



WILEY

Treating International Institutions as Social Environments

Author(s): Alastair Iain Johnston

Source: *International Studies Quarterly*, Vol. 45, No. 4 (Dec., 2001), pp. 487-515

Published by: Wiley on behalf of The International Studies Association

Stable URL: <http://www.jstor.org/stable/3096058>

Accessed: 22-05-2016 06:52 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://about.jstor.org/terms>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Oxford University Press, The International Studies Association, Wiley are collaborating with JSTOR to digitize, preserve and extend access to *International Studies Quarterly*

Treating International Institutions as Social Environments

ALASTAIR IAIN JOHNSTON

Harvard University

Socialization theory is a neglected source of explanations for cooperation in international relations. Neorealism treats socialization (or selection, more properly) as a process by which autistic non-balancers are weeded out of the anarchical international system. Contractual institutionalists ignore or downplay the possibilities of socialization in international institutions in part because of the difficulties in observing changes in interests and preferences. For constructivists socialization is a central concept. But to date it has been undertheorized, or more precisely, the microprocesses of socialization have been generally left unexamined. This article focuses on two basic microprocesses in socialization theory—persuasion and social influence—and develops propositions about the social conditions under which one might expect to observe cooperation in institutions. Socialization theories pose questions for both the structural-functional foundations of contractual institutionalist hypotheses about institutional design and cooperation, and notions of optimal group size for collective action.

It is fair to say that for most international relations theorists there are two main ways in which involvement in international institutions changes state behavior in more cooperative directions. The first is through material rewards and punishments: in pursuit of a (mostly) constant set of interests or preferences a state responds to positive and negative sanctions provided exogenously by the institution (rules, membership requirements, etc.) or by certain actors within the institution. The second is through changes in the domestic distributions of power among social groups pursuing (mostly) a constant set of interests or preferences such that different distributions lead to different aggregated state preferences.

Few would deny that these are plausible, observable, and probably quite frequent ways in which policies change direction after a state enters an international institution. But constructivists would expand this list, and ask if and how involvement in international institutions changes state behavior in the absence of these two conditions, and in the presence of conditions that are unique to social groups *qua* social groups, namely, socialization processes. How would one know if socialization processes were critical in producing cooperative behavior? Why it is important for IR theory to figure out an answer to this question?

Author's note: I am grateful for comments and criticisms from Marc Busch, Jeff Checkel, Martha Finnemore, Peter Katzenstein, Jeff Legro, Lisa Martin, Celeste Wallander, participants in Harvard's Ethics and International Relations Seminar, the research seminar at Stanford's Center for International Security and Cooperation, the Security Studies seminar at the Massachusetts Institute of Technology, the International Relations colloquium at the University of Virginia, and several anonymous reviewers.

This article starts from a very simple (and unoriginal) premise: actors who enter into a social interaction rarely emerge the same. For mainstream international relations theories this is both an uncontroversial statement and a rather radical one. It is uncontroversial because mainstream IR accepts that social interaction can change behavior through the imposition of exogenous constraints created by this interaction. Thus, for instance, neorealists claim that the imperatives of maximizing security in an anarchical environment tend to compel most states most of the time to balance against rising power. Contractual institutionalists also accept that social interaction inside institutions can change behavior (strategies) in cooperative directions by altering cost-benefit analyses as different institutional rules act on fixed preferences.

It is a radical statement for IR theory if one claims that the behavior of actors changes because of endogenous change in the normative characteristics and identities of the actors. Put differently, change in the behavior of the participants in a social interaction may have little to do with exogenous constraints on the individual and the group and a lot to do with socialization (Wendt, 1994:384). This is, essentially, the claim made by those involved in the "sociological turn" in IR theory. The implications for IR theory should not be underestimated: the claim focuses attention both on how cooperative norms are created under anarchy (something contractual institutionalism has little to say about) *and*, logically, therefore, why conflictual norms are not epiphenomena of anarchy (something that realist theories cannot fathom). Understanding how socialization works is central to testing the claim, put forward so succinctly by Wendt, that "anarchy is what states make of it." But constructivists, as Checkel (1998) has pointed out, have not been very successful in explaining the microprocesses about how precisely actors are exposed to, receive, process, and then act upon the normative arguments that predominate in particular social environments, such as international institutions.

Given its potential to provide new insights into the production of cooperative and conflictual norms in IR, it is important that those who work with the concept are crystal clear in explicating the microprocesses of socialization and then in systematically testing for their effects. For this reason, this article focuses on three interrelated themes: why socialization approaches may offer different insights into the conditions for cooperation in IR; why, in order to offer such insights, socialization approaches have to be much more precise about at least two different social microprocesses at the heart of conformity to norms; and why, empirically, explaining cooperation inside international institutions is an important test for the validity of socialization approaches. To these ends the article begins with a review of the status of socialization in IR theory, paying particular attention to the differences between contractual institutionalist and sociological approaches to explaining cooperation in institutions. It then defines socialization and disaggregates its microprocesses. Finally it offers some theoretical and methodological reasons why international institutions are useful places to look for evidence of how precisely socialization works.

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 (Schiefelin 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 ought to be a 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: the formation and change of preferences;¹ national identity formation; the creation, diffusion of, and compliance with international norms; and the effects of international institutions, for example.

It is curious, though, how undertheorized socialization is 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, of persuading, cajoling, or shaming them to accept, and hopefully internalize, new facts, figures, arguments, norms, and causal understandings about particular issues. The goal of diplomacy is often the socialization of others to accept in an axiomatic way novel understandings about world politics.²

Yet predominant IR theories either ignore the possibility of socialization or are unprepared or unwilling to theorize about it. Classical realism (and neoclassical 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, left 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 (Morgenthau, 1978:9). But by accepting the contingency of power and interests Morgenthau would logically have had to accept that the *realpolitik* impulses that characterize world politics are in fact not given, but learned, and that there can be, potentially, vast disjuncture between actors’ estimates of this world and the “real” world of material power distribution and *realpolitik* pursuits of interest. If this disjuncture can exist, then, in principle, the “real world” has less independent, predictable effects on actor behavior. As such the “realities” of anarchy, relative material power imbalances, and so forth are no longer so determinative. Yet for classical (and neoclassical) realism there is no obvious theory of socialization to explain variations across time and space in interpreting the meaning of power and interest.

Neorealism uses socialization to describe the homogenization of self-help balancing behavior among security-seeking states interacting under conditions of anarchy (Waltz, 1979:127–128). However, the neorealist process of homogenization is not really socialization in common-sense usage. Rather it is a process of selection and competition: 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) will tend to share *realpolitik* behavioral traits.³

Yet it is not obvious that this kind of selection even occurs. The death rates of states have declined dramatically in the twentieth century. Unsuccessful actors—those that eschew self-help and that do not balance internally or externally—simply do not disappear anymore. 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 retain heterogeneous traits and characteristics, supported by institutions and rules (e.g., norms against aggression, arms

¹ 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.

² 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).

³ For a sophisticated discussion of the neorealist concept of emulation see Resende-Santos, 1996.

control agreements, a concept of sovereignty that “equalizes” unequal actors, among others) somewhat analogous to those that support socially weak and “failed” individuals in many domestic societies.⁴ The characteristics of the system structure are thus more varied and complex than the simple tending-toward-balances anarchy of a neorealist world. In such an environment it should not be surprising if “socialization” leads to less homogenization in state characteristics and behavior than neorealism expects.

Contractual institutionalism generally does not focus on socialization processes in international relations per se. The notion that social interaction can change preferences and interests or fundamental security philosophies and ideologies is not a central concern. This is odd. 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, for instance.⁵ Instead, modeling usually assumes that preferences are fixed for any particular actor. Social interaction inside institutions is assumed to have little or no effect on the “identities” or “interests” of actors, or at least institutionalists are divided as to whether there are any effects.⁶ 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). For example, being enmeshed in an iterated but potentially finite prisoners’ dilemma does not make the D,C payoff less desirable, in principle. All it does is change the costs and benefits of pursuing these preferences.

The undersocialized nature of institutions in contractual institutionalist arguments is highlighted by the factors that contractualists *do* focus on when theorizing how cooperation is elicited inside institutions. The first is issue-linkage where (per)suasion is simply an effort to change the cost/benefit calculations of the defecting player with exogenous positive or negative incentives so as to secure cooperation. It does not change that player’s underlying desire to defect.⁷ The second is reputation. The desire to establish a trustworthy reputation for future exchanges can be an incentive to engage in norm-conforming, pro-social behavior (Kreps, 1992). Reputation in this sense is an instrument; the rewards come from the private material benefits of future exchange, not the social or social-psychological benefits accrued by cultivating a status and image that is rewarded by the group. The final factor—perhaps the most important one for contractualists—is information. Interaction in institutions can provide new information that can reduce uncertainty about the credibility of others’ commitments, and thus help actors’ expectations converge around some cooperative outcome (Martin, 1997). Information only affects beliefs about the strategic environment in which the actor is pursuing fixed preferences. If information has an effect on preferences it is mainly through its impact on elite change: information about the failure of some strategy, for instance, could lead to a loss of support

⁴ 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).

⁵ Morrow admits that the conspicuousness or prominence of an 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).

⁶ 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 is 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.

⁷ See Martin’s discussion of suasion games (1993:104).

for one set of elites pursuing preferences and their replacement by another set with different preferences.

There is sort of an infinite regress problem with much of the work on information, however. What makes the meaning of information conclusive enough to affect behavior unless there is prior agreement on the criteria for the credibility of information? What leads to prior agreement on these criteria? Presumably 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 often *is* problematic.

