The promise of
Institutionalist Theory
Robert O. Keohane
and
Lisa L. Martin

In his usual direct way, John J. Mearsheimer has sharpened the theoretical issues dividing realist from institutionalist theory, and for this service we are grateful. We are also pleased that he has read the institutionalist literature so thoroughly. He correctly asserts that liberal institutionalists treat states as rational egoists operating in a world in which agreements cannot be hierarchically enforced, and that institutionalists only expect interstate cooperation to occur if states have significant common interests. Hence institutionalist theory does not espouse the Wilsonian concept of collective security—which Charles and Clifford Kupchan refer to as “ideal collective security”—critiqued so well by I.L. Claude thirty years ago.¹

Nor does institutionalism embrace the aspirations to transform international relations put forward by some critical theorists. Like realism, institutionalist theory is utilitarian and rationalistic.²

However, Professor Mearsheimer’s version of realism has some rather serious flaws. Among them are its penchant for assertions that turn out to be incorrect; its propensity to privilege its own viewpoint, so that in the absence of decisive evidence either way it invariably seems to prevail; its failure to explicate the conditions for the operation of its generalizations; and its logical contradictions, escaped only through verbal sleight-of-hand. We will begin by pointing out such errors from his own recent articles in this journal, then


The authors thank Marc Busch, Chris Gelpi, Andrew Moravcsik, and Celeste Wallander for their valuable comments on an earlier version of this essay.


2. See Richard K. Ashley, “The Poverty of Neorealism,” International Organization, Vol. 38, No. 2 (Spring 1984), pp. 225–286. Ashley included Robert O. Keohane as one of the “neorealists” whose “orrery of errors” he rejected. The fact that Mearsheimer criticized institutionalism and critical theory in the same article should not, therefore, lead readers to believe that there is an intellectual affinity between these two schools of thought. However, the work of “constructivist” theorists such as Alexander Wendt eloquently makes a number of arguments that many institutionalists would accept.

© 1995 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.
examine his major claims about institutionalism. We consider the illusory divide between security and economic issues, the muddled question of "relative gains," and empirical work (admittedly in its early stages) that provides evidence of the significance of international institutions. We conclude that institutions sometimes matter, and that it is a worthy task of social science to discover how, and under what conditions, this is the case.

**The Fallacious Logic of Realism**

Five years ago Professor Mearsheimer forecast the imminent decline of NATO: "It is the Soviet threat that holds NATO together. Take away that offensive threat and the United States is likely to abandon the Continent, whereupon the defensive alliance it headed for forty years may disintegrate."3 At the same time, he predicted that "the EC is likely [due to the end of the Cold War] to grow weaker, not stronger with time."4 Yet now that both NATO and the European Community, now the European Union (EU), are expanding their memberships, and hardly in decline, he abandons specificity for the equally false but more difficult to falsify generalization that "institutions have minimal influence on state behavior and thus hold little prospect for promoting stability in a post–Cold War world."5

Professor Mearsheimer demands proof that international institutions matter. Yet he begins his article by reminding us that major governments recently have been emphasizing the value of international institutions; he could have added that they invest significant material and reputational resources in NATO, the EU, and also in organizations such as the General Agreement on Tariffs and Trade (GATT, recently strengthened to create the World Trade Organization) and the North American Free Trade Agreement (NAFTA). Not all international institutions command such resources from governments, but some do. How are we to account for the willingness of major states to invest resources in expanding international institutions, if such institutions are lacking in significance? Mearsheimer suggests that the answer lies in an ideological blindness of American policymakers, whose hostility toward realism drives them to the more congenial institutionalist framework (pp. 47–49). It is difficult to

---

square this assertion of a collective delusion with the dominant role of realist theory in policy discussions, or with realism's own precepts about the forces that drive state behavior. In light of states' investments in international institutions, it is fair to turn Mearsheimer's question around: could we not legitimated demand evidence either that leaders of governments are deluded or that NATO and the EU are designed to deceive unsophisticated observers? Mearsheimer assumes that his view is privileged, in the sense that we must accept realism unless overwhelmingly convincing evidence is presented for an alternative view; but the fact that states invest in international institutions make this stance quite problematic.

