preferences or strategies, as long as these strategies pertain to basic methods for achieving basic goals (e.g., multilateralism versus unilateralism as a “cause” of security). Much of the work on persuasion in political science has focused on what are, at root, rather minor changes in positions on public policy, not major changes in entire thought systems (see, for instance, Lupia and McCubbins 1998).

<sup>35</sup> Since one should expect variation in the durability of norms depending on the type of socialization microprocess, it does matter, then, whether one can observe internalization or not. Holding preferences constant for the purposes of modeling prevents one from exploring this important issue.

ated behaviors that actors in the environment are supposed to adopt and, ideally, internalize? In other words, what is the predominant ideology of the social environment? Second, what are the characteristics of individual agents involved in the social environment at time  $t$ ? How do these characteristics retard or propel the socialization process? Third, how do these agents then interact with this environment at time  $t + 1$ ? What are the policy processes through which newly socialized agents act upon the broader social environment?

The net effect of socialization, therefore, will be a function of the characteristics of the environment interacting with the characteristics of the agent in an ongoing tight feedback relationship, mediated by a foreign policy process. In IR one way of testing for socialization, then, is to use international institutions on the one hand and individuals and small groups involved in state policy processes on the other as, respectively, the social environment and individual agents of interest—the units of analysis, if you will. My reasoning is as follows.<sup>36</sup>

For the most part, when IR specialists or sociological institutionalists look for the effects of socialization, the unit of analysis has tended to be the state (or state elites in a fairly aggregated way) (Eyre and Suchman 1996; Meyer et al. 1997; Finnemore 1996a, 1996b; Waltz 1979). This presents problems when examining particular institutions as social environments since states as unitary actors do not participate in institutions; rather, state agents do, e.g., diplomats, decision makers, analysts, policy specialists, non-governmental agents of state principals, and so on. Moreover, treating the unitary state as actor presents problems when applying the most well-developed literature on socialization found typically in social psychology, sociology, communications theory, and even in political socialization theory. Most of this literature examines the effects of socialization on individuals or small groups.

A constructivist ontology allows (even demands) that the unit of socialization be the individual or small group. As Cederman points out, constructivism's 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 (Cederman 1997; Axelrod 1997b, 1997c; Hamman 1998). Thus, it matters how individual agents or small groups are socialized because their impacts on larger emer-

<sup>36</sup> I use a fairly loose definition of international institutions. Institutions are more or less formal organizations with identifiable names and with more or less obvious criteria for membership or participation. This allows one to differentiate among specific social environments with specific ideologies or normative “messages.”

gent properties of the social environment can be quite dramatic.<sup>37</sup> This focus on individuals and small groups also enables constructivists 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 1999:33).

There are good reasons, then, for studies of socialization to focus on the socialization of individuals, small groups, and, in turn, the effects of these agents on the foreign policy processes of states.<sup>38</sup>

But if these are appropriate units of analysis, why choose international institutions as the environments of socialization? After all, state actors experience a myriad of socializing environments from bilateral interactions at the state level, to intra-bureaucratic environments at the policy level, to training and work environments inside bureaucratic organizations themselves. Let me try to make the case. One of the critical claims constructivists make is that “anarchy is what states make of it” (Wendt 1992). In other words, material power structures do not determine state interests or practices, and thus *realpolitik* practice by unitary rational actors is not an immutable “fact” of international politics. In order to make this case, constructivists or their fellow travelers have, for the most part, underscored the empirical “deviations” from realist or material power interests theories: altruistic foreign aid (Lumsdaine 1993); weapons taboos (Price and Tannenwald 1996); “autistic” force postures in developing states (Eyre and Suchman 1996); “autistic” military doctrines (Kier 1997); and limits on the conduct of war (Legro 1995; Finnemore 1996c). These have been important cases that have gone far in undermining the mainstream realist edifice. But at some point the critique needs to go beyond so-called deviant cases to look at cases and phenomena that materialist realist theories claim they can explain; that is, constructivism is going to have to make the argument that *realpolitik* practice is a reflection of *realpolitik* ideology and norms.<sup>39</sup> To explore this claim against

<sup>37</sup> This is, after all, the point of much of the work on how transnational networks affect state behavior (Keck and Sikkink 1998; Evangelista 1999; Adler 1998), “teaching” and the diffusion of norms, and the creation of national interests (Finnemore 1996b). The roots of this complex adaptive systems approach, as it relates to normative structures in IR, go back to Durkheim’s work on the creation and re-creation of “social facts” through the interaction of individual normative agents. See Ruggie 1998:29.

<sup>38</sup> Ruggie calls this a focus on “innovative micro-practices,” a hallmark of constructivist research (Ruggie 1998:27).