Instead, contractualists often assume that the credibility of information rests on costliness to the provider of the information. In practice they see costliness mostly in terms of some loss of material welfare or political power. No doubt costs often take this form, but contractualists have no theoretical advantage here, no theory of the conditions under which new information will influence preferences, beliefs, or strategies and by how much.⁸ They often leave out the social context of information. The social origins of common definitions of costliness, essential for information to be credible, are unexamined. Yet empirically we know that the same information, even economic information, will be interpreted differently depending on whether it comes from “people like us” (the information is more authoritative and persuasive), or from a devalued “other” (Kuklinski and Hurley, 1996:127; Halpern, 1997; Valley, Moag, and Bazerman, 1998:230). Even in prisoners’ dilemma (PD) relationships, information about the other as an opportunist is not static. Hayward Alker (1996) reports on iterated PD games where after a string of mostly cooperative moves the players reinterpret the meaning of identical information. Defections that were interpreted as signals of the other’s malevolent or stupid nature before the cooperation streak were interpreted as situational or chalked up to random misperceptions afterward. 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 adversary).

To be fair, contractualist arguments do not *a priori* reject the possibility that information changes preferences instead of just beliefs about strategic environments. The advice is sometimes to test for both, but in practice the tendency is to discount the possibility of the former. This is primarily because preferences and changes in them are difficult to observe. What often may appear to be a change in preferences may, instead, be a change in strategies. Any likely source to which one might turn to “observe” preferences (e.g., from statements through to actions) could well be itself a product of strategic interaction, hence unrepresentative of true preferences. It is easier, therefore, to assume fixed preferences.⁹

This seems to be a reasonable, cautionary argument for a sound methodological choice. It does reveal, however, an implicit disciplining move that constrains

⁸ I just want to underscore that what constructivists focus on are 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 new information as it pertains to socialization (e.g., persuasion) is interesting precisely because it encourages 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 vs. unilateralism as a “cause” of security). This aligns constructivist work somewhat with more traditional work in political socialization that focuses on fundamental ideological dispositions (e.g., Beck and Jennings, 1991; Kinder and Sears, 1981; and Kinder and Sanders, 1996).

⁹ See Frieden, 1999, for a sophisticated statement of this argument.

efforts to think about changing preferences through new information acquired via social interaction inside an institution. Contractualists, I believe, overestimate the ease with which one can deduce preferences from some *a priori* features of an actor. 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. This is especially likely when the arena of action is not economics but security, politics, ideology, and culture where utilities and their metrics can vary dramatically. Thus, for instance, the preferences of military organizations (e.g., that they favor offensive doctrines and capabilities) cannot be pristinely deduced from some prior assumptions about the universal characteristics of military organizations (Kier, 1997).

And contractualists overestimate the difficulty of observing preferences and changes in them. To be sure, the 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 that might itself be strategic). But given the theoretical importance of the question it seems premature to give up trying to observe change (see Herrmann, 1988:180). Moreover, there are quite well-developed social and psychological survey and content analysis techniques that have been used for years in fields that take socialization seriously, including political socialization studies, and that do wrestle with validity and reliability questions. But most students of IR are not exposed to these technologies during their training.

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” (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, taken for granted.¹⁰ This naturally raises questions about which norms are internalized by agents, how and to what degree. Kratochwil and Ruggie (1986) 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. So some process of socialization must be going on.

Yet for much of the constructivist literature, socialization processes are unclear. There are a number of reasons for this. First, some constructivism inherits much of the epistemology of sociological institutionalism. A fair amount of empirical work in this regard has tended to focus on macrohistorical 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 (Price, 1998; Eyre and Suchman, 1996; Finnemore, 1996b). 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. Once actors are interacting inside institutions, the diffusion and homogenization of values in the “world polity” seems virtually automatic, even, and predictable. This leaves variation in the degree of socialization across units—the degree of contestation, normative “retardation,” the processes by which unit-level actors understand, process, interpret and act upon lessons that are “taught” by international institutions as agents—unexplained.¹¹ And it leaves the causal

¹⁰ I use the term *pro-norm* to indicate action that is consistent with the norm in question, whether done because the norm has been internalized or because some kind of consequentialist calculation makes it useful to follow. I do not mean that an actor is necessarily consciously “for” the norm.

¹¹ Even Finnemore’s detailed causal story of teaching often stops at the point where agents at the international level deliver norm-based lessons to rather passive students (1996b). In their discussion of the cognitive and social processes behind the evolution of security communities Adler and Barnett (1998) do not have much to say about resistance to such processes either. This characteristic of the literature is, perhaps, the natural result of a desire to show first of all that persuasion and socialization “matter.”

processes unexplicated.¹² This neglect in the literature is surprising, given constructivists' focus on reflective action by multiple agents: if this kind of agency exists in the diffusion of norms, what happens when it runs into reflective action by multiple agents at the receiving end of these "teaching" efforts?¹³

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. Here, however, there is a fair amount of variation in how the term is used. For some the term is 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 often borrow in some form or another from Habermas's theory of communicative action (Risse, 1997; Risse and Sikink, 1999:13; Black, 1999:102–103). 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 (Johnson, 1993:81). Thus, right from the start, bargaining involves argument and deliberation all in an effort to change the minds of others.¹⁴ As Hasenclever et al. put it, "the parties 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, 1997:176–177, emphasis mine; see also Knoke, 1994:3; James, 1998:7).

For others, 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 Sikink, 1999:13–14; Keck and Sikink, 1998:16).

There are a couple of issues 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. Habermasian approaches are unclear as to what constitutes a "convincing" argument. This is a huge requirement for argumentation and thus far constructivists have not really shown how debates over common knowledge, for example, "convince"

¹² See, e.g., 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 instance, anti-slavery norms in British diplomacy, but it is unclear why political leaders and government officials were persuaded by moral arguments (1990:494). Keck and Sikink (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 in a European identity. But it isn't clear why this persuasion ought to work on initial members who are somewhat illiberal and noncooperative. This neglect of microprocesses may change as scholars pick up on Finnemore and Sikink's insightful summary of some plausible causal processes (1998).

¹³ For similar critiques see Checkel, 1998:332, 335; Moravcsik, 1997:539; and Risse, 1997:2.

¹⁴ For an excellent exegesis of Habermas's theory of communicative action see Risse, 2000.

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?¹⁵ How would one know? The conditions seem to be quite demanding, involving a high degree of prior trust, empathy, honesty, and power equality. Some 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, identification leads to positive affect and positive affect leads to a greater probability that the arguments and interpretations of the other will be accepted as valid, and internalized. There is an endogeneity issue here, though nothing that complex adaptive systems epistemology would worry about. And the empirics have yet to be tested in much detail.

There is also a second, more important issue. 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—often the term is conflated with an entirely separate effect of social interaction that can also lead to pro-norm behavior in the absence of exogenous material threats or promises.¹⁶ This effect has been termed *social influence*. This term encompasses a number of subprocesses—backpatting, opprobrium or shaming, social liking, status maximization, etc.—where pro-norm behavior is rewarded with social and psychological markers from a reference group with which the actor believes it shares some level of identification. A focus on communicative action and/or the conflation of persuasion and social influence means that constructivists have a hard time distinguishing among the range of microprocesses that mediate between “teaching” attempts on the one hand and pro-norm behavior emanating from a foreign policy process on the other. This means they have a hard time explicating systematically the institutional conditions under which one or the other of these microprocesses might be at work. In the next section I focus on how one might distinguish between these microprocesses.

Socialization: Definitions and Microprocesses

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.” In Stryker and Statham’s words, “Socialization is the generic term used to refer to the processes by which the newcomer—the infant, rookie, or trainee, for example—becomes incorporated into organized patterns of interaction” (1985:325). Berger and Luckmann define the term as “the comprehensive and consistent induction of an individual into the objective world of a society or sector of it” (1966:130). Thus socialization is aimed at creating membership in a society where the intersubjective understandings of the society become taken for granted.

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” (1990:1). Siegal refers to political social-

¹⁵ See the conditions explicated by James (1998:7–11, 15–17)

¹⁶ Keck and Sikkink, for instance, refer to human rights networks being able to embarrass norm violators such that in order to save face they adjust their behavior (1998:24). But it is not clear why a norm violator would care about pressure that does not come with concrete threats of sanctions that affect wealth and relative military power. In one of their cases, for example, Argentina under military control in the 1970s, the differential effects of image per se and a desire to “restore the flow of military and economic aid” are not obvious (107).

ization 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). 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; see also Ikenberry and Kupchan, 1990:289–290).

There are a couple of common themes in the political science literature on socialization: the first is that socialization is most evidently directed at, or experienced by, novices and newcomers, whether they are children, inductees into a military, immigrants, or new states and their rulers. The second 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 “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 consequential terms. Why should one do X? “Because . . . ,” or “because X is the right thing to do . . . ,” or “because X is consistent with my social category or identity.”

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, this leaves a vast amount of pro-social behavior produced by neither process.

This leads to a key 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. But beyond persuasion, the literature on socialization (outside of IR theory) identifies a range of reasons why one might see pro-norm behavior in the absence of exogenous material (dis)incentives. Axelrod, for instance, lists identification (the degree to which an actor identifies with the 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-norm behavior (1997a:58–59). All of these depend on the acquisition of some kind of identification with or affective attachment to a group. Ikenberry and Kupchan list two routes to pro-norm behavior that do not involve persuasion: exogenous shocks that lead to elite transformation in a state; and exogenous material inducements that lead, over time (and somewhat mysteriously) to the internalization of norms that were once adopted for instrumental reasons (1990:290–292). 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 political personalities of younger members; or the political traits of parents are transmitted through a process of inheritance or mimicking (1991:744). Constructivism has tended to neglect many of these microprocesses.

Arguably these multiple processes boil down to two: persuasion and social influence. A critical question, then, is when and to what degree do these sepa-

rate processes help explain why actors change their behavior in pro-norm or pro-social ways. In practice these processes are likely to be interactive. But separating them out is important because the answer will help point to how durable pro-social conformity is over time and what kinds of institutional designs are most conducive to this durability. 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 purpose of modeling prevents one from exploring this important issue. Thus, broadly speaking, the speed, uniformity, and effectiveness of norm diffusion in international relations ought to depend a great deal on what kind of institutional social environment leads to what kind of socialization microprocess.

