International relations theory and multilateralism: the search for foundations
James A. Caporaso

Why has the concept of multilateralism not played a more prominent role in theories of international relations? The prima facie case for the importance of multilateral activity in the international realm would seem great. The world, we constantly tell ourselves, is increasingly drawn together. The Swedish economist Assar Lindbeck argues that most external effects of production and consumption are external not only to the household but also to the country in which they occur.\(^1\) According to many different indicators, interdependence is on the increase in nearly all parts of the world. International political economists talk about global indivisibilities, ranging from peace to pollution. Most important international problems—including pollution, energy, managing airline traffic, and maintaining rules for trade and investment—intrinsically involve many countries simultaneously. What makes a problem international is that often it cannot be dealt with effectively within the national arena. Costs and benefits spill into the external arena. These external effects are frequently so great that domestic goals cannot be accomplished without coordinated multilateral action.

This article was prepared for the Ford Foundation West Coast Workshop on Multilateralism, organized by John Gerard Ruggie. I am grateful to the Ford Foundation for financial support of the project. I also acknowledge the support of the Virginia and Prentice Bloedel Chair at the University of Washington as well as the Department of International Relations, Research School of Pacific Studies, Australian National University, Canberra. For their comments on earlier drafts of my article, I thank William Drake, Jeffry Frieden, Ronald Jepperson, Robert Keohane, Edgar Kiser, Stephen Krasner, Lisa Martin, George Modelski, Richard Sherman, Janice Thomson, Alexander Wendt, and several anonymous reviewers for International Organization and Columbia University Press.

Multilateralism: ignored by the world or by international relations theory?

The puzzling question motivating this article is “Why is multilateralism neglected in international relations theory?” The question contains an assumption—namely, that the treatment of multilateralism in the scholarly international relations literature is less than would be expected on the basis of its observed importance in the world.

Perhaps this assumption should be questioned. One possible reason for the paucity of theory concerning multilateralism is that there may be so little multilateralism in practice. Advocates of this view might cite the declining importance of the United Nations (UN), at least before 1990, the disaffections with the General Agreement on Tariffs and Trade (GATT), the rise of bilateralism in U. S. trade policy, and the numerous selective arrangements drawn up by the European Community (EC), such as the multifiber arrangements.

One scholar laments that “too many people in both countries [the United States and the Soviet Union] see the problems, and their solutions, as bilateral, overlooking the need for the additional cooperation of Europe in the multilateral world of the late twentieth century.” Another, a former member of the U. S. Council of Economic Advisers, argues that the United States has become “the bully of the world trading system” and that it has done so by “unilaterally redefining ‘unfair’ trade.” Finally, Inis Claude, Jr., reflecting on almost thirty years since the publication of Power and International Relations (1962), today accords much more weight and effectiveness to the balance of power, which he earlier contrasted with collective security and world government, the latter two options being inherently more multilateral. The balance-of-power system is a decentralized system of relations among powers. It acknowledges little if any debt (except for the tradition established by Hedley Bull) to world order or multilateral commitments.

Claude’s assessment is provocative and has at least some validity. He argues that during the last quarter-century, the stock of balance of power (presumably also in its bipolar form) has improved. Decision makers no longer believe that universal solutions are invariably best, that peace is indivisible (Europe has been far more peaceful than the Middle East and Southeast Asia), that local clashes should be met with worldwide responses, or that regional alliances are a bad thing.

Another explanation for inattention to multilateralism lies at the ideational level. In what ways, if any, does contemporary international relations theory

invite or discourage attention to multilateralism? It would be possible, albeit
difficult, for practitioners of liberal economic theory to explain wars, but it
would be incoherent with the theory to ask about state interests, unequal
exchange (in Arghiri Emmanuel’s sense5), and theories of development resting
on absolute immiseration. It would be difficult for a realist to explain extended
cooperation not tied to specific, identifiable gains, but it would not be
impossible.6 It would be impossible to motivate a realist argument about
goodwill, community spirit, or a state’s desire to minimize its power. What
theories make impossible to explain their practitioners of necessity avoid. What
they make merely difficult to explain often preoccupies most of their time.
Between that which is conceptually impossible and that which is natural (ergo
obvious), there lies the region of what is problematic. For realists, the problem
of cooperation is particularly salient. For liberals, why there is ever war,
protection, autarky, and Pareto-inferior behavior is often a puzzle so exotic as
to discourage inquiry.

Since I anticipate misunderstanding on this point, let me clarify my claim. I
am not arguing that multilateral activities and organizations have been ignored.
Indeed, multilateralism as a subject matter has been the grist for volumes. The
UN itself has spawned a large literature. Complex multilateral negotiations,
such as those concerning GATT and the Law of the Sea, have been studied.
And many processes that do not go under the name of multilateralism, such as
regional integration and coordination of economic policies, have occupied our
attention. My point is that multilateralism is not extensively employed as a
theoretical category and that it is rarely used as an explanatory concept.
Indeed, even in cases in which multilateralism provides the central conceptual
focus, cooperation or institutions usually turn out to do the important
theoretical work.

Multilateralism: definition and conceptualization

The terms “multilateralism” and “multilateral” suggest some linguistic consid-
erations. The noun comes in the form of an “ism,” suggesting a belief or
ideology rather than a straightforward state of affairs. Underlying John
Ruggie’s conception of multilateralism is the idea of “an architectural form,” a
deep organizing principle of international life.7 As an organizing principle, the
institution of multilateralism is distinguished from other forms by three
properties: indivisibility, generalized principles of conduct, and diffuse recipro-

5. Arghiri Emmanuel, Unequal Exchange: A Study in the Imperialism of Trade (New York:
6. Indeed, this is precisely what Robert O. Keohane does in After Hegemony: Cooperation and
7. See the following works of John Gerard Ruggie: “Unravelling the World Order: The United
States and the Future of Multilateralism,” mimeograph, University of California, San Diego, 1989;
and “Multilateralism: The Anatomy of an Institution,” in this issue of IO.
ity. These three properties should be treated as a coherent ensemble which is itself indivisible, rather than as additive, detachable indicators of multilateralism. Indivisibility can be thought of as the scope (both geographic and functional) over which costs and benefits are spread, given an action initiated in or among component units. If Germany experiences recession, are there consequences for Germans alone, the French, the members of the EC, or nationals in every corner of the earth? Generalized principles of conduct usually come in the form of norms exhorting general if not universal modes of relating to other states, rather than differentiating relations case-by-case on the basis of individual preferences, situational exigencies, or a priori particularistic grounds. Diffuse reciprocity adjusts the utilitarian lenses for the long view, emphasizing that actors expect to benefit in the long run and over many issues, rather than every time on every issue.

The distinction between multilateral institutions and the institution of multilateralism is cognizant of two levels of related international activity. Multilateral institutions focus attention on the formal organizational elements of international life and are characterized by permanent locations and postal addresses, distinct headquarters, and ongoing staffs and secretariats. The institution of multilateralism may manifest itself in concrete organizations, but its significance cuts more deeply. The institution of multilateralism is grounded in and appeals to the less formal, less codified habits, practices, ideas, and norms of international society. Bilateralism, imperial hierarchy, and multilateralism are alternative conceptions of how the world might be organized; they are not just different types of concrete organization.

There are at least two reasons for maintaining the distinction. The first, as Lisa Martin argues, is that multilateral institutions and the institution of multilateralism do not always mirror one another within a given issue-area. Depending on the structure of interests, one may be strong and the other weak. The second reason is that the two types of multilateralism may be related in complex cause-and-effect ways. Multilateral organizations may provide arenas within which actors learn to alter perceptions of interest and beliefs. The institution of multilateralism may in turn spawn, maintain, alter, and undermine specific organizations. Ernst Haas's study of learning within numerous international organizations, including the International Labour Organization and the World Bank, illustrates the first effect. Jean Monnet's

8. See Ruggie, "Multilateralism."
11. See the following works of Ernst B. Haas: Beyond the Nation-State: Functionalism and International Organization (Stanford, Calif.: Stanford University Press, 1964); and When Knowledge Is Power: Three Models of Change in International Organizations (Berkeley: University of California Press, 1990).
12. Ibid.
insistence that there be a *unité des faits* among numerous actors in international society before concrete regional organizations can take hold illustrates the second.\(^{13}\)

The term “multilateral” can refer to an organizing principle,\(^{14}\) an organization, or simply an activity. Any of the above can be considered multilateral when it involves cooperative activity among many countries. “Multilateralism,” as opposed to “multilateral,” is a belief that activities ought to be organized on a universal (or at least a many-sided) basis for a “relevant” group, such as the group of democracies. It may be a belief both in the existential sense of a claim about how the world works and in the normative sense that things should be done in a particular way. As such, multilateralism is an ideology “designed” to promote multilateral activity. It combines normative principles with advocacy and existential beliefs.