<sup>39</sup> I define *realpolitik* ideology, or strategic culture, fairly specifically to mean a worldview where the external environment is considered to be highly conflictual, where conflicts with other actors tend toward zero-sum, and where, given these conditions, the use of military force is likely to be quite efficacious in the resolution of conflicts. Vasquez calls this a power politics paradigm. I do not define *realpolitik* simply as the “prudent” pursuit of the power

realism's structural materialist argument requires setting up a critical test where both approaches spin out alternative but *competitive* propositions, predictions, expectations, and so on, to see which additional set of empirical observations is confirmed or disconfirmed.<sup>40</sup>

But there is an additional empirical implication that could provide an important test of constructivist versus material realist accounts of realpolitik: that is, if constructivists are right, realpolitik ideology and practice ought to be changeable—independent of material power distributions and “anarchy”—when actors are exposed to or socialized in counter-realpolitik ideologies. If materialist realist theories are right, realpolitik discourse is epiphenomenal to realpolitik practice, and neither should change in the presence of counter-realpolitik ideology.<sup>41</sup>

This is where international institutions come in. Constructivists suggest that international institutions in particular are often agents of counter-realpolitik socialization. They posit a link between the presence of partic-

---

interests of nation-states, as some realists do. This, it seems to me, is too vague and thus its presence or absence is empirically hard to falsify. Moreover, it does not really describe how realist theorists themselves believe decision makers socialized in anarchy understand this environment.

<sup>40</sup> Elsewhere I expressed some doubts about the wisdom of this kind of critical test for traditionally materialist realism versus ideational explanations (see Johnston 1996a). The reason was that materialist and ideational ontologies could be seen as incommensurate, that one had to subsume the other, that I believed the ideationally based realpolitik came prior to realpolitik pathologies emanating from structural anarchy, and hence that a critical test between theories derived separately from each made no sense. Desch picked up on this and suggested that I eschewed critical testing, period (Desch 1998:161). My earlier conclusion rested on empirical evidence that many critics did not find convincing, and evidence that I also indicated did not establish an especially strong test of my argument. (I did not reject the value of critical tests among competing ideationally based theories, but that is a minor debating point.) So in this book, the research design is premised on what I believe is a stronger critical test of materialist and ideational explanations for realpolitik pathologies in IR. This is necessary in order to establish whether materialist or ideational ontologies stand alone as an adequate basis for exploring realpolitik ideology and practice. For a very promising research program that shows how some realpolitik values—specifically a concern about relative gains—vary in the real world, see Rousseau 2002.

<sup>41</sup> Note that I do not accept that an actor's sensitivity to changes in relative power confirms material realism. I have argued elsewhere that one could argue that this sensitivity is ideationally rooted: Rousseau (2002) shows this empirically. Indeed, this is the whole point of testing for socialization. Similarly, when I conclude that cooperation occurs despite relative power concerns, this does not mean that I believe “relative power concerns” is a phenomenon exclusive to, hence confirming of, material realist arguments, or that socialization arguments necessarily expect cooperative behavior and a rejection of realpolitik pathologies. Socialization can go in both directions—actors can be socialized into or out of realpolitik practices. But to deal with the important charge that realpolitik ideology and practice are both epiphenomena of material structures, the critical test necessarily involves looking for evidence of non-realpolitik socialization.

ular normative structures embodied in institutions and the incorporation of these norms in behavior by the actor/agent at the unit level. It is in institutions where the interaction of activists, so-called norms entrepreneurs, is most likely, and where social conformity pressures are most concentrated. It is in international institutions where multilateral diplomacy with its emphasis on interpersonal communication, debating, and argumentation is manifest (Muldoon 1998:3). Institutions often have corporate identities, traits, missions, normative cores, and official discourses at odds with realpolitik axioms,<sup>42</sup> indeed at odds with the socialization pressures that neorealists argue come with being sovereign, insecure actors operating in anarchy. For example, some arms control institutions expose actors to an ideology where, inter alia, multilateral transparency is normatively better than unilateral non-transparency; disarming is better than arming as a basis of security; common security is better than unilateral security; and evidence of the potential for cooperative, joint gains in security in the international system is greater than evidence that the environment is a fixed, conflictual one. All of these axioms and assumptions challenge the core assumptions of realpolitik ideology. So, if there is any counter-realpolitik socialization going on, it ought to be happening in particular kinds of security institutions.<sup>43</sup> I do not mean to imply that institutions are the only fora in which socialization in IR goes on. Since the focus is on microprocesses, obviously state agents and principals in the policy process are exposed to a wide variety of socialization experiences and interactions inside their own states. I am simply interested in how broader non-realpolitik norms in international security might be diffused. Institutions are an obvious place to look.<sup>44</sup>

Note, however, that treating institutions as social environments means positing that different social environments vary in terms of their persuasiveness and social influence. This means asking how institutions as social environments vary in ways conducive to socialization. We need, then, a

<sup>42</sup> For a discussion of organizations and their goals, see Ness and Brechin 1988:247, 263–66. See also Muller’s discussion of the ideology of the non-proliferation regime and how the causal and principled ideas of the regime relate to its norms and proscriptive regulations (Muller 1993); Barnett and Adler on the role of international institutions in the construction of security communities (Barnett and Adler 1998:418–21); Alter’s discussion of the legitimacy of the European Court of Justice’s legal culture and doctrine and how this constrains states from challenging the ECJ even when its rulings run against state preferences (Alter 1998:134–35); and Schimmelfennig’s description of the goals of European institutions (Schimmelfennig 2002:7).