Persuasion

Persuasion has to do with cognition and the active assessment of the content of a particular message. As a microprocess of socialization, it involves changing minds, opinions, and attitudes about causality and affect (identity) in the absence of overtly material or mental coercion. It can lead to common knowledge, 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 that they share many cooperative interests. The point is, however, that the gap or distance between actors’ basic causal and affective understandings closes as a result of successful persuasion.

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

How persuasion works therefore is a focus of a great deal of research in communications theory, social psychology, and sociology. There is no obvious way of summarizing such a disparate and complex literature,¹⁷ but essentially there are three ways in which an actor is persuaded. First, s/he can engage in a high intensity process of cognition, reflection, and argument about the content of new information. The actor weighs evidence, puzzles through “counterattitudinal” arguments, and comes to conclusions different from those he/she began with; that is, the “merits” of the argument are persuasive, *given* internalized standards for evaluating truth claims. Arguments are more persuasive and more likely to affect behavior when they are considered systematically and, thus, linked to other attitudes and schema in a complex network of causal connections and cognitive cues (Wu and Shaffer, 1987:687; Petty, Wegener, and Fibrigar, 1997:616; Zimbardo and Leippe, 1991:192–197).

This process of cognition, linking one set of attitudes to another, is more likely to occur when the environment cues and allows for the actor to consider these connections. That is, it is less likely to be spontaneous than it is promoted. As Gibson has shown with political intolerance among Russian voters, intolerant

¹⁷ See Zimbardo and Leippe, 1991:127–167. Despite the volume of this literature, “To date there is precious little evidence specifying who can be talked out of what beliefs, and under what conditions” (Berger, 1995:8).

attitudes toward political opponents will change in more tolerant directions if counterattitudinal arguments are presented to respondents in ways that compel them to “think harder” about the implications of their initial attitudes. Thinking harder simply means people are cued, and have the time, to connect the implications of their initial attitude to outcomes that might affect their interests based on different sets of attitudes. Thus an initially intolerant view might change to a more tolerant one if the respondent is cued to think about the implications of cycles of intolerance for political stability or for opportunities for themselves to present their own political opinions in the face of opposition (Gibson, 1998:826–831). The probability of some change in attitudes through cognition increases in an iterated, cognition-rich environment where there is lots of new information that cues linkages to other attitudes and interests.

Second, the actor is persuaded because of her/his affect relationship to the persuader: here the persuadee looks for cues about the nature of this relationship to judge the legitimacy of counterattitudinal arguments. Thus information from in-groups is more convincing than that from out-groups. Information from culturally recognized authorities (e.g., scientists, doctors, religious leaders) is more convincing than that from less authoritative sources. This will be especially true for novices who have little information about an issue on which to rely for guidance (Zimbardo and Leippe, 1991:70; Gibson, 1998:821). Information from sources that are “liked” is more convincing than that from sources that are disliked. Liking will increase with more exposure, contact, and familiarity. The desire for social proofing means that information accepted through consensus or supermajority in a valued group will be more convincing than if the group were divided about how to interpret the message (Petty et al., 1997:612, 617, 623, 627, 629; Kuklinski and Hurley, 1996:129–131; Napier and Gershenfeld, 1987:159; Isen, 1987:206–210, 211; Axsom, Yates, and Chaiken, 1987:30–31).¹⁸

Third, the persuasiveness of a message may be a function of characteristics of the persuadee her/himself. This can refer to a range of variables from the cognitive-processing abilities of individuals in a group, to the strength of existing attitudes (usually these are stronger if developed through personal experience than if based on hearsay or indirect experience, for example), to what appears to be a deeply internalized desire to avoid appearing inconsistent, to the degree of independence an agent might have in relation to a principal. Thus, for example, an attitude associated with an explicit behavioral commitment made earlier will be more resistant to change later because actors experience discomfort at being viewed as hypocritical and inconsistent. Conversely, a new set of attitudes will be more persuasive if associated with a new, high-profile behavioral commitment (Cialdini, 1984; Wu and Shaffer, 1987:677). Thus a focus on the characteristics of the persuadee means looking at the individual features that can either retard or propel persuasion. All this means is that actors entering a social interaction bring with them particular prior traits that, interacting with the features of the social environment and other actors, leads to variation in the degree of attitudinal change.¹⁹

¹⁸ Using different language, Habermasian constructivists make a similar point: “trust in the authenticity of a speaker is a precondition for the persuasiveness of a moral argument” (Risse, 1997:16; see also Williams, 1997:291–292). Game theorists have come to a similar conclusion, only using another language. Lupia notes that persuasiveness rests basically on the persuadee’s belief that she or he shares common interests with the persuader and that the information the persuader is offering benefits both (1998). He does not specify what kind of information leads to the first belief. But it could, in principle, be anything from the list in the above paragraph.

¹⁹ Of course, persuasion in practice is likely to be a combination of all these microprocesses. Jorgensen et al. found in a study of televised political debates in Denmark, for example, that the most persuasive debaters were those who used a small number of extended, weighty discussions of specific qualitative examples. The use of these specific, straightforward and logical examples seemed to accentuate the authoritativeness of the debater and were easier for viewers to assess and adjudicate (see Jorgensen, Kock, and Rorbeck, 1998).

Lupia and McCubbins argue that all of these conditions and characteristics are simply indicators of more basic conditions for persuasion, namely, that the persuadee believes the persuader to be knowledgeable about an issue and that his or her intentions are trustworthy. The more certain the persuadee is about these beliefs, the more likely the persuader will be persuasive. Both these characteristics can be a function either of familiarity and extensive interaction that, over time, reveals them, *or* “external forces” that make it difficult or costly for the persuader to hide knowledge (or lack thereof) and trustworthiness (e.g., mechanisms for revealing knowledge, penalties for lying, costly actions that reveal the position of the persuader). Any other factors, such as ideology, identity, culture, and so forth, are only predictors of persuasion to the extent that they reveal information to the persuadee about the persuader’s knowledge and trustworthiness (1998).

Lupia and McCubbins present a rigorous formal model of persuasion that is probably correct in stripping the process down to these two pieces of perceived information. But this does not avoid the more interesting question about the empirical frequency with which *social* variables such as perceived ideology, identity, and/or cultural values are in fact the primary cues that people use to determine the degree of knowledge and trustworthiness of a persuader, and thus come prior to beliefs about knowledge and trustworthiness. On average is perceived shared identity between persuadee and persuader more likely to be used by the persuadee as an authoritative measure of a persuader’s knowledge and trustworthiness than other kinds of cues?

The answer has important implications for how social interactions lead to socialization and how different institutional designs might lead to different socialization paths. Lupia and McCubbins tend to focus, as befits their interest in signaling games, on the role of external forces in clarifying beliefs about the knowledge and trustworthiness of persuaders. They argue that since social and political environments are rarely ones where persuader and persuadee interact face to face over long periods of time, the familiarity/personal interaction route to beliefs about the persuader’s knowledge and trustworthiness tends to be less common. This may be true at the national level of persuasion (e.g., political messages from politicians aimed at masses of voters), but it is not necessarily true at the level of social interaction in international institutions among diplomats, specialists, and analysts. Here the first route—familiarity, iterated face-to-face social interaction—may be more common, hence affect based on identity, culture, and ideology may be more critical for persuasion than external forces and costly signals. Institutions, therefore, that are weak in terms of these external forces, nonetheless may create conditions conducive to persuasion—and convergence around group norms—even though there are few material incentives for the persuader to deceive and few material costs for the persuadee to defect from the group. I will come back to this at the end.

Persuasion in the end is a combination of all three processes above and it is hard to run controls that might isolate the effects of any one process. People are more likely to think hard and favorably about a proposition, for instance, when it comes from a high affect source, in part because affect helps kick in resistances to information from other sources (Mohr, 1996:81–82). On the other hand, one can identify ideal combinations that could, in principle, be tested. Given an effort by a persuader to provide information with a view to changing basic principled, causal, or factual understandings, there are certain kinds of social environments that ought to be especially conducive to persuasion. These conditions imply that certain institutional designs will be more effective for persuasion than others. These conditions occur

- when the actor is highly cognitively motivated to analyze counterattitudinal information (e.g., a very novel environment);

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

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

Assuming an actor enters the institution and its particular social environment with preferences and beliefs that are at odds with those of the group, if persuasion is at work one should expect to see (after exposure to this environment) the actor's convergence with these preferences and beliefs, and conformist behavior later in the interaction with the group that would not have been expected earlier on. In short, you should get increasing "comfort" levels with group values and normative practices even as the demands placed on the actor by the group for pro-social behavior increasingly violate the initial preferences and beliefs of the actor.

Social Influence

Social influence refers to a class of microprocesses that elicit pro-norm behavior through the distribution of social rewards and punishments. Rewards might include psychological well-being, status, a sense of belonging, and a sense of well-being derived from conformity with role expectations. Punishments might include shaming, shunning, exclusion, and demeaning, or dissonance derived from actions inconsistent with role and identity. The effect of (successful) social influence is an actor's conformity with the position advocated by a group as a result of "real or imagined group pressure" (Nemeth, 1987:237). The difference between social influence processes and persuasion is neatly summarized by the phrase Festinger used to describe compliance due to social pressure: "public conformity without private acceptance" (cited in Booster, 1995:96). Persuasion would entail public conformity with private acceptances. 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 can, instead, come 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" [cited in Betz, Skowronski, and Ostrom, 1996:116]). The rewards and punishments are social because only groups can provide them, and only groups whose approval an actor values will have this influence. Thus social influence rests on the "influenced" actor having 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]).