Definitions have implications for the broader conceptual framework and for theory. First, the term “multilateral” does not analytically presuppose any particular number of countries in the way that unilateral, bilateral, trilateral, and universal do. These terms now describe specific points on an underlying continuum from everyone going it alone to everyone participating. Multilateral suggests “many” actors, but it is unspecific as to what number constitutes many. “Many” could refer to anything from a minimum of three to a maximum of all. Multilateral refers to a region, rather than a point, on the continuum and thus can be analyzed in terms of gradations. When conceptualized in this manner, multilateral action is compatible with theories concerning thresholds in groups of less than universal membership.

Second, the definition of the term “multilateral” presumes cooperation. Not all cooperation is multilateral, but all multilateral activities include cooperation. In a Hobbesian war of each against all, we do not say that states behave in a multilateral fashion even though they are interacting in highly interdependent ways. However, it is consistent to think of some states as carrying out multilateral activities against others, such as under the auspices of a military alliance.

The question arises as to whether multilateralism is a means or an end, an instrument or an expression, or both. States or their agents are conceptualized as conscious, goal-seeking actors. As such, it is easy to see how multilateralism is one means among many to be used or ignored according to the instrumental calculus. In instrumental theories—that is, theories stressing given preferences and conscious choice—cooperation has been used to mean a process by which states actively adjust their policies to take into account the preferences of others. In game theory, a cooperative choice (which may be part of an overall strategy) generally represents an effort to take account of the other players’

---

interests, even if the dominant interest is to defect. This makes cooperation a “move,” an instrumental action that takes place within the overall context of a social dilemma.

While instrumental approaches to multilateralism are important, they do not exhaust the possibilities. Multilateral activities may also be an end, or consumption good, according to which states prefer to do things multilaterally. Stated in this way, the noninstrumental version may be hard to defend. How, for example, can one account for the vastly different forms of organization across different sectors by the same states? But there is something in between the ideal types of means and ends. In uncertain environments, states may be forced to make strong presumptions in favor of one or another approach. In this way, multilateralism may not always be thought out and chosen on the basis of exacting calculations of costs and benefits. To the extent that presumptions in favor of multilateralism exist, it becomes part of our ongoing, taken-for-granted understanding of international life.

As Ruggie argues, multilateralism is a demanding organizational form.\textsuperscript{15} It requires its participants to renounce temporary advantages and the temptation to define their interests narrowly in terms of national interests, and it also requires them to forgo ad hoc coalitions and to avoid policies based on situational exigencies and momentary constellations of interests. Yet it is by no means a rare organizational form in the world. This prompts the question, “What coherent theories do we have to account for multilateralism?” Since my approach to answering this question is unconventional, a word of explanation may be in order. My argument is that international relations theory, with few exceptions, does not offer “off-the-shelf” theories to explain multilateralism.\textsuperscript{16} This complicates the task and prompts me to look in part outside of the international relations literature. In what follows, I draw from numerous areas, including sociology, experimental psychology, organization theory, and game theory in an attempt to identify alternative foundations for theories of multilateralism. I identify the elements of three paradigms, pointing out the similarities with and implications for international relations theory.

**Three routes to multilateralism**

In this section, I explore three theoretical routes to understanding multilateral activities. The first route is provided by an individualist paradigm in which states “enter into” contractual relations with other states in a rational, self-interested way. The bare bones equipment of this approach includes states and state interests (both unproblematically given), capabilities, and strategies

\textsuperscript{15} Ibid.

\textsuperscript{16} For creative theoretical suggestions along these lines, see Martin, “Interests, Power, and Multilateralism”; and Robert O. Keohane, “Multilateralism: An Agenda for Research,” *International Journal* 45 (Fall 1990), pp. 731–64.
by which to interact with others. The second route to be explored is found in a loosely connected assemblage of writings labeled the social-communicative approach. The focus of this approach is still on the identities and powers of individual states, but the "interaction repertoires" of states include communication, persuasion, deliberation, and self-reflection. Ontologically, the state is separately constituted, but a bit more sociality is brought into its self-definition.

The exact cutoff point between the first and second approaches is not my concern here. As part of their project, radical individualists may seek to derive the second conception of states from the first and to explain even minimal elements of sociality among states as outgrowths of a model based on nothing more than autonomous, self-interested actors. Similarly, proponents of the second approach may argue that the first works only as long as it imports elements of their model as unexamined premises. They may ask, for example, how self-interested states can contract without some common understandings, rules, and practices. Rather than trying to solve this demarcation issue in the abstract, I pursue the implications of the distinction below.

The third perspective is provided by an institutional approach. Since there are numerous competing "new" institutional approaches, I will try to stake out the ground for the approach dealt with here. Institutionalism has ties to the second approach in its insistence on the importance of communication, reflection, discussion, learning, and interpretation. However, it departs from it in a number of ways. It is not necessarily methodologically individualist, does not treat preferences as exogenous, and does not understand social relations—including multilateral relations—solely as products of individual self-interested calculations. This is a negative way of describing the institutional approach. What the approach stands for in a positive sense will be outlined later.

The individualist paradigm

The individualist label is meant as a shorthand for a collection of theories that attempt to explain social behavior by appealing to characteristics of individual actors, particularly preferences and capabilities, and their strategic environments. The analogues in international relations theory are realism, neorealism, and game theory. Even some theories of international cooperation, institutions, and organization could be included insofar as they theorize these phenomena as lying outside the defining properties of states. This diverse collection is unified by its focus on states as conscious goal-seeking agents pursuing their interests within an external environment characterized by anarchy and the powers of other states. The paradigmatic question is how states pursue their goals given the constraints under which they operate. When goals are interdependent, the question assumes a strategic form: How can one state achieve what it wants, given the preferences and capacities of others?

Within this paradigm, the road to multilateralism logically involves two steps not necessarily temporally sequenced. First, there must be an explanation of
cooperation. Second, the theory must provide the conditions under which cooperation becomes multilateral. The explanation of one is not reducible to the other. The problem of cooperation is a problem precisely because state interests are independently given, often in conflict with one another, and pursued within an environment of anarchy. Furthermore, most of what we take as social (including rules, shared understandings, and collective beliefs) is not part of the representation of the problem of cooperation in either its verbal or game-theoretic form. In the standard version of noncooperative game theory, tur17 agents, preferences, and strategies are represented. Some rules are assumed (for example, if you are a prisoner, you should not grab the warden, thus changing the payoff matrix, or threaten to have your partner killed if he or she defects, thus changing the incentives for cooperation), but this amounts to very little in playing the game. Absent are norms, morality, and the ability to communicate, to promise, and to make binding commitments. To explain why states cooperate in the bilateral case seems difficult enough. To establish cooperation multilaterally is a problem compounded.

I mention these difficulties not to criticize at this point but simply to remind that the individualist approach does not make things easy for itself. tur18 Some approaches assume much of what they want to prove. Outcomes to be explained are then progressively revealed by uncovering the assumptions of the theory. Empirical research gives way to deductive elaborations. That is close to the opposite of what the individualist approach has done. Very little sociality is built into its central premises. If this approach works, it is more powerful in that it explains a complex social activity such as multilateralism without access by assumption to much of that sociality. It is as if, from a world of isolated, self-seeking individuals, complex social structure results. Below, I explore how several specific individualist models help explain multilateral actions.

The k group. Of the three standard solutions to the problem of cooperation—the k group, side-payments, and repetition of the game—the first seems to provide the safest footing for multilateralism. Repetition of the game (Kenneth Oye's "shadow of the future") tur19 may enhance the value of reputation and make it more attractive for participants to think about net benefits in the long run, but it provides little guidance as to how the time frame should affect the comprehensiveness of cooperative actions. Side-payments may help to bring potential defectors in line, but even these incentives would seem to presuppose an already existing pattern of cooperation.

18. I understand that the assumptions I have used to characterize the individualist approach can be and have been relaxed in both international relations theory proper and game theory. My approach here is a device used to see how far this simplified model can take us.
The logic of the $k$ group, however, holds some promise. Given an $n$-person social dilemma, we know that the dominant individual strategies, if played, yield a deficient equilibrium. In the international environment, which arguably models the $n$-person social dilemma, the defect option is better regardless of what the others do. But if all defect, as the familiar story goes, the outcomes are next to last in the preference schedules of the players. Now suppose, to counter this result, that among the $n$ players there exists some subset (called $k$) whose cooperation would ensure resolution of the dilemma, no matter what the others ($n - k$) did. This still would not ensure multilateral cooperation, since defect would still be the dominant strategy, unless individuals could be convinced that each contribution was crucial to resolving the dilemma. Deciding on the members of the subset would be difficult but critical, and some assurance would have to be given that others do not contribute or, more exactly, that $k$ group members did not think that others would contribute.