<sup>43</sup> Risse makes a similar point, suggesting that communicative action should be more frequent inside institutions than outside of them (Risse 1997:17).

<sup>44</sup> Or as Shambaugh put it: “The more provocative question is whether an actor’s preference, interests and identity can be altered initially as a result of its association with an international institution and vice versa” (Shambaugh 1997:8).

typology of institutional forms or institutional social environments. Unfortunately, we do not have one. One could imagine, though, at least several dimensions for coding institutions as social environments. Here I am borrowing and expanding on the typology of domestic institutions developed by Rogowski (1999):

1. membership: e.g., small and exclusive or large and inclusive
2. franchise: e.g., where the authoritativeness of members is equally allocated, or unevenly (though legitimately) allocated
3. decision rules: e.g., unanimity, consensus, majority, supermajority
4. mandate: e.g., 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 on these five dimensions) would thus create different kinds of social environments, leading to differences in the likelihood and degree of persuasion and social influence. For instance, to take one extreme ideal type (as I explain in chapter 3), persuasion is likely to be an especially prevalent and powerful socialization process when membership is small (social liking and in-group identity effects on the persuasiveness of the counter-attitudinal message are strongest); when the institutional franchise recognizes the special authoritativeness of a couple of actors (the authoritativeness of the messenger is likely to be high); when decision rules are based on consensus (this requires deliberation which, in turn, can trigger more flexible cognitive evaluation of new information); 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 so there is less pressure to represent the principal); and when autonomy of agents is high, e.g., when the issue is narrow or technical or when the principal just does not care much (when the principal is less attentive or relevant).<sup>45</sup> 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, ch. 5).

<sup>45</sup> Here Risse and I are, I think, moving along parallel tracks. He notes, for instance, that nonhierarchical and network-like international institutions “characterized by a high density of mostly informal interactions should allow for discursive and argumentative processes” (Risse 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” (Martin and Simmons 1998:742). See also MacLaren 1980 for similar hypotheses.

Conversely, given the microprocesses of social influence I outlined earlier, backpatting and opprobrium are more likely to be at work when membership is large (this maximizes the accumulation of backpatting/shaming markers); when the franchise is equally allocated (there are no obvious “authoritative” sources of information); when decision rules are majoritarian (an actor’s behavior is on record, and therefore consistency effects may be stronger); 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 mimicking, social influence, or persuasion had led to pro-social/pro-normative behavior in international institutions? First, as I noted earlier, one would have to show that social environments in institutions are conducive to mimicking, social influence, or persuasion. Second, one would have to show that after exposure to or involvement in a new social environment, attitudes or arguments for participation have indeed changed, converging with the normative/causal arguments that predominate in a particular social environment, or that they reflected social influence pressures emanating from that environment. Third, one would have to show that behavior had changed in ways consistent with these arguments. Finally, one would have to show that material side payments or threats were not present, or at least were not part of the decision to conform to pro-social norms. Broadly speaking, these are the tests this study sets out to conduct.

#### RESEARCH DESIGN ISSUES: WHY CHINA?

Given the research design requirements for analyzing the effects of socialization on cooperation, a strong case can be made for looking at China. Precisely because counter-realpolitik institutions may be critical environments for counter-realpolitik socialization, an easy case can be made for studying arms control institutions and China. In many instances, these institutions embody a non-realpolitik, even anti-realpolitik, ideology centered on the notion of common security (though, admittedly, in uneasy tensions with sovereignty-centric axioms as well). China is at one and the same time a “novice” and a hard-realpolitik state. This status is ideal for testing for socialization since this is precisely the kind of state where the effects of socialization (if there are any) should be easiest to observe given the potential contrast between a China that has not participated and a China that has participated in these institutions.<sup>46</sup> And it is the kind of

<sup>46</sup> I do not use socialization here in a normative sense. Nor do I believe that China prior to entry into these institutions was in some sense unsocialized. It was merely differently socialized.

state where resistances to this kind of socialization should be greatest, given the prevalence of hard-realpolitik worldviews. Together this means that China is a “least likely” case for socialization arguments, but one where the effects of exposure to international institutions should be relatively easy to observe.

Noviceness is, unfortunately, undertheorized in IR. If socialization is 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 newly liberated or created states, or recently isolated states. This suggests where to look for “most likely cases” for the purposes of theory testing: newly decolonized states from the 1950s on; 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 equivalent to “primacy effects,” where early experience and information will have out-of-proportionate effects on inferences drawn from later experiences and information.<sup>47</sup> New states literally have had to set up foreign policy institutions, determine what their foreign policy interests are on a range of novel issue areas, decide in which of a myriad social environments in IR they should participate (e.g., which institutions, which communities of states—middle power, major power, developed, developing), and which competitive and cooperative relations to foster.<sup>48</sup> It is precisely these kinds of states where one ought to expect socialization effects from involvement in international institutions to be greatest. China is not exactly a novice in the same way as 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 US and other developed states.