There is considerable evidence that identification with a group can generate a range of cognitive and social pressures to conform. But, like persuasion, the microprocesses of social influence are multiple, complex, and still the subject of

much debate. Generally, however, the literature on social influence has isolated a number of possibilities:

- cognitive discomfort associated with perceived divergence from group norms generates strong internal pressures to conform to the group's practice, that is, the trauma to self-esteem from this divergence can motivate an actor to reduce discrepancies through greater conformity;²⁰
- the sense of comfort that comes from interacting with others with whom she/he is perceived to share traits (social liking) leads to an increased willingness to comply with the requests of friends (Cialdini, 1984, 1987); and
- the discomfort with being perceived as inconsistent or hypocritical in relation to past actions and commitments, and conversely the positive moods with being viewed as consistent with one's self-professed identity, leads people whose consistency is challenged to respond by greater conformist behavior (consistency theory). Membership in a group usually entails "on the record" statements or behaviors of commitment (e.g., pledges of loyalty, participation in group activities, commitments to fulfill a membership requirement). These behaviors, even if relatively minor, establish a baseline identity such that behavior that diverges from these identity markers are discomforting inconsistencies. The more the identity-conforming behavior is repeated, the more extreme, and tenaciously held, the actor beliefs and attitudes become, thus reinforcing his/her commitment to the group (Petty et al., 1997:612, 620; see also Cialdini, 1984, 1987).

The most important microprocess of social influence, or at least most relevant to international relations theory given the prevalence of status language in interstate discourse, is the desire to maximize status, honor, prestige—diffuse reputation or image—and the desire to avoid a loss of status, shaming, or humiliation and other social sanctions. Status refers to "an individual's standing in the hierarchy of a group based on criteria such as prestige, honor, and deference." Typically, status is closely related to others' "expectations of ability or competent performance" (Lovaglia, 1995:402). Choi offers a useful definition: "An individual's status is communal certification of his or her relative proficiency in conventions" (1993:113). Thus, competency or proficiency need not mean a mechanical ability to do some task, but can mean a high ability to represent some normative ideal. A competent nonproliferator is, in the eyes of an antiproliferation community, a responsible actor and a consistent, effective proponent of nonproliferation norms. Image is the public manifestation of status. Image refers to the package of perceptions and impressions one believes one creates through status-consistent behavior.

There are numerous motivations behind maximizing status. Often status brings with it power, wealth, and deference, and vice versa. Gilpin, for instance, refers to prestige as "a reputation for power." States are at the top of the status hierarchy because of their economic and military power (1981:30–33). Moreover, in Gilpin's view, status is highly coercive: status markers are forced out of subordinate states through superior power, often through military victory. This is fine as far as it goes, but often status markers and immediate material gains are not correlated. For example, status markers such as citations, medals, or public recognition may have no obvious material reward. Moreover, the desire to maximize status need not entail efforts to defeat others to seize status: it can entail group-conforming behavior designed to "buy" status. The reward is psychological well-being from backpatting; the punishment is psychological anxiety from opprobrium.

²⁰ On SIT and the psychological discomforts of nonconformity see Turner, 1987; Gerard and Orive, 1987; Stryker and Statham, 1985; Barnum, 1997; Axelrod, 1997a.

A second possible motivation is to maximize reputational effects attributed to particular status markers. Here status is an instrument: a good image can encourage actors to deal with you in other arenas, can help build trust leading to reciprocity and decentralized (uninstitutionalized) cooperation (Kreps, 1992). In this sense, image can also be used deceptively—one might want a positive image to convince other states to cooperate, setting them up for the sucker's payoff in some exploitative prisoners' dilemma game. There are two problems with this conceptualization, however. The first is, as Frank points out, if people know about this instrumentality, then an actor's image or reputation as a cooperators has no advantage (1988). So it is in the actor's interest to make cooperation automatic, deeply socialized, in order to make the reputation for cooperation credible. But then no advantages can be accrued, since deception is abandoned. The second problem is that instrumentality assumes the actor is seeking some concrete, calculable benefit from having a good image, an image that can be translated into leverage in some explicit, linked, immediate issue area. Yet often there are no obvious concrete benefits, or they are quite diffuse and vague. Indeed, sometimes there are concrete material costs. In this case, sensitivity to image may be related to identity.

This is the third reason for a concern about status. A particular high status image may be considered a good in and of itself. Frank argues that the desire to maximize prestige and status has physiological and psychological benefits (1985:32). Harre attributes the drive to people's "deep sense of their own dignity, and a craving for recognition as beings of worth in the opinions of other of their kind." To be fulfilled, this desire necessarily depends on *public* affirmation of one's status (1979:3, 22). Hatch notes that "the underlying motivation is to achieve a sense of personal accomplishment or fulfillment, and the individual does so by engaging in activities exhibiting qualities that are defined by the society as meritorious" (1989:349). Franck argues, in reference of the fact that most states abide by most institutional legal commitments most of the time in IR, conformist behavior is due mainly to a desire to be a member of a club and to benefit from the status of membership (1990:38).

An actor will be sensitive to arguments that her/his behavior is consistent or inconsistent with their self-identity as a high-status actor. This sensitivity ought to depend as well on who is making these arguments. The more the audience or reference group is legitimate, that is, the more it consists of actors whose opinions matter, the greater the effect of backpatting and opprobrium (Dittmer and Kim, 1993:9, 14–15). The legitimacy of the audience is a function of self-identification. Actors more easily dismiss the criticisms of enemies and adversaries than they do of friends and allies. If, for example, an actor completely rejected the social norms of a particular group, then no matter what the size of that group it could not generate backpatting or shaming effects. Thus the strength of backpatting and opprobrium depends on two related factors: the nature of the actor's self-categorization, and which other actors, by virtue of this self-identification, become important, legitimate observers of behavior. Changes in identities mean that different audiences matter differently.

All of this hinges, of course, on an intersubjectively agreed upon notion of what socially valuable behavior looks like. I would argue, then, that the production of positive and negative social sanctions sufficient to induce cooperation in the absence of material side-payments or threats rests on two tiers. First, there must be an intersubjective normative consensus about what "good" behavior looks like.²¹ Without this shared standard, then the "fact" of some particular action will have no agreed interpretation, and consequently it will have no

²¹ Franck makes a similar point about symbolic validation of participation in international institutions (1990:117).

meaning, generating no shaming or backpatting effects. Thus, while social influence is not as direct or pristine an example of socialization (persuasion leading to changes in preferences), it could be considered a secondary socialization process because it requires at least some prior change to an actor's understanding of the group's normative preferences.

Second, even if there is a shared interpretation of the meaning of a particular behavior, these actions will not generate social pressures if they are unobserved and private. Thus the second layer is a forum or institution that makes acting a particular way public and observable. The forum could be something as loose as a process where voluntary reporting on some agreed commitment is scrutinized, where defectors would stand out by either not submitting a report or by submitting shoddy and incomplete ones. Or it could be something as strict as a multilateral negotiation process where actors are required to state bargaining positions, justify them, and then "vote" in some form on the proposed solution.

Thus constructivists and institutionalists are both right. Constructivists are right that socially induced cooperation requires shared understandings of what appropriate behavior looks like. But this may not be enough without an institutional structure that provides information about the degree to which actors are behaving in ways consistent with this shared understanding.²² This information makes the distance between an actor's behavior and the socially approved standard public. It is this distance that generates backpatting and shaming effects. In principle, the larger the relevant audience of cooperators, the more powerful these effects are.

The converse of social backpatting is shaming or opprobrium derived from violating status-related norms and practices. It is widely accepted in a number of subfields that fear of opprobrium is a motivation for group conformity, even if suboptimal from a welfare perspective. As Oran Young remarks *a propos* of international institutions, "Policy makers, like private individuals, are sensitive to the social opprobrium that accompanies violations of widely accepted behavioral prescriptions. They are, in short, motivated by a desire to avoid the sense of shame or social disgrace that commonly befalls those who break widely accepted rules" (1992:176–177; see also DiMaggio and Powell, 1991:4). The specific micro-processes that compel people to avoid opprobrium are similar to those that encourage the accumulation of backpatting.

It is important to note that pro-social behavior motivated by status maximization is not altruistic or pro-group *per se*. Rather it reflects an actor's egoistic pursuit of social rewards and avoidance of social sanctions (Batson, 1987:65). But these rewards and sanctions cannot exist without the prior existence of a group and without a common understanding of the value or meaning that the group places on putative status markers. This much, at least, must be shared by the actor and the group.

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

Assume, for the moment, that the actor in question has internalized a traditionally realpolitik concern about shielding relative power (military and economic) from potentially constraining commitments to international regimes. One can model the actor's diplomacy using a simple N-person's prisoners' dilemma model (Figure 1). The C line represents the payoffs to the actor who cooperates

²² Keohane notes, for instance, that one of the things international institutions do is provide a forum in which an actor's conformity with group standards can be evaluated. He links this to a more instrumental notion of reputation than I do here, however (1984:94)

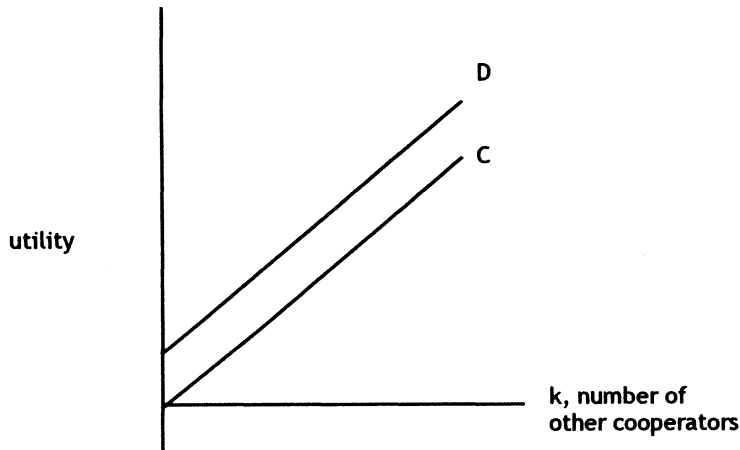


FIG. 1. Free-riding.

when exactly k other players cooperates, in the sense of limiting their military activities. The D line represents the payoffs to the actor from defection when exactly k other players cooperates, in the sense of limiting their military activities. If the realpolitik actor were the only cooperator (e.g., if it were the only one trying to reduce some global security “bad” while others continued to maximize their relative power) it would be constraining its relative power while having little effect on stabilizing the security environment. If it did not cooperate while others also defected, then, although it could not derive any benefits from the cooperation of others, it would be better off than if it unilaterally cooperated. Thus it would pay not to cooperate even if there were no other cooperators. This payoff from defection would hold even as the number of cooperators increased. As these players contributed to a public good the actor would benefit from the provision of this good, but by free-riding it would not incur the cost of providing its share of the good. Thus the payoff line from defection will always be greater than the payoff line from cooperation.