While $k$ groups do provide some basis for multilateral cooperation, this approach needs a mechanism to identify or designate who the members of the group are so as to ensure that no one else is perceived as a possible contributor and to establish arrangements for enforcing these conditions. There is another, more important complication, however. The smaller the $k$ group, the easier it is to cooperate but the less multilateral the arrangement will be. The larger the $k$ group, the more multilateral the cooperative arrangement might be but the more difficult it is to pull off cooperation. Solutions may be available to get around the problem, but such solutions make assumptions and introduce variables that take us to the other side of my classificatory divide—from individualist, interest-oriented to social-communicative theories.

Repetition of the game. A second approach to multilateral cooperation is to relax the assumption that the game is played only once and to allow for an indefinitely large number of plays, more or less as in real life. If the game is played only once or is played a finite, known number of times, the incentive to defect is dominant. If the game is repeated an unknown number of times or if uncertainty is present in other ways, the possibility of benefits from long-term cooperation must be weighed against the possible gains of a defect move in the short term.20

Lengthening the time frame over which individuals interact has several effects. One is that it opens the door to concepts such as "dynamic cooperation," in which a comprehensive strategy—that is, a series of related moves or a plan for choosing moves over the entire game—rather than a single move is the object of equilibrium thinking. Another is that it allows for variability in the value of reputation and trustworthiness, as well as conditional if implicit

promising and reciprocity. Clearly, the introduction of these concepts allows for a much richer analysis than the one-play games do.

In addition to their increased complexity, iterated games are more supportive of cooperative outcomes. Both Russell Hardin and Michael Taylor demonstrate "the possibility of cooperation," as the title of Taylor’s book cautiously puts it.\(^{21}\) As Robert Keohane argues, "The essential reason for this difference [between single-play and iterated games] is that, in multiple-play Prisoners’ Dilemma, defection is in the long run unrewarding, since the short-run gains thereby obtained will normally be outweighed by the mutual punishment that will ensue over the long run."\(^{22}\)

Robert Axelrod has shown that a mixed strategy of selective cooperation and defection can be a winning strategy.\(^{23}\) He also demonstrated with considerable success that conditional cooperation is a winning strategy and one that is collectively stable—that is, resistant to any invasion by a different strategy. Axelrod’s experiments were carried out in a setting in which one could single out a partner, identify defections and cooperative moves, and bring sanctions or rewards to the appropriate party. The results depended, probably to a considerable extent, on the memory of particular dyads and on the reputations of their constituent members.

Rudolph Schuessler developed a different form of the game, an iterated prisoners’ dilemma with an exit option after each play of the game.\(^{24}\) By constructing the game in this fashion, he was able to explore the conditions of cooperation where there were no memory or reputational effects and where defectors could not be identified, traced, and punished. Instead, if a player felt exploited, he or she could simply exit. Even under these spare conditions, closer to an anonymous *Gesellschaft* rather than *Gemeinschaft*, Schuessler was able to demonstrate considerable cooperation.

The problems begin when we ask how this cooperative behavior of dyads generalizes to become multilateral. Both Axelrod’s and Schuessler’s analyses were of dyadic interactions within a large population. Many individuals were involved and they were all playing a prisoners’ dilemma game, but this was not the same thing as an \(n\)-person prisoners’ dilemma game. In the language of this article, it was aggregate bilateralism rather than multilateralism. Axelrod’s and Schuessler’s players were cooperating with specific relevant others, not with the group.

By contrast, in *The Possibility of Cooperation*, Taylor focused on \(n\)-person games, particularly prisoners’ dilemma. He developed a supergame consisting

---


of "a countably infinite number of iterations of a single constituent game" in which a player's payoffs were a function of his or her strategy (cooperate or defect) in the constituent game and of the number of other players who cooperated in that game. Here, the incentives to cooperate or defect were in part (perhaps in large part) determined by an unspecified number of other players with whom a given player may or may not have been in direct contact. Taylor's conclusion was that "Cooperation amongst a relatively large number of players is 'less likely' to occur than Cooperation amongst a small number." Part of the reason has to do with the obvious fact that since more actors exist, more interests are involved and have to be taken into account. But the main reason, according to Taylor, is that "Cooperation can be sustained only if conditional Cooperators are present and conditional Cooperators must be able to monitor the behavior of others." The difficulties and costs of monitoring, which I discuss below, go up as the number of actors goes up.

Transaction costs, multilateralism, and regimes. Transaction costs are all the costs incurred in exchange, including the costs of acquiring information, bargaining, and enforcement, as well as the opportunity cost of the time allocated to these activities. In a sense, the idea of transaction costs takes us beyond noncooperative games. In the standard version of these games, agents, payoffs, and rules are given by assumption. These parameters are not problematized and do not function as costs of playing the game. In addition, the idea of enforcement costs is specifically excluded in noncooperative games. Nevertheless, we can think of transaction costs in a thinner sense as the costs of discovering relevant agents and their preferences, of negotiating, of identifying defection or cooperation, and of bringing rewards and penalties to bear on the relevant parties.

What is the relationship between the number of actors involved in a potential multilateral scheme and the costs of transacting? The costs of transacting almost certainly increase with an increase in actors. The costs of identifying the relevant others, of discovering their preferences and strategies, and of devising policies that are capable of discriminating among defectors and cooperators all go up. Just how or by what function this increase takes place is difficult to say. A linear relation suggests that complications are directly proportional to the number of actors involved. An exponential relation suggests that costs rise in reaction to the number of pairs or n-tuples involved. Another possibility is that while costs rise, there are dampening factors as n increases. Taylor suggests

26. Ibid., p. 105.
27. Ibid.
29. Here, I am speaking of rewards and penalties resulting from different moves within the game, not from resources outside the game. The latter would amount to an enforcement mechanism.
that a critical concept for cooperation in large groups is "conditional cooperation," defined as the strategy of cooperating on the condition that others cooperate. Conditional cooperators have to monitor the moves of others; otherwise, their own moves cannot be conditional. However, as Taylor points out, conditional cooperators need to know only that a certain fraction of others cooperated in a prior move, not which ones in particular cooperated. While this may decrease information costs, it does not eliminate them.

What relationship might we expect between multilateral organizations and transaction costs? Ruggie points out that sometimes institutions are multilateral because to organize them differently would be more costly. If the costs are large for each pair of states and if the activities generating these costs are essentially the same (that is, repetitive), large gains may be expected through multilateral organization. Some of these gains may be Smithian, deriving from the same advantages that make factory production more efficient than individual production, advantages such as decreasing the amount of time and resources wasted in moving between different points of production. Keohane refers to political economies of scale. The marginal costs of handling an additional agreement (that is, another bilateral agreement) within a multilateral organization will be lower than the average cost.

But multilateral activity without an organization to facilitate and enforce agreements brings up all of the problems that haunt international political cooperation in the first place: absence of trust, weak and unreliable information, incentives to defect, and reneging on agreements when it is convenient. While transaction costs may be lower for multilateral than for bilateral agreements, the arrangements in this case may still be bilateral. Such nonoptimal outcomes are not rare in international politics.

One proposed way around this problem is provided by the literature on international regimes. In After Hegemony, Keohane makes the case for regimes lowering transaction costs. At the simplest level, they make it cheaper for governments to negotiate agreements. They provide administrative help, an ongoing forum in which representatives of different states can meet, and a set of rules and procedures for dealing with problems. In addition, regimes provide valuable informational services that facilitate multilateral contact. They collect information, standardize conceptual categories (defining the characteristics of a nontariff barrier, for example), codify rules and practices, and attempt to increase the transparency of both cooperative and defecting moves. Although GATT is a relatively small international organization, it is hard to imagine how individual states would manage their global trade relations bilaterally. Finally, since regimes provide a framework of rules, a continuing organization in which

31. Ibid.
33. Keohane, After Hegemony, p. 90.
34. Ibid.
states come back for more every day, they facilitate sanctioning. Even if the actual rewards and penalties are carried out by the states themselves, regimes make the whole idea of a “violation of rules” intelligible and meaningful in some intersubjective sense.

The idea of transaction costs provides a bridge of sorts between the individualist approach and the social-communicative approach. As noted above, transaction costs are all the costs associated with pulling off exchanges, costs that are usually assumed to be zero in the perfectly competitive ideal. The concept of transaction costs contains within it much material that is anterior to and outside of exchange and strategic interaction per se. Insofar as this material has to do with the making and implementation of binding contracts, the transaction costs approach overlaps with other approaches.