A few simple statistics will suffice in showing the pace of China’s integration into international institutional life. Figure 1.2, for instance, shows China’s shift, particularly after Mao’s death in 1976, from being a “novice” in international institutional life to being a participant at levels nearing those of most major developing and developed states.<sup>49</sup>

<sup>47</sup> See Choi (1993:52–53).

<sup>48</sup> See, for instance, the argument in Chafetz, Abramson, and Grillot (1996). The authors argue that the debate in Ukraine in the 1990s over whether to sign the 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 de-nuclearize and join the NPT).

<sup>49</sup> This figure includes all “conventional” international bodies, identified as IO types A–D in the Union of International Associations’ *Yearbook of International Organizations*.

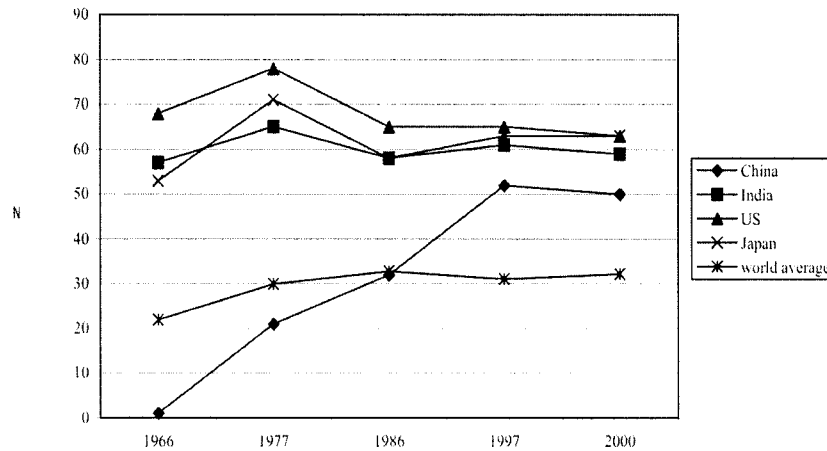


Figure 1.2. Comparative involvement of international governmental organizations. Source: Union of International Associations (2000/2001).

For another view of this change in Chinese participation rates, figure 1.3 uses level of development as a predictor of membership in international organizations for all states in the international system. The assumption here is that more resource-constrained states with fewer linkages to the global economy should be less involved in political institutions as well. High levels of development are associated with high levels of interdependence, hence with a high demand for institutions that can regulate these interactions. Thus, GDP/capita can act as a proxy indicator for a demand for institutions.<sup>50</sup> The figure shows that over the 1990s, China became increasingly *overinvolved* in international organizations given its level of development. Prior to the 1990s, China's participation rates fell below the regression line. That is, for its level of development China was underinvolved in international organizations (IOs). Put differently, its demand for institutions was lower than it should have been, given its level of development. Beginning in the 1990s, however, China became overinvolved. In essence, in the 1990s China moved from well below to well above the regression line.

In security institutions, China's increasing participation rates over the 1980s and 1990s are equally impressive. Figure 1.4 indicates that China's

<sup>50</sup> I used GDP/capita as a predictor for the number of IGO memberships for each state for which both sets of data were available, and entered these data into an ordinary least squares (OLS) regression equation for the years listed in figure 2.4. The regression equations are as follows: 1977,  $y = .001x + 38.238$ ,  $R^2 = .21$   $p = 0.00$ ; 1985,  $y = .001x + 36.437$ ,  $R^2 = .207$   $p = 0.00$ ; 1989,  $y = .001x + 41.421$ ,  $R^2 = .228$   $p = 0.00$ ; 1997,  $y = .001x + 38.501$ ,  $R^2 = .163$   $p = 0.00$ ; 2000,  $y = .001x + 38.225$ ,  $R^2 = .172$   $p = 0.00$ .

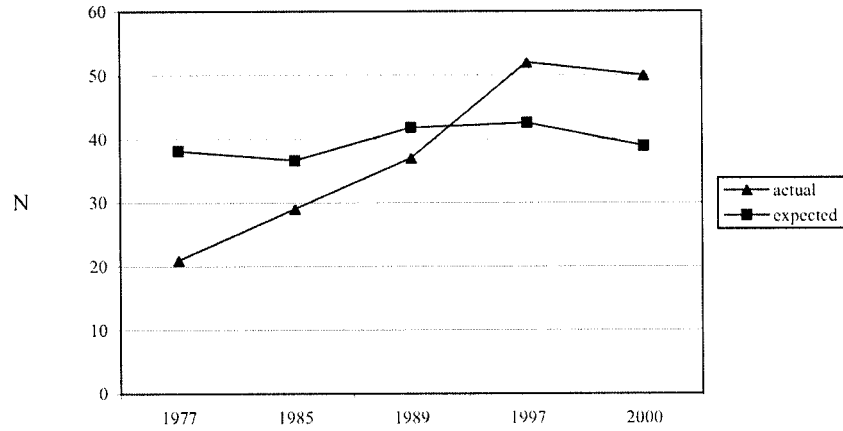


Figure. 1.3. China's actual and expected membership in international organizations, based on OLS regression of development data as a predictor of IO memberships. Sources of data: Union of International Associations (2000/2001) and World Bank.