However, if this actor is also sensitive to social rewards and punishments, then social interaction can induce caution in the pursuit of a defection strategy that might have an adverse effect on status. Within international organizations and institutions the participating/cooperating audience can be relatively large. While the opportunities to free-ride are potentially greater—given the number of potential cooperators—the scrutiny of each player is more intense and state behavior is often more transparent than in bilateral relations, due to the rules of these institutions. In this context, a concern about image has two very different effects on a realpolitik actor’s payoff structure, corresponding to the effects of backpatting and opprobrium.

Backpatting is a benefit incurred from being seen as a cooperator or an active pro-social member of a group. An actor receives recognition, praise, and normative support for its involvement in the process. Backpatting can reaffirm an actor’s self-valuation, its self-categorization as a high-status actor, with concomitant payoffs for self- and public legitimation. *Ceteris paribus*, as the size of the cooperating audience grows, the actor accrues more backpatting benefits. Thus for every additional member of the institution, a potential defector receives a certain added payoff from backpatting as long as it cooperates. The benefits are cumulative. As Figure 2 indicates this increases the slope of the payoffs from cooperation (from C to C’).

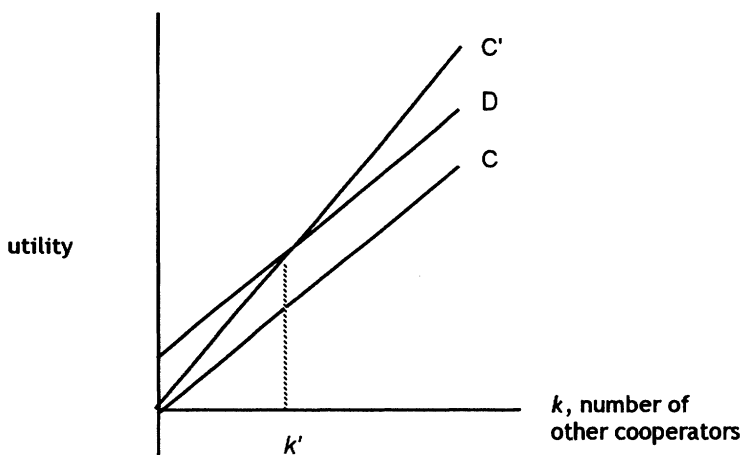


FIG. 2. Backpatting.

Opprobrium, of course, carries social costs—a denial of the prior status and prestige of the actor—as well as psychological ones—a denial of the actor's identity as one deserving of backpatting. Opprobrium can also be modeled as an accumulation of shaming markers that diminishes the value of free-riding as the number of participants/cooperators in a regime increases. A certain social cost is incurred with each additional participant/observer in the reference group. As the group increases the criticisms accumulate, and this increases the costs of defection. The effect, as shown in Figure 3, is to depress the slope of the payoffs from defection (from D to D'). At a certain point, an increase in the slope of the payoffs from cooperation and/or a decrease in the slope of the payoffs from defection may create a crossover point in the two lines. This is the point where the size of the audience (k') is such that the backpatting benefits and opprobrium costs change the cost-benefit analysis. It is at this point that it begins to pay to cooperate as the size of the audience increases.

When backpatting benefits and (implicit or threatened) opprobrium costs are combined this can dramatically reduce the size of the audience needed to make it pay to cooperate (this is shown by k' in Figure 4).

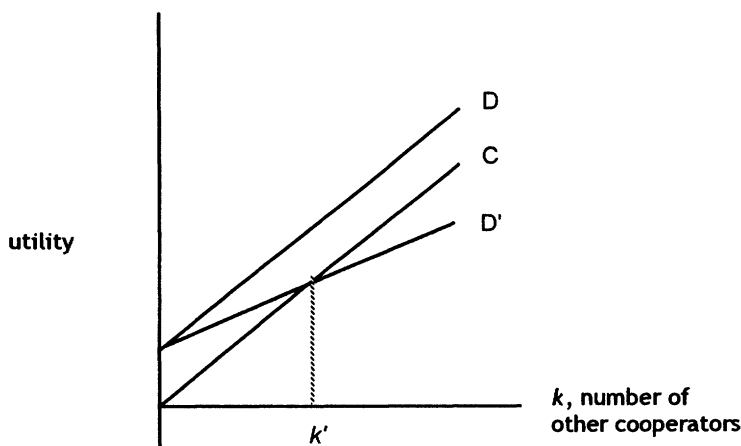


FIG. 3. Opprobrium.

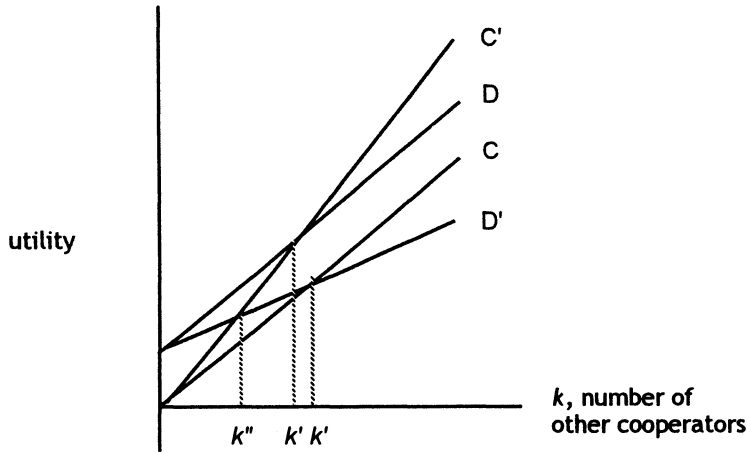


FIG. 4. Backpatting and opprobrium.

Note that the *net* effects of social influence on the cost-benefit calculus of cooperation in an institution appear to be similar to the provision of material side-payments and sanctions. It is important to point out, however, that backpatting and shaming change this cost-benefit calculus in a very different way than side-payment or sanctions. Typically (though not always) side-payments or sanctions, whether provided by the institution or by a key player or players in the institution, have a constant effect on an actor's utility regardless of how many others backpat or shame. Put graphically (Figure 5), the effect of the side-payments and sanctions is to raise the entire C payoff line and/or depress the entire D payoff line, respectively, while not changing their slopes, such that C' payoff line ends up above the D payoff line, or D' payoff line ends up below the C payoff line. A sanction for defection (imposed by an enforcer or hegemon, for instance) is equally costly regardless of the size of the group cooperating. Coop-

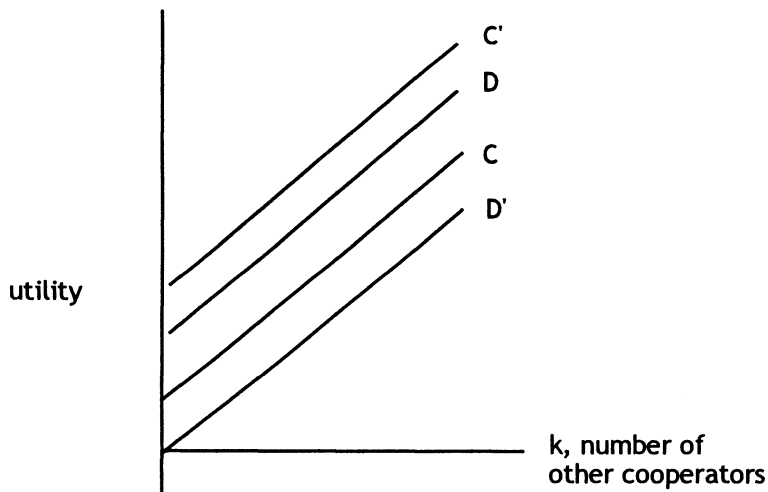


FIG. 5. The effects of material side-payments and sanctions.

eration brings higher utility than defection regardless of what other members of the group do. Thus the audience size and its legitimacy are irrelevant.

Put differently, cooperation as a result of material side-payments or sanctions is not a social effect of the institution. Cooperation due to social influence, however, is a social effect *only*, and would not exist without interaction with a group. Therefore backpatting and opprobrium are uniquely social inducements to cooperation, requiring a forum for social interaction in which the behavior of the potential free-rider is judged. Backpatting and opprobrium lose their impact outside of a social group. The fact that this forum is legitimately designed to promote cooperation accentuates the legitimacy and weight of the social backpatting and opprobrium directed at potential defectors. These forms of social influence would not carry any weight if unilaterally directed by one actor at another in a bilateral, institutionless relationship.

But how would one know if social influence had led to pro-social/pro-normative behavior in international institutions? Controlling for effects on relative power and the presence or absence of material side-payments and punishments, if social influence is at work, one should expect to see the following effects:

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

Empirical Issues

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 agentlike “teaching” properties, what are the norms and associated behaviors that actors in the environment are supposed to adopt and, hopefully, internalize? In other words, what is the predominant ideology in 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, mutually constitutive or feedback relationship and mediated by a policy process. One way of testing for socialization in IR, 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. My reasoning is as follows:

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 obvious 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, and nongovernmental agents of state

principals). Moreover, treating the unitary state as actor presents problems when applying the socialization literature found in social psychology, sociology, communications theory, and even in political socialization theory. For most of this literature the unit of analysis is the individual or small group.

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 (1997; see also Axelrod, 1997b, 1997c; Hamman, 1998).²³ Thus it matters how individual agents or small groups are socialized because their impacts on larger emergent properties of the social environment can be quite dramatic.²⁴ 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).

Thus, there are good reasons for studies of socialization to "go micro" and focus on the socialization of individuals, small groups, and, in turn, the effects of these agents on the foreign policy processes of states.²⁵

But why choose international institutions as the "agentlike" environments for 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.

There are a couple of reasons to focus on institutions. The most obvious is that because sociological approaches offer a clear alternative to contractual approaches to explaining cooperation, it makes sense to focus on the forum or arena on which contractualism itself focuses—institutions under anarchy.