In a world of zero transaction costs, there would be no need for international institutions, multilateral or otherwise. It is the recognition of positive transaction costs that leads us to inquire about the role of institutions in solving problems of exchange. In this regard, Lisa Martin asks a straightforward but productive question: “What is the instrumental value of multilateral norms under different configurations of state interest?” Taking only the interests of states as primary data, she explores what implications these strategic considerations have for multilateral institutions and the institution of multilateralism.

While Martin focuses on several types of strategic structures, the contrast between coordination and collaboration games is instructive. In coordination games, actors have a strong incentive to reach an agreement and do not have an incentive to depart from it once it has been reached. The constellation of interests given by coordination games does not require a strong, centralized, multilateral institution to enforce solutions, since the temptation for opportunistnic behavior is small to nonexistent. By contrast, in collaboration games such as prisoners’ dilemma, there is a strong incentive to defect and a consequent need for monitoring and enforcement mechanisms to prevent unilateral defection. But as Martin argues, multilateral norms are likely to be dysfunctional for collaboration problems. This is because the norms of diffuse reciprocity and indivisibility require unconditional cooperation and may discourage the specific detection and selective punishment required by tit-for-tat strategies. For this reason, opportunistic behavior and free riding can be more effectively dealt with through bilateral arrangements than through multilateral ones.

In broad terms, then, Martin argues that there is a tension between the highly generalized norms of multilateralism on the one hand and the specific knowledge and actions required to enforce complex agreements on the other. Beth Yarbrough and Robert Yarbrough also exploit this tension by arguing that bilateralism can be seen as an enforcement mechanism for dealing with the

35. Martin, “Interests, Power, and Multilateralism.”
36. Ibid.
relentless opportunistic pressures inherent in prisoners' dilemma structures. Bilateralism implies that individual countries bear responsibility for detecting violations of agreements and for imposing sanctions. Imposing sanctions, in turn, requires some capacity to inflict costs for noncooperation. Yarbrough and Yarbrough suggest that the exchange of economic hostages may provide the basis for imposing penalties. "Economic hostages" are defined as binding precommitments to invest in assets that are specific to a transaction and to a partner. Japanese automobiles fitted with pollution devices required by U. S. regulations are one example.

While the theoretical point here is well taken, I am skeptical about generalizing the conditions of economic exchange to multilateral settings. First, we should not underestimate the difficulty of hostage exchange in political terms. A great deal of intragovernmental coordination and power would be required to tailor regulatory policy to the specifications of foreign trade. This suggests a world in which "domestic" regulatory policy is driven by the desire to set up sanctioning mechanisms in the event of the breakdown of foreign exchange. Second, we can question the hostage argument on its own terms. To discourage opportunism, the link between countries must show high asset specificity, since this condition indicates high opportunity costs. But if this condition exists, by definition it works against multilateral generalization. A country cannot have a different regulatory environment in the same issue-area for each partner. By contrast, if assets are standardized (the condition suggested by policy convergence), the costs of shifting to another partner will be low.

A way out of this dilemma might be to tailor regulations to each partner in specific (different) issue-areas. This procedure faces numerous hurdles. Surely regulatory policies will overlap and interfere with one another, limiting the degrees of freedom for policymakers. Moreover, specific assets are sunk costs, which slow adjustment to ever-shifting comparative advantage.

In this section, I have examined the path to multilateral cooperation provided by approaches based on individualist, state-centered theory. Since these approaches build very little sociality into their premises, they are of limited value in explaining multilateral cooperation. The sparsest version of this theory draws on noncooperative games in which the rules, the agents, and the preferences, beliefs, and choices of agents are established by assumption. In the one-shot dilemma game, defection is a dominant strategy. By selectively relaxing the assumptions, complexity and realism can be introduced. The game can be repeated so as to allow signaling (a form of communication), interdependent strategies (such as tit-for-tat), and various values for reputation. Cooperative results can be and are produced. Indeed, some argue that they are robust.

38. See, for example, Schuessler, "Exit Threats and Cooperation Under Anonymity."
Nevertheless, extensive multilateral cooperation is fragile and compounded by increasing the numbers of players. Iterated games imply cooperation but not necessarily multilateral cooperation. As the number of actors increases and power diffuses, the size of the solution set can change.\(^3\) Further, as John Orbell and his colleagues show, cooperative solutions are often behaviorally dependent on a consensus far in excess of the level of agreement rationally needed to produce optimal results.\(^4\) While results could be produced by a \(k\) group (a subset of \(n\)), how does a \(k\) group know that it is one and who should be a member of it? The existence of a minimal contributing set (a small group capable of providing a public good) is not a mechanical question relating to the distribution of capabilities. It requires consciousness about the common situation and shared interests as well as positive actions regarding sharing information and forming coalitions. Moreover, as Taylor emphasizes, even if a \(k\) group exists it is likely to be the case that it is in the interest of some members to get others to join in place of themselves: "In those cases where subsets of the players find it collectively worthwhile to provide the public good, there arises a quite different strategic problem, which results from some players having an incentive to ensure that the subset which provides the public good does not include themselves."\(^1\) Thus, problems of free riding and strategic preferences do not disappear just because a privileged group exists.

The idea of transaction costs, applied in conjunction with regimes, perhaps takes us further toward understanding multilateral institutions. However, it also takes us further away from approaches based purely on interest and strategic choice and closer to those based on communication and making binding commitments.

**Social-communicative approaches**

In making the transition from individualist to social-communicative approaches, we move from representation of autonomous agents engaged in strategic interaction to a view that progressively incorporates social structure and communication. The change is not qualitative but one of degree.

The social approach does not throw out individual rationality; it situates it and broadens it. Individual intentionality is embedded in social relations in which communication, shared beliefs, norms, and identity commitments are present. Thus, the focus is not only on individual choice but on how the choosing agent reflects, discusses, trusts and distrusts, tries to build consensus, alters others' perceptions of the world, and, in general, uses his or her capacities as a social being to identify problems, solve them, and shape the


\(^1\) Taylor, *The Possibility of Cooperation*, p. 82.
environment. With respect to Jon Elster’s metaphors of market and forum, this approach is closer to the forum.\textsuperscript{42} With respect to Anatol Rapoport’s distinction between games and debates, it is closer to debates.\textsuperscript{43} Within game theory proper, it draws more on cooperative than noncooperative game theory. In the remainder of this section, I explore several routes to multilateralism that rely on the social-communicative approach.

\textit{Discussion and persuasion.} In noncooperative game theory, whether a Nash equilibrium is reached can depend to a considerable extent on the ability to coordinate policies. One way for this coordination to come about is through preplay communication. Often it is the case that an actor wants the opponent to play his or her best move in response to the actor’s own move.\textsuperscript{44}

In experimental social psychology, studies concerning how actors deal with social dilemmas have demonstrated that group discussion greatly increases the incidence of cooperation. It does so without any help from a leviathan or central enforcement agency, iteration of the game, side-payments, or other narrowly self-interested reasons. Indeed, some scholars argue, discussion will have an effect on cooperation despite the fact that the cooperative choice is strictly dominated by the defect choice.\textsuperscript{45}

The exact mechanism by which discussion works is not clear. By itself it is no panacea. Many discussions and negotiations in international relations produce no agreements at all, and sometimes it is not even possible to get to the negotiating table. If the structure of interests is “deadlock,” no amount of discussion alone will produce cooperation; one side must effectively persuade the other to see interests differently. Theoretically, discussion may work to alter preferences, to create a feeling of shared identity, to encourage norms, or to facilitate promising behavior.

How important is discussion in international relations theory? According to the predominant model of strategic interaction in noncooperative games, states act on their interests given the interests and strategy choices of others. The emphasis is on arriving at the best outcomes, possibly one from which there will be no rational incentive to depart. Since not all outcomes have this property, attention is paid to devising mechanisms of monitoring and verification to prevent “cheating” or “shirking.” This orientation suggests little role for discussion.


\textsuperscript{44} See Joseph Farrell, “Communication, Coordination and Nash Equilibrium,” \textit{Economic Letters}, vol. 27, 1988, p. 209. If there is only one Nash equilibrium, the equilibrium outcome might result without communication. I am indebted to Richard Sherman for this point.