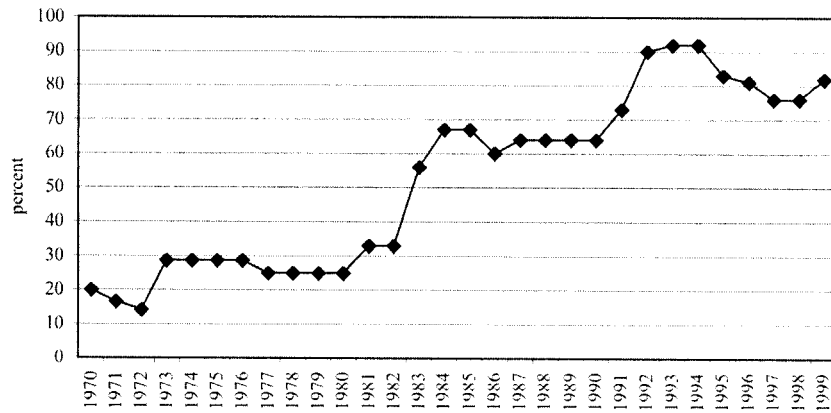


Figure 1.4. China's arms control treaty accessions as a percentage of eligible accessions.

accession to multilateral arms control agreements as a percentage of all possible agreements it was eligible to sign jumped rapidly, beginning in the early 1980s.

This has not simply been a function of an increasing number of international security institutions: the rate of increase in China's participation has been faster than the rate at which new international security institu-

TABLE 1.1  
China's Arms Control Treaty Accessions

<i>Treaty</i>	<i>Date of Signature</i>
Geneva Protocols	1952
Latin American Nuclear Weapons-Free Zone	1973
Convention on Certain Conventional Weapons (CCCW)	1981
Antarctic Treaty	1983
Outer Space Treaty	1983
Biological Weapons Convention	1984
Convention on Assistance in Case of Nuclear Accident	1986
Convention on Early Notification of Nuclear Accident	1986
South Pacific Nuclear Weapons-Free Zone	1987
Convention on the Physical Protection of Nuclear Material	1989
Seabed Treaty	1991
Non-Proliferation Treaty	1992
Chemical Weapons Convention	1993
Convention on Nuclear Safety	1994
London Convention on Nuclear Dumping	1994
CCCW Protocol II (landmines) and Protocol IV (lasers)	1996
Africa Nuclear Weapons-Free Zone	1996
Comprehensive Test Ban Treaty	1996
Southeast Asia Nuclear Weapons-Free Zone	1999*

\* Indicated it would be the first nuclear power to sign the treaty protocol.

tions have been created. From 1982 to 1996, for example, the total number of possible treaties increased from nine to eighteen, or an increase of 100 percent. China's accessions rose from three to fifteen, or a jump of 400 percent. By 1996, the only eligible treaties that China had not signed on to were the Partial Nuclear Test Ban Treaty (PTBT) and the Environmental Modification Treaty. In 1986 China publicly pledged to end atmospheric testing (which it had done in practice in the early 1980s), thus, in effect, unilaterally committing itself to the major provision of the PTBT. The Comprehensive Test Ban Treaty (CTBT), in any event, makes the PTBT irrelevant (see table 1.1).

These data suggest, then, that China's noviceness in the 1980s, and its increasing involvement in international institutions over the 1990s, created necessary conditions, at least, for socialization effects from international

social environments. If materialist realist theories are right, however, this pattern of participation should be irrelevant to the reasons for China's cooperation. There should be no socialization effects of a non-realpolitik kind on the PRC. Indeed, Chinese decision makers' realpolitik suspicions about entrapment in multilateral security commitments should not change. To the extent that materialist realist arguments can explain increases in Chinese participation in international institutions, one should find evidence that, at best, all relevant actors in China see arms control institutions as tools for balancing against US power. There should not be much internal debate on this score. All this should be especially true after the collapse of the Soviet Union in 1991 and the emergence of the United States as the sole military-political superpower (see figure 1.5).<sup>51</sup> In a (nearly) unipolar world, according to neorealists, one ought to see candidate poles such as China trying to balance against the United States, eschewing arms control commitments that might place constraints on its relative power capabilities (Layne 1993; Waltz 1997).

Thus, China is a hard case for non-realpolitik socialization and, hence, an easy case for structural realism. It is a hard case as well precisely because of the fairly deeply ingrained realpolitik worldview among China's elites, reinforced by an account of modern history where China has been a victim at the hands of militarily and economically more powerful states in a highly competitive and dangerous international system (Johnston 1996a, 1998a; Christensen 1996; Deng 1998; Callahan 2004). Any evidence for socialization effects in the China case, then, would go a long way in confirming the analytic value of socialization, and in highlighting the analytic flaws of materialist realism and its claims about the epiphenomenality of realpolitik ideology.