But another powerful, though less obvious, reason is that in the late twentieth and early twenty-first century, international institutions are likely to be the one arena of inter-state activity where the effects of "anarchy" are likely to be checked. Even contractualists make this claim. They argue, of course, that institutions are a rational response to the interest in cooperation when anarchy makes such cooperation difficult to enforce. The argument is that anarchy—institutionless—will make it too dangerous for states to explore cooperation. This claim is rooted in neorealist logic even though, unlike neorealism, it assumes institutional effects are not necessarily epiphenomena of power distributions.

Constructivists, on the other hand, will have to argue that one of the important checks on anarchy is actor socialization in non-realpolitik directions. To do this, they have to also argue that realpolitik behavior is a product of socialization in realpolitik ideology. Put differently, one of the critical claims constructivists make is that "anarchy is what states make of it." 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

²³ For research tractability, however, it makes sense to look at separate parts of these feedback loops, as I suggest here, separating out the institutional "teaching" from the "individual socialization" from the impact on state policy from the impact on institutional ideologies, each process endogenized and exogenized at different stages.

²⁴ This is, after all, the point of much of the work on how transnational networks affect state behavior (Keck and Sikkink, 1998; Evangelista, 1999), "teaching" and the diffusion of norms, and the creation of national interests (Finnemore, 1996b). 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 also Ruggie, 1998:29).

²⁵ Ruggie calls this a focus on "innovative micro-practices," a hallmark of constructivist research (1998:27).

politics. In order to make this case, constructivists and their fellow-travelers have, for the most part, underscored the empirical “deviations” from realist or material power-interests theories: altruistic foreign aid; weapons taboos; “autistic” force postures in developing states; “autistic” military doctrines; and limits on the conduct of war (Lumsdaine, 1993; Price and Tannenwald, 1996; Eyre and Suchman, 1996; Kier, 1997; 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 realist theories claim they can explain; that is, constructivists are going to have to examine the argument that realpolitik practice is a reflection of realpolitik ideology and norms.²⁶ Critics of constructivism have a simple answer to this issue: realpolitik ideology is simply an epiphenomenon of anarchical material structures, the ideational superstructure that one should expect to find if states are trying to ensure their security in an uncertain world of shifting power distributions.

But the story cannot end there because two radically alternative explanations for the same phenomenon exist. The proper next step is to set up a critical test where one spins out alternative but *competitive* propositions and expectations from the two sets of explanations to see which additional set of empirical observations is confirmed or disconfirmed. One additional empirical implication that could provide an important test of constructivist versus material realist accounts of realpolitik is the phenomenon of counter-realpolitik socialization. If constructivist arguments 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.²⁷

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 particular 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 norm entrepreneurs, is most likely, and where social conformity pressures are most concentrated. Institutions often have corporate identities, traits, missions, normative cores, and official discourses at odds with realpolitik axioms, indeed at odds with the socialization pressures that many realists argue come with being sovereign, insecure actors operating in anarchy.²⁸ Where else, indeed, would state agents who have internalized real-

²⁶ 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 (1993:86–120). I do not define realpolitik simply as the “prudent” pursuit of the power 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.

²⁷ Note that I do not accept that an actor’s sensitivity to changes in relative power confirms material realism. I’ve argued elsewhere that one could argue this sensitivity is ideationally rooted. 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.

²⁸ For a discussion of organizations and their “goals” see Ness and Brechin, 1988:247, 263–266. See also Muller’s discussion of the ideology of the nonproliferation regime and how the causal and principled ideas of the regime relate to its norms and proscriptive regulations (1993). See also 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 (1998:134–135).

politik ideologies be exposed to alternative “theories” about the nature of world politics and the routes to security? They are unlikely to be exposed to these sorts of arguments in the domestic policy process or inside their organizations or through contact with the myriad channels through which the state constructs a competitive, often zero-sum picture of the external world (e.g., education systems, propaganda systems). Quite literally, for many of these agents the only sustained exposure to counter-realpolitik arguments and normative structures is often in international security institutions whose own ideology emphasizes cooperation, transparency, confidence-building, and demilitarization.

Thus, for example, some arms control institutions expose actors to an ideology where *inter alia*: multilateral transparency is normatively better than unilateral nontransparency; where disarming is better than arming as a basis of security; where common security is better than unilateral security; and where 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 counterattitudinal socialization going on, it ought to be happening in particular kinds of security institutions.²⁹ I do not mean to imply that institutions are the only fora in which socialization in IR occurs. 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. The question is simply how broader non-realpolitik norms in international security might be diffused. Institutions are an obvious first place to look, especially when testing the hypothesis that socialization occurs in the first place.³⁰

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

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

Recalling my earlier discussion of the conditions under which different kinds of socialization effects will occur, then, different institutional designs (combinations of measures on these five dimensions) should 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 persuasion is likely to be the most prevalent and powerful socialization process when membership is small (social liking and in-group identity effects on the persuasiveness of counterattitudinal messages are strongest); when franchise recognizes

²⁹ Risse makes a similar point, suggesting that communicative action should be more frequent inside institutions than outside of them (1997:17).

³⁰ 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” (1997:8).

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 where cognition effects will be strongest); when the institution's mandate is deliberative (this, again requires active complex cognition; agents may also be more autonomous since there is no obvious distribution of benefits at stake and thus there is less pressure to represent the principal); and when the autonomy of agents is high (e.g., when the issue is narrowly technical or when the principal just doesn't care much, or when the principal is less attentive or relevant). All these design-dependent effects will be enhanced for novices who are exposed to the environment over long periods of time (Zimbardo and Leippe, 1991: ch. 5).

Conversely, 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" or "persuasive" sources of new information); when decision rules are majoritarian (behavior is on record and consistency effects may be stronger); when the mandate involves negotiations over the distribution of benefits; and when the autonomy of agents is low (agents have to represent principals thus reducing the effects of persuasion on agents).

But how would one know if persuasion or social influence had led to pro-social/pro-normative behavior in international institutions? First, as I noted above, one would have to show that social environments in institutions are conducive to persuasion or social influence. 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.

Implications

If my general arguments about socialization are plausible, a focus on institutions as social environments raises at least two implications that challenge predominant arguments in IR theory about the conditions for cooperation inside international institutions.

The first implication has to do with persuasion and institutional design. Typically, contractual institutionalists argue that efficient institutional design depends on the type of cooperation problem (e.g., a PD-type problem requires information (monitoring) and sanctions; an assurance problem primarily requires reassurance information). The flip side is that one can identify inefficient institutional designs for a particular cooperation problem as well (e.g., an institution that is designed only to provide assurance information but has no monitoring or sanctioning capacity would be inefficient for resolving PD-type problems). Additionally, Downs et al. argue that so-called transformational institutions (inclusive institutions that bring genuine cooperators and potential defectors together in an effort to instill norms and obligations in the latter) are less likely to provide efficient solutions than a strategic construction approach. This latter approach to institutional design stresses exclusive memberships of true believers where decisions are made on the basis of super-majority rules. The gradual inclusion of potential defectors under these conditions ensures that the preferences of the true believers predominate as the institution evolves. Their critique of the transformational approach rests explicitly on skepticism that the *preferences* of potential defectors can change through social interaction (Downs, Rocke, and Barsoom, 1998).

It is not clear whether this skepticism rests on empirical evidence or simply on the methodological difficulties of assuming and then trying to observe preference change. In any event, if one relaxes that assumption then one is forced to revisit the contractual institutionalists' notions of efficient institutional design. An institution that appears inefficient to contractual institutionalists (e.g., an assurance institution for a PD problem) may actually be efficient for the cooperation problem at hand. If, say, a player with PD preferences can be socialized (persuaded) to internalize stag hunt preferences through interaction in a social environment with no material sanctioning or side-payments, then "assurance" institutions may work in PD-like cooperation problems. An efficient institution might then be reconceived as the *design* and *process* most likely to produce the most effective environments for socializing actors in alternative definitions of interest. As I have argued, the literature on socialization microprocesses suggests such an institution may have to be informal, weakly institutionalized, consensus-based—the opposite of an institutional design that contractualists believe is effective for dealing with PD problems.

The second implication comes from arguments about social influence and has to do with the problem of collective action inside institutions. Social influence effects may provide insights into how groups resolve the collective action problem that hinders resolving collective action problems. That is, traditionally scholars have argued that a critical solution to free-riding is to offer material side-payments (and sanctions) to make collective action pay for the individual. The conundrum has been, however, that offering side-payments is itself a collective action problem. Who will take up the burden of offering side-payments? Hege-mons and activists are usually part of the answer to this puzzle (though why activists should exist in the first place is hard for collective action theorists to specify *a priori*). Social rewards and punishments, however, are a particularly interesting kind of incentive to overcome collective inaction. They are relatively cheap to create, but are infused with a great deal of value. This means that new status markers can be manufactured and distributed without necessarily diminishing their value. In principle any member of a group therefore can provide social side-payments at relatively low cost, indeed at zero cost if the member can also receive these kinds of side-payments for providing them to others. This is, after all, what backpatting entails—a mutual, virtuous circle of bestowing and receiving social rewards. Cheap, but social, talk, then, can indeed be cheap to produce but nonetheless still be considered credible precisely because of its social value.³¹ Thus, because status markers are so highly valued, it doesn't take much of a "costly commitment" by providers of these markers to establish the credibility of promises to bestow, or threats to retract, these markers. All this suggests, then, is that one reason why collective action problems are often less frequent and debilitating than theorists expect (Green and Shapiro, 1994:72–97) may have to do with the fact that actors are also motivated by the desire to maximize social rewards and that these are relatively easy for groups to produce and distribute.