In the descriptive world of international relations, a plausible case can be made for the importance of discussion. For example, Jeffrey Frankel and Katherine Rockett suggest that the ability of discussion to improve cooperative outcomes among decision makers depends on its ability to produce consensus on the “true” model of economic reality.\(^46\) However, they emphasize that the “assumption that policy makers agree on the true model has little, if any, empirical basis.”\(^47\) Still others, such as Robert Putnam and C. Randall Henning, see discussion as quite central.\(^48\)

In their case study of the 1978 Bonn summit meeting, Putnam and Henning analyzed the economic policy preferences and interests of seven advanced capitalist countries (Britain, Canada, France, Germany, Italy, Japan, and the United States) and found that the policies adopted were different as a result of negotiations.\(^49\) The core of the Bonn accord involved an agreement on the part of Germany and Japan to reflate their domestic economies in return for the U. S. commitment to raise domestic oil prices to world levels.\(^50\) As the authors pointed out, the interests of the states differed, with each having its own target levels regarding outputs, employment, fiscal policy, interest rates, and energy policy. Moreover, the decision makers’ beliefs about which kinds of policies were likely to lead to targeted goals differed markedly. Sometimes these differences had to do with variations in the results of forecasting models. A West German model, for example, predicted growth for 1977 in the 4.5 to 5.5 percent range, a much stronger performance than that forecasted by the Organization for Economic Cooperation and Development (OECD). Sometimes these differences stemmed from theoretical disputes about important parameters in macroeconomic models. How much growth would result from a tax cut of quantity \(x\) or a money supply increase of size \(y\)? How much unemployment would result from decreasing inflation by a given amount?

The extensive case material provided by Putnam and Henning documents in detail the importance of deliberation, negotiation, and attempts to change the beliefs of others about the macroeconomic environment. Interests, power, and strategic interaction were not irrelevant, but the ability to change perceptions of interest using technical knowledge was important as well: “In practice, negotiation (especially internationally) is more than mere bargaining over a fixed, known payoff matrix. Much of what actually happens is attempted persuasion. Behaviorally at least, our account of the events of 1977–78 suggests that it is highly misleading to disregard the degree to which policy makers,

---


\(^{47}\) Ibid., p. 318.


\(^{49}\) Ibid., p. 13.

\(^{50}\) Ibid., pp. 18–19.
negotiating internationally, actually try to convince one another that their respective models of the world and even their respective preferences are mistaken. Under uncertainty, international communications and persuasion can change minds, move the undecided, and hearten those in the minority, domestically speaking.”

Minimal contributing sets. A minimal contributing set (MCS) is defined as the smallest number of actors who together could provide a public good if they were willing to do so. As the earlier discussion of \( k \) groups indicated, for some subset of \( n \) (all actors), total benefits exceed total costs, regardless of what the others do. While the MCS is formally identical to the \( k \) group, the literature suggests that there are somewhat different mechanisms to bring these minimal groups into existence.

Within the international relations literature, much attention has been directed toward identifying the conditions that make \( k \) groups possible. The investigations have pointed scholars in the direction of structural properties of the international system, properties such as the number of important actors and the concentration either of aggregate power or sectoral power. Thus, shares of world trade, investment, monetary reserves, and energy supplies have functioned as indicators of hegemony or of the concentration of power in the hands of a few states.

Proponents of hegemonic stability theory do not claim that the concentration of power automatically translates into the provision of public goods by the hegemon. The existence of a hegemon may not even be necessary, let alone sufficient. At most, they claim that hegemonic distributions of power are conducive to the provision of international public goods. The stability of the Bretton Woods system from 1945 to 1971 was facilitated by the overwhelming economic power of the United States, just as the dominance of the gold standard was associated with Britain’s hegemony during the second half of the nineteenth century.

It is in this context that the experimental literature on MCSs is relevant. It attempts to provide some of the missing links between aggregate structural facts such as the international distribution of power and process-level phenomena. Among other things, the literature focuses on discussion, the dissemination of information (for example, about the identity of members of the MCS), and optimal levels of provision. For public goods with specific threshold provision points, it is important to identify the relevant actors and the required amount of contribution. It may also be important to convince MCS members that each of their contributions is essential; otherwise, there would be a temptation to free ride.

51. Ibid., p. 110.
52. Keohane, After Hegemony, p. 31.
In the experiments set up by Alphons van de Kragt and his associates, subjects were placed in a social dilemma and given a choice between cooperating (contributing) and defecting (free riding). If they cooperated, the group would be awarded a bonus to be shared by all (a public good). If they did not cooperate, there was no guarantee of a bonus. If the bonus were given, however, they would share in the benefits at no cost. Subjects were divided into twenty-four groups, half of which allowed discussion and half of which did not. In all twelve discussion groups, the members contributed to the public good; and in ten of these discussion groups, they contributed the optimal amount.

According to the designers of this experiment, the most important factor is "criticalness." Each member of the group is motivated to contribute when the contribution of each is seen as a sine qua non for provision of the public good. Individuals are placed in the same relation to goods as they are in private markets: either pay and receive or else pay nothing and receive nothing. If members can be convinced that some others will not contribute (thus falling short of the minimal set), their contributions are in effect equilibrium moves from which there is no rational incentive to depart.

While the structural problem addressed by MCS is the same as that addressed by the k group, the resolution for the latter is considerably different. Emphasis is placed on discussion and explicit coordination of interests. The ability to designate explicitly who is and is not a member of the k group is critical. The logic of the solution is limited to public goods with explicit provision points where participants can be brought into contact with one another, discuss, and designate contributors.

Norms, promising, and group identity. In exploring the effect of norms, promising behavior, and group identity on cooperation in social dilemmas, I discuss the results of laboratory experiments constructed to show the social dimensions of decision making in n-person dilemmas. The general form of the experiments is that of a social dilemma in which the set of individually rational moves not to cooperate (not to contribute to the public good) is dominant and therefore produces a deficient equilibrium. I should add that the decision situations are anonymous and limited to one time frame (no iteration); therefore, tit-for-tat strategies are not permitted, and the value of reputation is zero. Although the processes leading up to decisions are social, the actual decisions are completely private. While norms are ultimately dismissed as a key mechanism explaining cooperation in these experiments, group identity and promising are not.

55. Ibid., p. 114.
56. Ibid., p. 116.
57. See Orbell, Dawes, and van de Kragt, "The Limits of Multilateral Promising."
The first experiment was designed by Robyn Dawes, Alphons van de Kragt, and John Orbell to determine whether group identity independently affects cooperation, apart from self-interest, reputation, mutual altruism, or even conscience.58 The subjects were divided into various discussion groups and were told that they could make a contribution to their own group if they wished but not to other groups. The researchers found that individuals gave to their group even when it was not in their self-interest to do so, but they adopted no generalized norm about giving: “Our experiments have led us to conclude that cooperation rates can be radically affected by one factor in particular that is independent of the consequences for the choosing individual. That factor is group identity. Such identity—or solidarity—can be established and consequently enhance cooperation responding in the absence of any expectation of future reciprocity, or current reward or punishment, or even reputational consequences among other group members; moreover, this identity operates independently of the dictates of conscience.”59

In another experiment, the same researchers attempted to extend their analysis of cooperation by exploring the psychological and group dynamics involved in promise-making and promise-keeping behavior. Starting from a baseline of bilateral promising, they attempted to generalize the logic to multilateral promising in groups of up to fourteen subjects. It turned out that the generalization was not at all straightforward. What they found was that promises in large groups were kept when all group members made them but were much less binding when only some members made them.60

The central question is how to make sense of the fact that the investigators found no evidence that promising increases cooperation in cases in which promises were not made universally. A commonsense approach to promising and group identity is to think of the two phenomena as continuously related to cooperative behavior. A little promising or group identity is better than none; a lot is better than a little; and so on. But this commonsense approach does not square with the step-level function observed in the experimental results. Why is unanimity so important, even when the supply of public goods is continuous or when the provision threshold is short of unanimity?

A number of hypotheses come to mind. First, it may be that individuals believe that cooperation will only occur when everyone promises. The hypothesis leaves this belief itself unexplained, but it is suggestive. The fear of an agreement unraveling might highlight the importance of unanimity (“one for all, all for one”; “united we stand, divided we fall”). In addition, unanimity provides a prominent solution in Thomas Schelling’s sense—that is, a point which can be singled out and around which expectations can converge.61

59. Ibid., p. 86.
60. Orbell, Dawes, and van de Kragt, “The Limits of Multilateral Promising.”
Second, group identity may somehow be triggered by unanimous consent. If everyone agrees, feelings of solidarity among group members are likely to be high. Still, it is hard to accept that a few noncooperative individuals could have so disproportionate an effect. Third, and this is not completely distinguishable from the first hypothesis, universality in promising may simply describe the condition under which individuals perceive their own promises as binding. This amounts to saying that a promise is an offer of cooperation contingent on everyone else also cooperating.