If contractual institutionalist arguments are right, then pro-social or cooperative Chinese behavior should be a product of one of three factors: (1) exogenous material incentives or disincentives constraining a hard-realpolitik China from pursuing its prisoners' dilemma (or worse, deadlock) prefer-

<sup>51</sup> If one uses the International Institute for Strategic Studies data, the United States lead in military power increases still more. In 1987 the United States accounted for 39 percent of world military expenditures in 1985 dollars, the Soviet Union for 17 percent, and China for 2.6 percent. In 1992, these figures stood at 44 percent, 7.1 percent, and 4 percent respectively figures are about one-third of the actual expenditures. Regardless which data are used, the United States does not constitute a unipole using Modelski's categorization of polarity (a unipole controls 50 percent of world expenditures or more, in bipolarity two states combined possess more than 50 percent of the expenditures with each having 25 percent or more, and any other distribution is a multipolar one) (Modelski 1974). But neither can one call the system multipolar. Thompson calls this kind of distribution near unipolarity (Thompson 1988). From the perspective of military power distribution, then, since the early 1990s the PRC ought to have been preoccupied with balancing against US power.

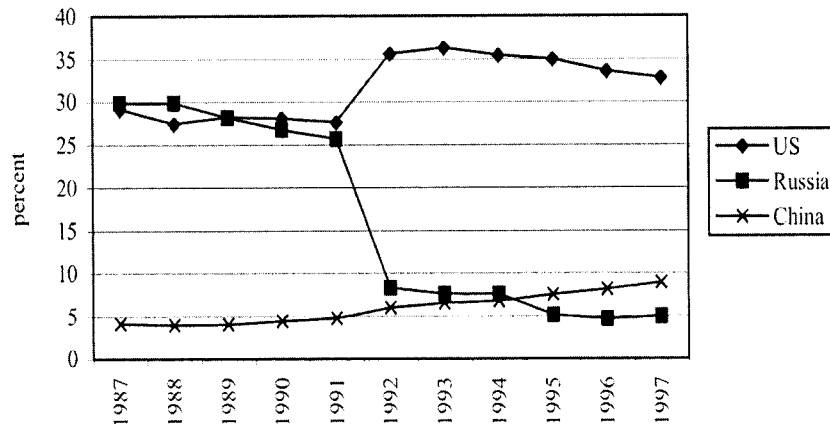


Figure 1.5. US, Russian, and Chinese shares of world military expenditures. Source: Arms Control and Disarmament Agency, *World Military Expenditures*.

ences; (2) new information that reassures a PD or “deadlock” China that it cannot be exploited or entrapped in the arms control institutions (e.g., that participation is essentially costless); (3) or changes in who makes decisions either because of policy failures or because effective involvement in the institution requires a shift in the locus of decision making.

If constructivists are right, any prosaically or cooperative behavior that emerges from China’s cooperation in the arms control institutions should be a function either of mimicking the discourse and practices of counter-realist institutions, or of the inherent desire to acquire status markers as a cooperator in a global system increasingly managed by international institutions, or of changes in a preference for multilateralist outcomes (in which case there should be a convergence over time in the security ideology that Chinese decision makers take to these institutions and that is promoted by the institutions themselves).

The China case is useful from a research design perspective for one other reason, namely, another prominent source of change in behavior toward international security institutions, domestic political change, is essentially held constant. As I noted earlier, contractual institutionalist and choice-theoretic work, when it does look for changes in elite preferences and foreign policy behavior, often looks first to changes in domestic political alignments, elite transformation, and so on. One would be hard-pressed, however, to find dramatic changes in domestic political configurations, elite political ideology, or, indeed, in the composition of the top leadership over much of the period under study in this book. From 1978 through to the mid-1990s, Deng Xiaoping was China’s preeminent leader,

the first among equals in foreign policy strategy. The Communist Party of China retained its monopoly on power throughout this period. Deng's successors, while of a very different generation, nonetheless were politically socialized inside the party and state apparatuses. One simply does not see, therefore, the kinds of changes in governments, official ideologies, or rearrangements of political power relationships among relevant domestic interest groups that can precede rapid change in foreign policy. This is not to say that domestic institutions, particularly those organizations involved in security policy processes, are irrelevant, or that these have not changed at all. Indeed they have, and their evolution is not irrelevant to explaining the evolution in Chinese arms control policy, as I will discuss in chapter 2. But this domestic institutional development does not really constitute a major change in domestic political alignments where, as contractual institutionalist and liberal theory suggest, one should normally look for major changes in foreign policy.

#### CASE SELECTION AND ANALYSIS

The preceding discussion establishes why China's involvement in international security institutions is an effective general case for studying the effects of socialization.

Together, these design issues suggest three general sets of empirical expectations. Assuming an actor enters the institution and its particular social environment with realpolitik preferences and causal and principled beliefs, and assuming the institution's ideology embodies causal and principled beliefs that are generally inconsistent with realpolitik ones, the following are three sets of plausible empirical expectations.