Following from this argument about collective action, social influence arguments also challenge the conventional wisdom about the optimal size of institutions and groups. From a contractual institutionalist perspective, *ceteris paribus*, more actors makes cooperation more difficult (collective action problems, problems of monitoring and punishing defection, etc.). Transaction costs increase with more actors. Decentralized institutions are therefore handicapped in dealing with "problems of transaction costs and opportunism" (Abbott and Snidal,

³¹ This is not dissimilar to Johnson's argument that cheap talk, in the context of persuasion whereby interests and identities converge inside a social relationship, establishes focal points that are necessary to reduce the strategic indeterminacy of bargaining games (see Johnson, 1993).

1998:15). From a social influence perspective, however, more may be better. Status backpatting and opprobrium effects are likely to be stronger when the “audience” or reference group is larger.

References

- ABBOTT, K. W., AND D. SNIDAL (1998) Why States Act Through Formal International Organizations. *Journal of Conflict Resolution* 42:3–32.
- ADLER, E. (1998) “Seeds of Peaceful Change: The OSCE’s Security Community-Building Model.” In *Security Communities*, edited by E. Adler and M. Barnett, pp. 119–160. Cambridge: Cambridge University Press.
- ADLER, E., AND M. BARNETT (1998) “A Framework for the Study of Security Communities.” In *Security Communities*, edited by E. Adler and M. Barnett, pp. 3–28. Cambridge: Cambridge University Press.
- ALKER, H. (1996) “Beneath Tit for Tat.” In *Rediscoveries and Reformulations: Humanistic Methodologies for International Studies*, edited by H. Alker, pp. 303–331. Cambridge: Cambridge University Press.
- ALTER, K. (1998) Who Are the “Masters of the Treaty”? European Governments and the European Court of Justice. *International Organization* 52:121–147.
- AXELROD, R. (1997a) “Promoting Norms: An Evolutionary Approach to Norms.” In *The Complexity of Cooperation*, edited by R. Axelrod, pp. 44–68. Princeton, NJ: Princeton University Press.
- AXELROD, R. (1997b) “A Model for the Emergence of New Political Actors.” In *The Complexity of Cooperation*, edited by R. Axelrod, pp. 124–144. Princeton, NJ: Princeton University Press.
- AXELROD, R. (1997c) “The Dissemination of Culture: A Model with Local Convergence and Global Polarization.” In *The Complexity of Cooperation*, edited by R. Axelrod, pp. 148–177. Princeton, NJ: Princeton University Press.
- AXSOM, D., S., YATES, AND S. CHAIKEN (1987) Audience Response Cues as a Heuristic Cue in Persuasion. *Journal of Personality and Social Psychology* 53:30–40.
- BARNUM, C. (1997) A Reformulated Social Identity Theory. *Advances in Group Processes* 14:29–57.
- BATSON, C. D. (1987) Prosocial Motivation: Is It Ever Truly Altruistic? *Advances in Experimental Social Psychology* 20:65–122.
- BECK, P. A., AND M. K. JENNINGS (1991) Family Traditions, Political Periods, and the Development of Partisan Orientations. *Journal of Politics* 53:742–763.
- BERGER, C. R. (1995) “Inscrutable Goals, Uncertain Plans and the Production of Communicative Action.” In *Communication and Social Influence Processes*, edited by C. R. Berger and M. Burgoon, pp. 1–28. East Lansing: Michigan State University Press.
- BERGER, P. L., AND T. LUCKMANN (1966) *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. New York: Anchor Books.
- BETZ, A. L., J. K. SKOWRONSKI, AND T. M. OSTROM (1996) Shared Realities: Social Influence and Stimulus Memory. *Social Cognition* 14:113–140.
- BLACK, D. (1999) “The Long and Winding Road: International Norms and Domestic Political Change in South Africa.” In *The Power of Human Rights: International Norms and Domestic Change*, edited by T. Risse and K. Sikkink, pp. 78–108. Cambridge: Cambridge University Press.
- BOOSTER, F. J. (1995) “Commentary on Compliance-gaining Message Behavior Research.” In *Communication and Social Influence Processes*, edited by C. R. Berger and M. Burgoon, pp. 91–113. East Lansing: Michigan State University Press.
- BRENNER, T. (1998) Can Evolutionary Algorithms Describe Learning Processes? *Journal of Evolutionary Economics* 8:271–283.
- CEDERMAN, L.-E. (1997) *Emergent Actors in World Politics: How States and Nations Develop and Dissolve*. Princeton, NJ: Princeton University Press.
- CHAYES, A., AND A. H. CHAYES (1996) *The New Sovereignty: Compliance with International Regulatory Agreements*. Cambridge, MA: Harvard University Press.
- CHECKEL, J. T. (1998) The Constructivist Turn in International Relations Theory. *World Politics* 50:324–348.
- CHOI, Y. B. (1993) *Paradigms and Conventions: Uncertainty, Decision Making, and Entrepreneurship*. Ann Arbor: University of Michigan Press.
- CIALDINI, R. (1984) *Influence: The New Psychology of Modern Persuasion*. New York: Quill Books.
- CIALDINI, R. (1987) “Compliance Principles of Compliance Professionals: Psychologists of Necessity.” In *Social Influence: The Ontario Symposium*, vol. 5, edited by M. P. Zanna, J. M. Olson, and C. P. Herman, pp. 165–184. Hillsdale, NJ: Lawrence Erlbaum.
- DiMAGGIO, P. J., AND W. W. POWELL (1991) Introduction to *The New Institutionalism in Organizational Analysis*, edited by W. W. Powell and P. J. DiMaggio, pp. 1–38. Chicago: University of Chicago Press.

- DITTMER, L., AND S. S. KIM, EDs. (1993) *China's Quest for National Identity*. Ithaca, NY: Cornell University Press.
- DOWNs, G. W., D. M. ROCKE, AND P. N. BARsOOM (1998) Managing the Evolution of Multilateralism. *International Organization* 52:397–419.
- EVANGELISTA, M. (1999) *Unarmed Forces: The Transnational Movement to End the Cold War*. Ithaca, NY: Cornell University Press.
- EYRE, D., AND M. sUCHMAN (1996) "Status, Norms, and the Proliferation of Conventional Weapons." In *The Culture of National Security*, edited by P. J. Katzenstein, pp. 79–113. New York: Columbia University Press.
- FARKAS, A. (1998) *State Learning and International Change*. Ann Arbor: University of Michigan Press.
- FINNEMORE, M. (1996a) Norms, Culture, and World Politics: Insights from Sociology's Institutionalism. *International Organization* 50:325–347.
- FINNEMORE, M. (1996b) *National Interests in International Society*. Ithaca, NY: Cornell University Press.
- FINNEMORE, M. (1996c) "Constructing Norms of Humanitarian Intervention." In *The Culture of National Security: Norms and Identities in World Politics*, edited by P. J. Katzenstein, pp. 153–185. New York: Columbia University Press.
- FINNEMORE, M., AND K. sIKKINK (1998) International Norm Dynamics and Political Change. *International Organization* 52:887–917.
- FRANCK, T. M. (1990) *The Power of Legitimacy Among Nations*. New York: Oxford University Press.
- FRANK, R. (1985) *Choosing the Right Pond: Human Behavior and the Quest for Status*. New York: Oxford University Press.
- FRANK, R. (1988) *Passions Within Reason: The Strategic Role of the Emotions*. New York: W. W. Norton.
- FREEDMAN, P. E., AND A. FREEDMAN (1981) "Political Learning." In *The Handbook of Political Behavior*, vol. 1, edited by S. Long, pp. 255–303. New York: Plenum Press.
- FRIEDEN, J. (1999) "Actors and Preferences in International Relations." In *Strategic Choice in International Relations*, edited by D. Lake and R. Powell, pp. 39–76. Princeton, NJ: Princeton University Press.
- GERARD, H. B., AND R. ORIVE (1987) The Dynamics of Opinion Formation. *Advances in Experimental Social Psychology* 20:171–202.
- GIBSON, J. L. (1998) A Sober Second Thought: An Experiment in Persuading Russians to Tolerate. *American Journal of Political Science* 42:819–850.
- GILPIN, R. (1981) *War and Change in World Politics*. Cambridge: Cambridge University Press.
- GREEN, D. P., AND I. SHAPIRO (1994) *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science*. New Haven, CT: Yale University Press.
- HAAS, P. (1998) "Constructing Multilateral Environmental Governance: The Evolution of Multilateral Environmental Governance Since 1972." Paper presented at the Center for International Affairs, Harvard University, April 16.
- HALPERN, J. J. (1997) Elements of a Script for Friendship in Transactions. *Journal of Conflict Resolution* 41:835–868.
- HAMMAN, H. L. (1998) "Remodeling International Relations: New Tools from New Science?" In *International Relations in a Constructed World*, edited by V. Kubalkova, N. Onuf, and P. Kowert, pp. 173–192. Armonk, NY: M. E. Sharpe.
- HARRE, R. (1979) *Social Being: A Theory for Social Psychology*. Oxford: Basil Blackwell.
- HASENCLEVER, A., P. MAYER, AND V. RITTBERGER (1997) *Theories of International Regimes*. Cambridge: Cambridge University Press.
- HATCH, E. (1989) Theories of Social Honor. *American Anthropologist* 91:341–353.
- HERRMANN, R. (1988) The Empirical Challenge of the Cognitive Revolution: A Strategy for Drawing Inferences About Perceptions. *International Studies Quarterly* 32:175–203.
- ICHILOV, O. (1990) Introduction to *Political Socialization, Citizenship Education and Democracy*, edited by O. Ichilov, pp. 1–8. New York: Teachers College Press.
- IKENBERRY, G. J., AND C. KUPCHAN (1990) Socialization and Hegemonic Power. *International Organization* 44:283–315.
- ISEN, A. M. (1987) Positive Affect, Cognition and Social Behavior. *Advances in Experimental Social Psychology* 20:203–253.
- JAMES, M. R. (1998) "Communicative Action and the Logics of Group Conflict." Paper prepared for the American Political Science Association Annual Meeting, Boston, September.
- JOHNSON, J. (1993) Is Talk Really Cheap? Prompting Conversation Between Critical Theory and Rational Choice. *American Political Science Review* 87:74–86.
- JORGENSEN, C., C. KOCK, AND L. RORBECH (1998) Rhetoric That Shifts Votes: An Exploratory Study of Persuasion in Issue-oriented Public Debates. *Political Communication* 15:283–299.