While these interpretations are currently unresolved, let us explore some of the differences between bilateral and multilateral promises. In the bilateral case, the payoffs present clear information about the benefits and costs of different choices. If one of the two actors refuses to make a promise, this yields an outcome of "both defect." In the multilateral case, however, there will almost always be levels of cooperation that are less than universal but nonetheless yield payoffs better than at the point where all defect.

This point is illustrated in Figure 1, which is adapted from the work of Orbell, Dawes, and van de Kragt. There are many points along the $x$ axis

---

**FIGURE 1. The structure of incentives in multilateral exchange**


---

62. Orbell, Dawes, and van de Kragt, "The Limits of Multilateral Promising."
(number of cooperators) where the gains from cooperation (y axis) equal and then exceed the gains from all defecting. This is true even if cooperation is dominated by defection at the individual level (the defect line has the same slope but a higher intercept than the cooperate line). Any point to the right of k is socially superior to any point to the left of it. Yet the entire region between k and n - 1 on the horizontal axis describes a span where cooperation is not forthcoming as a result of promising.63

Orbell and his colleagues tentatively conclude that perhaps the answer lies within the realm of uncertainty, cognitive limitations, and the difficulty of specifying a prominent rule, one that is “simple enough to be generally recognized, and general enough to apply to the many n–prisoners’ dilemma game parameters that are possible.”64 The economically attractive rule “Your promise is ethically binding when any number more than k (including yourself) have also promised” would do nicely, but it may be beyond the cognitive competence of the actors. If so, we can see why members of the group might push toward universality: it eliminates free riding, diminishes the obstacles that envy and relative deprivation might throw up, and is equitable. However, a multilateral cooperative arrangement that has no mechanisms for dealing with local defections would be extremely fragile.

**Institutional approaches**

To use the term “institutional” today is to evoke a host of competing conceptions about what institutions are and how they function.66 First, there is the old (but still alive) institutional economics of John Commons, Clarence Ayers, and Thorstein Veblen.67 Second, there is the much more recent “new institutionalism” of neoclassical economics, sometimes called “the new economics of organization.” This approach, which highlights principal-agent problems

63. This does not mean that cooperation does not occur in this region. It just means that there are no significant differences in cooperation across groups that promise (short of universal promising) and groups that do not.
64. Orbell, Dawes, and van de Kragt, “The Limits of Multilateral Promising,” p. 265.
65. Ibid.
67. For a description of this approach, see the special issue of *Journal of Economic Issues*, vol. 21, no. 3, September 1987.
and focuses on transaction costs, is discussed in the writings of Douglass North and Oliver Williamson. Third, there is the "new institutionalism" of political science described by James March and Johan Olsen. Proponents of this third approach argue for the relative autonomy of political institutions and the importance of beliefs and norms.

The three institutional approaches are not carbon copies or modified versions of one another. They often differ in philosophical assumptions, core theoretical concerns, research programs, and method. The institutionalism of Commons, Ayers, and Veblen looks to Charles Pearce for his theory of semiotics and to German historicists such as Wilhelm Dilthey. The new economics of organization, tracing its roots to Ronald Coase's 1937 article entitled "The Nature of the Firm," interprets economic institutions as reducing transaction costs involved in certain types of market exchange. The institutionalism of March and Olsen reclaims parts of a much older political science tradition while enriching it theoretically through insights drawn from organization theory and the work of recent social scientists such as Theda Skocpol, Stephen Krasner, and Peter Katzenstein.

I will first describe the major components of the institutional approach employed here and then try to draw out some implications for multilateralism. What I describe borrows selectively from the three approaches discussed above, but the final version is not reducible to any one of them.

The starting point for wanting to develop an institutional paradigm is a dissatisfaction with rational choice, pluralist, and international systemic approaches to international relations theory. Individualistic rational choice approaches focus on individuals attempting to maximize their utility functions. Pluralist approaches view policies and behavior as results of the pulling and hauling of pressure groups, and international systemic approaches explain the range of state options in terms of variations in international structure. All of


69. See March and Olsen, "The New Institutionalism."


these approaches are valuable in their own right yet are inadequate for understanding multilateral cooperation. Most choice-theoretic approaches take preferences and rules as given, neglecting the ways in which institutions shape preferences. International systemic approaches help us understand how states overcome collective action problems. They tell us, given a particular constellation of interests, whether a cooperative solution is likely. But they tell us little about the content of that solution. Pluralism, in its traditional versions at least, provides a good account of the social forces underlying decision making, but it treats political institutions as transmission belts conveying demands through the policy process rather than as independent variables in their own right.

These dissatisfactions set the stage for the institutional approach outlined here. I argue that there are three crucial components of the approach. The first is ontological and relates to the status of entities, particularly individual agents and institutions. The second is theoretical and is concerned with the specification of the proper relations among preferences, institutions, norms, and ideas. The third is interpretive and has to do with how we are to understand cooperation.

Ontologically, the institutional approach may accept elements of methodological holism. While the debate regarding whether groups (or individuals) are “real” is singularly unproductive, the issue of which levels of organized complexity “count” is not unimportant. The institutional approach assumes that enduring structures and patterns of rule are important. It seeks to explain individual behavior by reference to “institutional facts” rather than to characteristics of individuals per se. This statement often prompts confusion, since it suggests to some that institutional explanations involve an overriding of rational choice (or intentional) explanations by anonymous institutional forces. I argue that this is not the proper way to understand the relationship between “micro” and “macro” in institutional theory. Institutional theory is not

74. This is not a trivial point. It is one that is passed over too quickly in the collective action literature. This literature, of which hegemonic stability theory is one expression, generally assumes some form of mixed-motive game in which outcomes (payoffs) can be improved through coordination of behavior. That is, it assumes the existence of some contingent pairs of strategies which, if played, will yield better outcomes than the noncoordinated solution would provide. If the structure of interests is zero-sum, a hegemonic distribution of power would have markedly different consequences.


76. Institutional approaches need not be methodologically holist, however. They may be committed to the view that individuals are the ultimate units of society and still treat complex social structures and institutions as describing important emergent effects. In this sense, the ontological argument may be misleading in that it is easy to glide from the ontological position that individuals are the ultimate units to the theoretical position that all causation has an individual locus. I am indebted to discussions with Ronald Jepperson for this point; see his article entitled “Institutions, Institutional Effects, and Institutionalism,” in Powell and DiMaggio, The New Institutionalism in Organizational Analysis, pp. 143–63.
primarily about nonrational, habit-driven, unthinking behavior. It is mostly about quite purposeful, goal-oriented behavior in which the identities, preferences, beliefs, and behavior of microunits are given a structural determination.

The institutional approach attaches primacy to institutions, norms, and roles that shape and constrain, as well as facilitate, certain kinds of activity. While the individual enters the world with numerous options, the structure of choice and certain persistent configurations of choice are part of the "given" environment. The individual may provide the microcomponent of institutional theory, but social relations and institutions are not seen as products of freely choosing individuals; instead, agency is given a structural determination.

The strong version of this structural determination argument is that individual actions merely "instantiate" the reproduction of structure. The subjects of structural analysis are particular roles and locations that are "occupied" and "executed" by agents. Structures are deemed to be the appropriate focus because they are more durable than individuals and because they preserve consistent properties in the face of greater variety among their occupants. A less extreme version of the structural determination argument is that structures "loosely determine," "condition," or "establish boundaries" within which actions transpire. Individual agents have some room to maneuver, some space within which to exercise choice, but the boundaries are controlled by underlying structures. This less extreme version allows for a compromise between the view that institutions are by-products of individual utilitarian calculations and the opposite view that normatively oriented individual agents mindlessly reproduce structures. Institutions are treated as a level of organized complexity: they are distinct from the sum of individuals composing them, yet they rest on a basis of human actions that are continually contested, are only partly propelled by norms and role expectations, and always reflect a tension between the desires of individuals and the needs of institutions. An institution is no more unproblematic than an autonomous individual. As Ruggie argues, the "collective situation" must continually be "negotiated by the parties concerned."

77. In "International Institutions," p. 384, Keohane stresses that institutions constrain activity, shape expectations, and prescribe roles.

78. For this viewpoint, see Anthony Giddens, Central Problems in Social Theory (Berkeley: University of California Press, 1979), pp. 49-95.

79. The phrase "the needs of institutions" may raise some eyebrows, especially for methodological individualists. What I have in mind here is not some refied entity with motives similar to individuals. I am simply thinking of types of organized complexity that may be quite important for a collectivity even though they are unimportant or perhaps damaging for the individual. Complex divisions of labor may be alienating for the individual but crucial for societal survival. Hierarchy may be undesirable for everyone in the firm yet necessary for competitive profit levels. The theoretical source supporting this viewpoint is evolutionary theory where the forces operating on institutions are selective pressures in the environment. Thus, institutional decline, survival, and change can be seen as results of a blind trial-and-error model with selective winnowing by environmental forces.