First, if some of China's increasing cooperation in international institutions is a function of path-dependent mimicking, we should expect to see: no change in the causal arguments behind decisions to cooperate; rather one should see arguments that reflect short-term conformity because of the novelty of a situation, or a desire to acquire more information about the institution to reduce operational uncertainty. Discourse and behavior should reflect the constraints of the linguistic and organizational procedures of the institution, with little obvious cost-benefit calculus other than a short-term desire to "learn the ropes."

Second, if social influence (in this case, status concerns and desire to maximize backpatting and minimize opprobrium) is at work, one should expect to see: commitments to participate and join that take place in the absence of material side payments or threats of sanctions; arguments for joining or participating that stress backpatting and image benefits, diffuse reputation benefits, and opprobrium costs; arguments that stress the "in-

evitability” or lack of choice in participation; and, as a first cut, initial bargaining positions, if stuck to, that would put the state in a distinct minority, and isolate it from the reference group—commitments to pro-social behavior would be made only when it was clear that noncommitment would be highly isolating.

Finally, if persuasion is at work, one should expect to see (after exposure to this environment): arguments about participation that include declining concern about detrimental effects of participation on relative capabilities and security; heightened concern about beneficial effects for global, regional, and national security; and conformist behavior later in the process that could not be expected earlier on, given the initial causal understandings of the Chinese participants. In short, you should get increasing comfort levels even as the process encroaches more on the autonomy and unilateral security options of the state. Moreover, this more fundamental support for multilateral security institutions should become more obvious over time, especially for those who have participated directly in these processes. If persuasion is not at work, one should expect to see the prevalence of free riding and relative power arguments in discussions of participation in security institutions. The key terms in these institutions’ discourses should not show up in Chinese deliberations. Indeed, one should find even harder *realpolitik* arguments in the policy process as China faced a unipolar, US-dominated international structure after 1991. Contractual institutionalism would also expect the prevalence of free riding and relative power arguments, but one should also see arguments about the value of side payments, or about short-term reputational gains/losses that are explicitly linked to other issue areas, or about information that indicates the costs of cooperation were lower than expected. These, then, are the general propositions that I test in this book.

The book centers on how mimicking, social influence, and persuasion effects work their way through the policy process to produce policy change. Choosing the main empirical cases, however, requires some care. I focus mainly on five institutions. In most of these cases, the level of China’s ultimate cooperation was unexpected by most observers at the time, cooperation was not necessarily in China’s relative power interests, and/or there were no obvious material incentives (such as side payments) to encourage participation. For the analysis of mimicking, I focus mainly on the organizational and ideational effects of China’s participation in the Conference on Disarmament in Geneva from 1980 on. For the analysis of social influence, I look at the Comprehensive Test Ban Treaty (1994–1996) and the Convention on Certain Conventional Weapons landmines protocol negotiations (1995–1996). In this regard I also look at the Ottawa Treaty banning anti-personnel landmines (1996–1997), a treaty that China did not join despite overwhelming support for the treaty from other

states. I look at the details of this case as a guard against selecting on the dependent variable. As for persuasion, I examine China's participation in the ASEAN Regional Forum (1994–2000) and related regional security dialogues. In addition, I develop a number of other empirical observations one might expect to see if these socialization processes are at work.

Let me explain these choices a bit more.

The universe of arms control institutions and treaties in which China could participate or has participated is not very large. Moreover, most of China's accession to or participation in arms control institutions and treaties can be classified as very low cost. For example, the Seabed Treaty and Outer Space Treaty—banning nuclear weapons deployments in the sea or in outer space—were essentially costless for the PRC since there is no evidence of any intention to deploy or actual deployment of nuclear weapons in these places. China's decision to join the NPT appears to have been affected to a large degree by the opportunities accession opened up for access to US nuclear power technology and for the export of Chinese nuclear-related technology. These sorts of cases of cooperation, while important for the substantive effect on levels of dealing with the problems at hand, are not surprising, and they are uninteresting for testing socialization arguments.<sup>52</sup>

There are, however, a small number of cases where China has essentially agreed to cooperate even though there are potential relative power costs to doing so, in the absence of obvious material side payments or sanctions. These are the interesting cases. China's participation in the CTBT and the Landmine Protocols of 1996, and in the ASEAN Regional Forum—the only formal multilateral security institution in East Asia—are these kinds of cases.

At the end of each chapter, I also look at additional empirical implications and cases, so as not to select wholly on the original dependent variable. Not all of these are confirming cases; in some cases the change in Chinese behavior is not in the cooperative direction. A socialization argument would hypothesize that in these cases, the socialization effects are weaker, due either to stronger resistance (more powerful relative power effects) or to weaker persuasion and social influence conditions. These short cases are worth looking at to see if socialization effects are indeed more constrained than they were in the CTBT, landmines, and ARF cases. Thus, the research design is eclectic and varies to some degree across the three main chapters.

<sup>52</sup> For a good discussion of the more or less standard explanations for China's cooperation in institutions of all kinds (territorial security, domestic stability, information gathering, management of relations with great powers, economic development, and prestige), see Lan-taigne 2005.