- KECK, M. E., AND K. SIKKINK (1998) *Activists Beyond Borders: Advocacy Networks in International Politics*. Ithaca, NY: Cornell University Press.
- KEOHANE, R. O. (1984) *After Hegemony*. Princeton, NJ: Princeton University Press.
- KIER, E. (1997) *Imagining War: French and British Military Doctrine Between the Wars*. Princeton, NJ: Princeton University Press.
- KINDER, D. R., AND L. M. SANDERS (1996) *Divided by Color: Racial Politics and Democratic Ideals*. Chicago: University of Chicago Press.
- KINDER, D. R., AND D. O. SEARS (1981) Prejudice and Politics: Symbolic Racism Versus Racial Threats to the Good Life. *Journal of Personality and Social Psychology* 40:414–431.
- KNOKE, D. (1994) *Political Networks: The Structural Perspective*. London: Cambridge University Press.
- KRATOCHWIL, F. V., AND J. G. RUGGIE (1986) International Organization: A State of the Art on an Art of the State. *International Organization* 40:753–775.
- KREPS, D. M. (1992) “Corporate Culture and Economic Theory.” In *Perspectives on Positive Political Economy*, edited by J. E. Alt and K. A. Shepsle, pp. 90–143. London: Cambridge University Press.
- KUKLINSKI, J. H., AND N. L. HURLEY (1996) “It’s a Matter of Interpretation.” In *Political Persuasion and Attitude Change*, edited by D. C. Mutz, P. M. Sniderman, and R. Brody, pp. 125–144. Ann Arbor: University of Michigan Press.
- LAKE, D. A., AND R. POWELL (1999) “International Relations: A Strategic-Choice Approach.” In *Strategic Choice and International Relations*, edited by D. A. Lake and R. Powell, pp. 3–38. Princeton, NJ: Princeton University Press.
- LEGRO, J. W. (1995) *Cooperation Under Fire: Anglo-German Restraint During World War II*. Ithaca, NY: Cornell University Press.
- LOVAGLIA, M. J. (1995) Power and Status: Exchange, Attribution, and Expectation States. *Small Group Research* 26:400–426.
- LUMSDAINE, D. (1993) *Moral Vision in International Politics: The Foreign Aid Regime, 1949–1989*. Princeton, NJ: Princeton University Press.
- LUPIA, A. (1998) “Who Can Persuade Whom: How Simple Cues Affect Political Attitudes.” (<http://weber.ucsd.edu/~alupia/AJPS97.html>)
- LUPIA, A., AND M. D. MCCUBBINS (1998) *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge: Cambridge University Press.
- MARTIN, L. L. (1993) “The Rational Choice State of Multilateralism.” In *Multilateralism Matters: The Theory and Praxis of an Institutional Form*, edited by J. G. Ruggie, pp. 91–121. New York: Columbia University Press.
- MARTIN, L. L. (1997) “An Institutional View: International Institutions and State Strategies.” Paper prepared for the Conference on International Order in the 21st Century, McGill University, May 16–18.
- MEYER, J. W., ET AL. (1997) The Structuring of a World Environmental Regime, 1870–1990. *International Organization* 51: 623–651.
- MOHR, L. B. (1996) *The Causes of Human Behavior: Implications for Theory and Methods in the Social Sciences*. Ann Arbor: University of Michigan Press.
- MORAVCSIK, A. (1995) Explaining International Human Rights Regimes: Liberal Theory and Western Europe. *European Journal of International Relations* 1:157–190.
- MORAVCSIK, A. (1997) Taking Preferences Seriously: A Liberal Theory of International Politics. *International Organization* 51:513–553.
- MORGENTHAU, H. (1978) *Politics Among Nations: The Struggle for Power and Peace*, 5th ed. New York: Alfred A. Knopf.
- MORROW, J. D. (1994) *Game Theory for Political Scientists*. Princeton, NJ: Princeton University Press.
- MULLER, H. (1993) “The Internalization of Principles, Norms, and Rules by Governments: The Case of Security Regimes.” In *Regime Theory and International Relations*, edited by V. Rittenberger, pp. 361–388. Oxford: Clarendon Press.
- MUTZ, D. C., P. M. SNIDERMAN, AND R. A. BRODY (1996) “Political Persuasion: The Birth of a Field of Study.” In *Political Persuasion and Attitude Change*, edited by D. C. Mutz, P. M. Sniderman, and R. Brody, pp. 1–14. Ann Arbor: University of Michigan Press.
- NADELMANN, E. A. (1990) Global Prohibition Regimes: The Evolution of Norms in International Society. *International Organization* 44:479–526.
- NAPIER, R. W., AND M. K. GERSHENFELD (1987) *Groups: Theory and Experience*, 4th ed. Boston: Houghton Mifflin.
- NEMETH, C. J. (1987) “Influence Processes, Problem Solving and Creativity.” In *Social Influence: The Ontario Symposium*, vol. 5, edited by M. P. Zanna, J. M. Olson, and C. P. Herman, pp. 237–246. Hillsdale, NJ: Lawrence Erlbaum.

- NESS, G. D., AND S. R. BRECHIN (1988) Bridging the Gap: International Organizations as Organizations. *International Organization* 42:245–273.
- NISBETT, R. E., AND D. COHEN (1996) *Culture of Honor: The Psychology of Violence in the South*. Boulder, CO: Westview Press.
- ONUF, N. (1998) “Constructivism: A User’s Manual.” In *International Relations in a Constructed World*, edited by V. Kubalkova, N. Onuf, and P. Kowert, pp. 58–78. Armonk, NY: M. E. Sharpe.
- PETTY, R. E., D. T. WEGENER, AND L. R. FABRIGAR (1997) Attitudes and Attitude Change. *Annual Review of Psychology* 48:609–647.
- PRICE, R. (1998) Reversing the Gunsights: Transnational Civil Society Targets Landmines. *International Organization* 52:613–644.
- PRICE, R., AND N. TANNENWALD (1996) “Norms and Deterrence: The Nuclear and Chemical Weapons Taboos.” In *The Culture of National Security: Norms and Identity in World Politics*, edited by P. J. Katzenstein, pp. 114–152. New York: Columbia University Press.
- RESENDE-SANTOS, J. (1996) Anarchy and the Emulation of Military Systems: Military Organization and Technology in South America, 1870–1930. *Security Studies* 5:193–260.
- RISSE, T. (1997) “Let’s Talk.” Paper presented to the American Political Science Association Annual Conference, Washington, D.C., August.
- RISSE, T. (2000) Let’s Argue: Communicative Action in World Politics. *International Organization* 54:1–39.
- RISSE, T., AND K. SIKKINK (1999) “The Socialization of International Human Rights Norms into Domestic Practices: Introduction.” In *The Power of Human Rights: International Norms and Domestic Change*, edited by T. Risse and K. Sikkink, pp. 1–38. Cambridge: Cambridge University Press.
- ROGOWSKI, R. (1999) “Institutions as Constraints on Strategic Choice.” In *Strategic Choice and International Relations*, edited by D. A. Lake and R. Powell, pp. 115–136. Princeton, NJ: Princeton University Press.
- RUGGIE, J. G. (1998) *Constructing the World Polity: Essays on International Institutionalization*. London: Routledge Press.
- RULE, B. G., AND G. L. BISANZ (1987) “Goals and Strategies of Persuasion: A Cognitive Schema for Understanding Social Events.” In *Social Influence: The Ontario Symposium*, vol. 5, edited by M. P. Zanna, J. M. Olson, and C. P. Herman, pp. 185–206. Hillsdale, NJ: Lawrence Erlbaum.
- SCHIEFFELIN, B. B., AND E. OCHS, EDs. (1986) *Language Socialization Across Cultures*. Cambridge: Cambridge University Press.
- SHAMBAUGH, G. E. (1997) “Constructivism, Entrepreneurship and the Power to Socialize Rogue States.” Paper presented to the American Political Science Association Annual Meeting, Washington, D.C., August.
- STRYKER, S., AND A. STATHAM (1985) “Symbolic Interaction and Role Theory.” In *The Handbook of Social Psychology*, vol. 1, edited by G. Lindzey and E. Aronson, pp. 311–378. New York: Random House.
- SUSSKIND, L. (1994) *Environmental Diplomacy: Negotiating More Effective Global Agreements*. London: Oxford University Press.
- TURNER, J. C. (1987) *Rediscovering the Social Group*. Oxford: Basil Blackwell.
- VALLEY, K. L., J. MOAG, AND M. H. BAZERMAN (1998) “A Matter of Trust”: Effects of Communication on the Efficiency and Distribution of Outcomes. *Journal of Economic Behavior and Organization* 34:211–238.
- VASQUEZ, J. A. (1993) *The War Puzzle*. Cambridge: Cambridge University Press.
- WALLANDER, C. (1999) *Mortal Friends, Best Enemies: German-Russian Cooperation After the Cold War*. Ithaca, NY: Cornell University Press.
- WALTZ, K. (1979) *Theory of International Relations*. Reading, MA: Addison-Wesley.
- WENDT, A. (1994) Collective Identity Formation and the International State. *American Political Science Review* 88:384–396.
- WILLIAMS, M. C. (1997) “The Institutions of Security.” *Conflict and Cooperation* 32:287–307.
- WU, C. H., AND D. R. SHAFFER (1987) Susceptibility to Persuasive Appeals as a Function of Source Credibility and Prior Experience with Attitudinal Object. *Journal of Personality and Social Psychology* 52:677–688.
- YOUNG, O. (1992) “The Effectiveness of International Institutions: Hard Cases and Critical Variables.” In *Governance Without Government: Order and Change in World Politics*, edited by J. N. Rosenau and E.-O. Czempel, pp.160–194. Cambridge: Cambridge University Press.
- ZIMBARDO, P. G., AND M. R. LEIPPE (1991) *The Psychology of Attitude Change and Social Influence*. New York: McGraw-Hill.