The second major claim of institutionalism concerns the relationships among preferences, norms, and beliefs. The approach requires us to rethink the conventional relationships. We can begin by viewing the question of relationships as a type of causal question: Which variable is most important? The basic conceptual issues are then resolved. What remains is to estimate the parameters and determine the share of the variance attributable to each factor. It is important to recognize that both preferences and norms can function as motives, with beliefs supplying information about how to relate means to ends. Both neoclassical economics and realist international relations theory approach their explanatory questions in similar ways. Agents (individuals, states) have preferences (or national interests) and act on them with resources (endowments, national capabilities) at hand. The economic equation is more complete, since it deals with technology (which concerns ideas about combining factor inputs) and offers some insight into decision making in the face of risk and uncertainty. Realist international relations theory provides almost nothing in terms of rigorous theory and research about beliefs. However, international relations scholars outside the realist school are contributing to our knowledge of how ideas and beliefs affect outcomes.

For the kind of institutionalism proposed here, the relative importance of norms, preferences, and beliefs is of less concern than the question of which of these factors are taken as exogenous and which are taken as endogenous. Conventional rational choice models in neoclassical economics start with exogenous preferences, a given distribution of endowments, and a given technology. The behavior of agents is explained by showing that they are responsive to changes in costs and benefits at the margin. In the institutional approach, however, norms, beliefs, and rules occupy a more central position. Individuals come to politics not only with preferences for particular outcomes but also with shared and divisive values and variously developed beliefs about the political process. In addition, politics and individual preferences undergo change: not only do individuals "act out" their preferences politically, but as March and Olsen point out, the political process is a forum within which their

81. In referring to realist international relations theory here, I am not including the literature on psychological images, cognitive structures and perception, and perceptual distortion in crisis decision making. Without devaluing this literature, it strikes me that its purpose is more to identify the sources of perceptual distortion than to argue how better means-ends knowledge or alternative interpretive models can facilitate cooperative outcomes.

preferences and beliefs change "as in the rest of life, through a combination of
education, indoctrination, and experiences." 83

Institutions are thought to be important with respect to preferences, beliefs,
and norms in numerous ways. They help shape preferences by changing the
payoff matrix (for example, making it easier to punish free riders and defectors
within an institutional context) and by offering an environment in which
socialization and learning can occur. The continuous contact, exchange of
information, and education about various cognitive styles that take place within
multilateral institutions serve to alter perceptions about both the content and
the means to achieve private interests. In addition to altering preferences,
institutions provide information, increase trust, and reduce uncertainty about
the actions of others. By so doing, they increase the capacity of separate agents
to achieve interdependent goals by easing coordination problems. Finally,
institutions promote the adherence to norms. In the institutional approach,
norms are not treated as utilities (or as second-order utilities) but as
prescriptions lying outside of preference structures. The content of norms has
less to do with what agents want concretely than with how they ought to behave
in certain situations and what goals they ought to pursue. The UN prohibition
on the use of force to acquire territory, people, or resources is a norm. It
proscribes the use of force for these ends, even though such ends are
represented in the preference schedules of some decision makers.

The third major component of the institutional approach has to do with the
proper way to understand cooperation. According to individualist theories,
such as neorealism, cooperation is as an outgrowth of individual desires,
capacities, and choices. Cooperation is instrumental. While the social atoms of
neoclassical economics are individuals, the primitive social units of realism and
neorealism are states. These states and their identities and ends are instrument-
tally defined as separate from the institutions of international society. As
Alexander Wendt and Raymond Duvall point out, this has important conse-
quences: "The predominance of neorealism in contemporary international
relations theory, with its focus on the conscious choices of self-interested
actors, has meant that the story of international institutions and order is
currently being told overwhelmingly in terms of the strategic problem of
constructing 'cooperation under anarchy' rather than in terms of an interna-
tional society, albeit an essentially anarchical one." 84

Institutional theorists see complex patterns of cooperation already embed-
ded within states and the interstate system. There are shared elements of
international society (such as common language and norms), diplomatic rules
prescribing how states should represent themselves, and rules implied by the

84. Alexander Wendt and Raymond Duvall, "Institutions and International Order," in
Ernst-Otto Czempiel and James N. Rosenau, eds., Global Changes and Theoretical Challenges
very concept of sovereignty. Yet sovereignty, as several have pointed out, is a concept that is laden with social meaning and rests on rules about mutual recognition, noninterference in domestic affairs, and rights of foreign diplomats.  

The social conception of the interstate system implies a different understanding of cooperation. The fundamental starting point is not atomistic interaction within anarchy. The basic question is not how to cooperate and to derive rules, norms, and sociality from a rule-free, normless state of nature. Rather, the starting point is a social conception of the actors, and the basic questions have to do with how the system of states can reproduce itself, what tensions it incorporates, and what capacity it has for altering its structures and rules to deal with changing environmental pressures.

To summarize the above discussion, proponents of the institutional approach do not view rules, norms, and habits of cooperation exclusively as something external to agents (states), something that agents "bump into" or "run up against" as they interact with one another. Instead, they recognize that these practices are often constitutive of the identities and powers of agents in the first place. They also recognize that institutions are not necessarily chosen on a rational basis but instead are the products (residual products) of unconscious trial-and-error activity, coupled with selective pressures in the environment. Whether this unconscious activity is at the cultural level of developing shared symbols and meaning or is hard-wired into states themselves is beyond the scope of this article. However, both conceptions of rules have their advocates.

I have proposed three core principles of institutionalism, the first one ontological, the second theoretical, and the third interpretive. These principles have important implications for how institutionalists identify questions concerning multilateralism, the method of analysis, and the substantive core of their research program.

The core question is decidedly not the state of nature problématique. The "problem of cooperation" is not representable primarily as a game of strategic interaction, although this dimension of interstate behavior surely exists. The emphasis shifts from strategic interaction with given (and fixed) utilities to a model of debate, communication, persuasion, argument, and discursive legiti-


86. The work of Richard Ashley and Friedrich Kratochwil seems to be grounded in culturally acquired norms and practices. According to Hayek's conception of rules, however, our rule-based behavior has its origins in evolutionary processes and is the result of blind trial and error and selective retention. Consequently, rules are arrived at unconsciously and are "hard-wired" in the biological organism, at least in part. For a discussion of Hayek's theory of rules, see Anna Elizabeth Galeotti, "Individualism, Social Rules, Tradition: The Case of Friedrich A. Hayek," Political Theory 15 (May 1987), pp. 163–81.
mation. Indeed, as Friedrich Kratochwil notes, “Most of our arguments concerning policy or rights are not so much about the determination of the likely result, given a certain distribution of ‘preferences,’ as they are debates over which preferences deserve priority over others, which ones ought to be changed, and which judgements deserve our assent. Here the overall persuasive ‘weight’ of claims rather than their logical necessity or aggregation is at issue.”

The international system, then, is not just a collection of independent states in interaction. Sovereignty is not a concept that is sensibly applied to a single state or to numerous states in isolation from one another. It is inherently a relational concept. Multilateralism as an organizing principle would have to focus on the constitutive principles of the state system and to draw out its implicit and sometimes hidden sociality. With respect to multilateral activity, institutionalists heavily emphasize the discursive, deliberative, and persuasive aspects of communication and argument. The interstate system is a forum as well as a chessboard, and its actors debate, argue, and justify as well as signal moves.

Another implication of institutionalism runs along methodological lines. This implication is clarified by contrasting the institutional approach presented here with economic theories of institutions in the Coase-Williamson tradition. The analytic starting point of the institutional approach requires, in addition to data about power and interests, some extant institutional data, including data about rules, social structures, values, and common understandings that are “historically given.” Economic theories of institutions stress the overall fit of institutions to the environment. The emergence of and changes in organizations are efficient responses to environmental challenges. Institutional arguments often stress the contingent, path-dependent nature of institutional change. These arguments generally assume a narrative form in which timing and sequences matter.

These two different approaches suggest two different views of history. The economic orientation relies on efficiency and has an equifinal structure. Institutions may offer some frictional resistance, but in the end they can be incorporated into the general equilibrium model of allocative efficiency. In addition, the model suggests similar (if not uniform) institutional outcomes.