In sum, I think a strong case can be made for looking at change in China's involvement in international security institutions over the 1980s and 1990s in order to probe the effects of socialization. China's prior realpolitik strategic culture, its noviceness, its rapid move into institutions that embody anti-realpolitik normative goals, and the ability to control or test for the effects of materialist and domestic political independent variables together create conditions for about as good a test of socialization microprocesses as one can hope for in the real world.

The one condition that is not so ideal is, of course, access to the policy process. Since much of the evidence for the effects of socialization microprocesses necessarily comes from arguments for or against pro-social behavior, the more the access to the details of policy making, the better. The Chinese cases make this difficult, and I am terribly envious of my colleagues who have looked at socialization microprocesses in European institutions. By necessity, then, the data for the case studies come from an eclectic mix of sources. Some are open-source analyses appearing in specialists' articles in journals or in papers written for a range of non-governmental and international conferences and fora. Some are open-sourced documents and information circulated in these institutions by Chinese actors. Some are internal circulation analyses and documents, not technically secret but nonetheless on average likely to reflect more authoritative views and arguments than official government statements. Just as important, however, I have relied on over 120 interviews with arms control specialists from China, the United States, Canada, and Singapore, most of whom have been involved in the policy processes or interagency discussions of their respective countries. I am obviously constrained in accessing the policy process, especially in a system that has developed a term, "asymmetric transparency" (*bu dui cheng tou ming du*) specifically to justify its lack of openness on security questions. The IR subfield has tended to slight or undervalue interviews. There are a number of reasons, but probably one of the key ones is a distrust that agents are willing or able to accurately report on their intentions behind an action. Such reporting may be deliberately deceptive, or exaggerated, or overly modest due to the personality or cognitive abilities of the interviewee. Often, instead, the researcher's preference is, in the face of this interpretive uncertainty, to deduce intentions from prior theoretical assumptions about the organizational affiliation of the actor, or about his or her material interest. As I noted earlier, this is problematic on empirical grounds. And it biases the search for the effects of socialization on interests, desires, preferences, and intentions right from the start. Yes, intentions and/or notions of appropriateness are difficult to observe. But if interviewing is done carefully with attention paid to where the interviewee fits into the decision process, with follow-ups, with careful wording of questions, with sensitivity to the

interpersonal dynamics between interviewer and interviewee, and with triangulation interviews with others, one can reduce some of the measurement error that inheres in using face-to-face self-reporting of intentions behind actions.

## CONCLUSION

This chapter has tried to establish the theoretical importance of testing for the effects of socialization microprocesses in world politics. Socialization infuses, sometimes explicitly, sometimes implicitly, much of the debate in international relations theory about the origins of, and changes in, actor interests in IR. It also infuses, sometimes explicitly, sometimes implicitly, policy debates in a number of states, including the United States, over how to deal with rising powers, “rogue” states, and other potentially “revisionist” actors in IR. Yet there has been a great deal of confusion or neglect in both spheres about how precisely socialization is supposed to work.

Various versions of realism hold that states are socialized by their exposure to an international anarchical environment, but only in one direction—toward realpolitik definitions of interest and practices. Those that maladapt are likely to suffer a reduction in security, or worst of all, elimination as a state. This socialization process ensures that, over time, there is a convergence in the system around realpolitik behavioral pathologies. Most realisms, however, rely more on a selection argument than a socialization argument, and they exaggerate the homogeneity of the socialization process.

Contractual institutionalism—while eschewing the term *socialization*—nonetheless holds out the possibility that extended social interaction in institutions (e.g., iterated PD games) provides new information that can change “beliefs” about the interests, intentions, and capabilities of other actors. This is, however, a socialization argument without using socialization language, and there is no reason why this new information, as Alker points out, cannot also lead to redefinitions of identity and interest (Alker 1996). Nor is there any a priori reason for contractual institutionalism to downplay social rewards and punishments as, potentially, just as important as exogenously provided material side payments and sanctions in eliciting cooperation.

Constructivists, of course, focus on socialization. It is the central dynamic process for constructivist theorizing. Socialization is the process through which identities are constituted through social interaction. With some notable exceptions,<sup>53</sup> however, much constructivist empirical work

<sup>53</sup> See the special issue on socialization in European institutions in *International Organization* 59:4 (September 2005).

has tended to focus more on correlations between the norms promoted by norms entrepreneurs (individuals or institutions) at the international level and pro-normative behavior by states as actors at the unit level. This work brackets the microprocesses of socialization, and thus downplays variations in the effectiveness of norms entrepreneurs, variations in the degree of unit-level resistance, and variations in the kinds of behavioral responses at the unit level.

This chapter, then, has gone to some lengths to establish why the individual and small group makes sense as a unit of analysis in testing for socialization. The chapter has also established the case for a new look at international institutions as critical social environments in which socialization effects are likely to be observable. And I have argued why China—a state whose leaders have traditionally been hostile to relative power-constraining institutions—provides some useful hard or least likely cases with which to examine how institutional social environments might socialize the foreign policy agents of the state. On, then, to the details.