90. For an important exception, see Douglass C. North, “Institutions and Their Consequences for Economic Performance,” in Karen Schweers Cook and Margaret Levi, eds., The Limits of Rationality (Chicago: University of Chicago Press, 1990), pp. 383-401. On p. 392 of this chapter, North recognizes that the neoclassical economics literature on institutions focuses primarily on institutions as efficient solutions to problems of economic organization.
Given the starting point described in terms of preferences, endowments, technology, and transaction costs, analysts can predict outcomes, including institutional outcomes. Since the environment in which economic institutions operate is competitive, there are strong selective pressures toward institutional homogeneity. Thus, it does not matter much where the analyst starts and what sequences occur. Institutions respond to changes in the major independent variables independently of the sequences in which they occur. The method of comparative statics drives out narrative. By contrast, the institutional approach is more attuned to variations in initial conditions and to the sequences in which particular events and processes occur. It speaks the language of eras, conjunctures, and historically produced choices. At the microlevel, individuals may still be described in instrumental terms, but their future choices, their perceptions of what is possible, and their beliefs and standard operating procedures are products of past historical choices.

The paradigms of the first, or functional, approach are modernization theory and economic theories of history in which efficiency considerations determine outcomes. An example of the second, or path-dependence, approach is Alexander Gerschenkron’s theory, which emphasizes how important the timing of entry into the international market is for development, particularly as it relates to the role of the state in the economy. The stress on historical contingency and path-dependent behavior suggests that many different institutional worlds are possible. What we observe at any point in time is not necessarily efficient (compared with what might have been chosen had other historical contingencies intervened). There is even less reason for believing that uniformity of institutional outcomes is likely, since the model producing these outcomes is highly sensitive to minor variations in its parameters and thus does not have robust properties.

These two views of history have implications for our understanding of the multilateral organization of our contemporary world. Although social scientists do not like to engage explicitly in counterfactuals, it is necessary to do so. To say that something occurred the way it did because of $x$ implies that things would have been different in the absence of $x$. To pose the issue in this way prompts us to ask several questions: Is multilateralism an inevitable organizational form of the modern state system? Is it the institution that would be recreated if we could replay history with a number of different initial conditions, choice points, and historical trajectories? For example, what would have happened, Ruggie asks, if Germany had prevailed in World War II? Would Schachtian bilateralism have been discarded in favor of multilateralism? The apparent differences among the ideas of prewar Germany, postwar Japan, the Soviet Union, and the United States with respect to organizing the world political economy should at least give pause.

92. See Ruggie, “Multilateralism.”
There is a research program implied by the institutional approach. It is reflected in the numerous puzzles and theoretical challenges suggested by the approach, in the unanswered questions regarding the importance of norms and beliefs, in the tough transition from assumption to theory regarding the independent importance of the institutional level of analysis, and in the task of developing rigorous methodologies to explore theoretical issues historically, as I have used that term in this article.

In this research program, norms, ideas, and social purposes will occupy a central location. Mainstream international relations theory relies heavily on power, interest, and anarchy for its explanatory foundation. Indeed, since interest (national interest) is partly determined by the structure of the international system (anarchy and the distribution of power), only two factors can independently vary. Scholars such as John Ruggie, Hayward Alker, and Richard Ashley want to question this paradigm. Ruggie does not accept accounts of regimes that rest on power distribution explanations only. The concentration of power may make it easier to solve collective action problems, but it tells us nothing about the content of the regime. For this, we must introduce ideas, norms, and social purposes.

The institutionalist research program is already being carried out, although not all scholars would subscribe to the label, nor would they necessarily comply with all of the components of the program elaborated here. Ernst Haas’s work on the importance of consensual knowledge, common beliefs, learning, and perceptual change in international organizational environments is exemplary. Working within a domestic context, Judith Goldstein has tried to demonstrate how key norms and ideas about free trade have been incorporated into national laws and institutions. The transfer of responsibility for commercial policy from the legislative to the executive branch, the evolution of the International Trade Commission, and the distancing of trade policy from public opinion have all served to preserve free trade, despite periods of high protectionist demands. Goldstein’s work shows that multilateralism is not just an international affair; it has crucial domestic components as well. Finally, Alexander Wendt’s work focuses on the ways in which power politics, generally taken as an unproblematic given of the modern world, is a construction owing much to processes of socialization, learning, and cognitive change. Our present international


94. See the following works of Ernst B. Haas: Beyond the Nation-State; Tangle of Hopes: American Commitments and World Order (Englewood Cliffs, N. J.: Prentice-Hall, 1969); and When Knowledge Is Power.


system is not eternal; it has a history, and it has a future that we are ill-equipped to theorize, given the static nature of many of our theories.

Conclusion

The basic argument of this article builds on the premise that multilateralism has been relatively neglected in international relations theory. While multilateral organizations and practices have been examined empirically, their sources have not been the subject of much concern in realist and neorealist theory. Part of the neglect stems from an implicit ontology that takes states as the social atoms of international society and explores how their interaction might support large-member organizational forms. Nevertheless, individualist approaches have met with some success, as the discussion in an earlier section of this article demonstrates.

An important debate in international relations is taking place between proponents of rationalist theory, which is exemplified by neorealism, and proponents of reflectivism, which is closer to institutionalism as described in this article. Both groups are interested in explaining the sociality of states—that is, the manner in which states acknowledge membership in and contribute to international society. However, each addresses the sociality of states in a different way.

The theoretical project launched by the neorealists involves the demonstration that cooperative behavior among many players can emerge as a result of self-interested strategic interactions and can do so within a class of games that is itself noncooperative in its paucity of communications, trust, and third-party enforcement. This is not a project that is doomed to failure. Some significant advances have been made. In addition, there are some research programs broadly related to neorealism that have not been explored here. Theories of public goods could be examined to see if the scope of externalities might provide insight into the question of the generality of cooperation. The economic theory of clubs, which deals with impure public goods and congestion effects, could be examined with the same question in mind. And the bodies of literature on tactical and party linkages have not been explored for the returns they might offer.

There are limitations to neorealist approaches, however. Neorealism underestimates the extent to which cooperation depends on a prior set of unacknowledged claims about the embeddedness of cooperative habits, shared values, and taken-for-granted rules. Further, its assumption that preferences are exogenously given reduces multilateralism to a question of strategic interac-

97. See Keohane, “International Institutions.”
98. See Taylor, The Possibility of Cooperation; and Schuessler, “Exit Threats and Cooperation Under Anonymity.”
tion, making it difficult to comprehend multilateralism propelled by collective beliefs, presumptive habits, and shared values. Finally, the absence of a historical (narrative) approach discourages the exploration of counterfactuals and lends support to the view that arrangements, including institutional arrangements, are what they are either because they represent functional responses to environmental challenges or because they reflect the prevailing power distribution.

Reflectivists reject the state of nature as the appropriate starting point even for heuristic purposes. If states are characterized only by interests and strategies, cooperative outcomes will not occur. Shared understandings regarding the rules of the game, the nature of permissible plays, the linkages between choices and outcomes, and the nature of agents involved in the game are important preconditions. To say this is to acknowledge that shared understandings and communicative rationality are as important as instrumental rationality.\textsuperscript{100}

Instead of deriving sociality from the state of nature, the task of reflectionists is to show how socially defined states, operating within given institutional sites, engage in behavior that is both competitive and cooperative. The problem is how to explain institutions and sociality given some data on extant and prior institutions and sociality. Thus, in trying to deal with the problem of the distribution of wealth internationally, Naeem Inayatullah starts from a conception of state sovereignty that is highly social in its mutual recognition and ongoing practices.\textsuperscript{101} The work of Robert Jackson is also exemplary of the tradition in which the social aspects of state sovereignty are taken seriously.\textsuperscript{102}

In summary, I have premised this article on a broad correspondence between individualist, instrumental approaches and neorealism on the one hand and between institutional approaches and reflectivism on the other. Without rejecting individualist, instrumental approaches, I have tried to identify the limits of their usefulness and to open up a theoretical space for institutional analysis. While an institutional approach to multilateralism should not banish individuals, intentional behavior, and strategic interaction, it highlights different things. It departs from the preference-action paradigm by introducing structure and behavior that draw on habitual, preconscious, taken-for-granted understandings. As such, the institutional approach "defocalizes interest."\textsuperscript{103} It


\textsuperscript{102} See the following works of Robert H. Jackson: "Quasi-States, Dual Regimes, and Neoclassical Theory"; and "Negative Sovereignty in Sub-Saharan Africa," \textit{Review of International Studies} 12 (October 1986), pp. 247-64.

might try to understand the emergence of multilateralism as a product of the power, resources, and beliefs of important actors and the reproduction of multilateral institutions in terms of organizational inertia, socialization to system norms, and adaptation to the “needs of the institution.”

All three approaches examined here—the individualist, social-communicative, and institutional approaches—are only partly worked out, particularly in their relation to multilateralism. None by itself is a panacea. The questions addressed are big ones, sometimes even daunting. But the promise is considerable, too. Hopefully, this article has identified some of the tasks and pointed to the challenges and difficulties on the way ahead.