Institutionalism and realism differ in a number of other respects, one of the most significant of which concerns how they approach social science. A central fault of Mearsheimer's realism as a scientific theory—rather than as rhetoric—is that the conditions for the operation of its "grim picture of world politics" (p. 9) typically are not well-specified. Realism is replete with global generalizations, lacking qualifications about the conditions under which they may be valid. Let us consider two examples from Mearsheimer's own article. First, Mearsheimer writes that "states in a realist world . . . must be motivated primarily by relative gains concerns when considering cooperation" (p. 12, emphasis added). But he later admits that this proposition may be false when the threat of aggressive war is low—for instance, when defensive technologies (such as secure second-strike nuclear forces) are prevalent (pp. 23–25). Second, in Mearsheimer's realist world, "every state would like to be the most formidable military power in the system" (p. 12). But since no one thinks that Switzerland, Argentina, or contemporary Britain actually seeks to become "the most formidable military power," what Mearsheimer presumably means to argue is that states with sufficient capabilities always pursue this goal. Even this statement is often false: for example, the United States during the interwar period could reasonably have expected to become the most powerful state in the world, but did not seek such a position. Confronted with such contradictions and anomalies, realism typically retreats from universal rhetoric to post hoc and ad hoc qualifications, taking into account geography, history, perceptions, and domestic politics.

Institutionalism, in contrast, seeks to state in advance the conditions under which its propositions apply. Our theory may therefore have less appeal to those who require simple "truths," but purportedly scientific theories should specify the conditions under which the theory is expected to hold a priori. As Mearsheimer indicates, when state elites do not foresee self-interested benefits from cooperation, we do not expect cooperation to occur, nor the institutions
that facilitate cooperation to develop. When states can jointly benefit from cooperation, on the other hand, we expect governments to attempt to construct such institutions. Institutions can provide information, reduce transaction costs, make commitments more credible, establish focal points for coordination, and in general facilitate the operation of reciprocity. By seeking to specify the conditions under which institutions can have an impact and cooperation can occur, institutionalist theory shows under what conditions realist propositions are valid. It is in this sense that institutionalism claims to subsume realism.

Realism’s proclivity for bold, unqualified generalizations not only generates anomalies but gets its proponents into logical difficulties. Mearsheimer holds that “institutions have no independent effect on state behavior” (p. 7); that NATO is an institution (p. 13); and that NATO played a role in preventing World War III and helping the West win the Cold War (pp. 13–14). These propositions sound like a classically fallacious syllogism, until one recognizes that there is an escape clause: “NATO was basically a manifestation of the bipolar distribution of power in Europe during the Cold War, and it was that balance of power, not NATO per se, that provided the key to maintaining stability on the continent” (p. 14). But liberal institutionalists, who see institutions as rooted in the realities of power and interest, do not argue that NATO could have maintained stability under any imaginable conditions. What we argue is that institutions make a significant difference in conjunction with power realities. Institutions are important “independently” only in the ordinary sense used in social science: controlling for the effects of power and interests, it matters whether they exist. They also have an interactive effect, meaning that their impact on outcomes varies, depending on the nature of power and interests. Mearsheimer is forced to admit the truth of institutional effects with regard to NATO, although for rhetorical purposes he shifts his ground to attack a view that we do not hold: that institutions can prevent war regardless of the structure in which they operate.

Hence Mearsheimer’s version of realism is replete with analytical problems. However, it is not our duty here to correct realism’s copy-book. In the rest of this brief response, therefore, we focus on the promise of institutionalist theory, and the research directions that we hope will help to realize that promise.

**Political Economy vs. Security and the Issue of Relative Gains**

Although Mearsheimer has provided an admirable summary of several aspects of institutionalist theory, his version of our argument requires correction on
two major points. First, Mearsheimer asserts that institutionalist theory is based on "the assumption that international politics can be divided into two realms—security and political economy—and that liberal institutionalism mainly applies to the latter" (pp. 15-16). Although some institutionalists have made this assertion, it is not the predominant view of the institutionalist literature, and we certainly do not accept it. Secondly, in contrast to Mearsheimer's assertion, our focus is not exclusively on "cheating." Situations of coordination, in which cheating is not a problem but distributional issues are serious, are equally important, although they were underemphasized (but not absent) in the early institutionalist literature.

THE PURPORTED SECURITY VS. POLITICAL ECONOMY DIVIDE

Mearsheimer's assertion that institutionalism employs a "neat dividing line" to separate political economy from security issues is surprising, in view of the attention that he devotes to the volume edited by Kenneth Oye, *Cooperation Under Anarchy*. A major argument of *Cooperation Under Anarchy* is that institutionalist theory can be applied to both security and political economy issues. As Robert Axelrod and Robert O. Keohane wrote:

It has often been noted that military-security issues display more of the characteristics associated with anarchy than do political-economic ones. Charles Lipson, for instance, has recently observed that political-economic relationships are typically more institutionalized than military-security ones. This does not mean, however, that analysis of these two sets of issues requires two separate analytical frameworks. Indeed, one of the major purposes of the present collection is to show that a single framework can throw light on both [emphasis added].

We share Mearsheimer's view that there is no clean analytical line between economic and security issues, although we do not base our view on the overarching role of relative gains. Institutionalist theory should be highly applicable to security issues because its argument revolves around the role of institutions in providing information. This argument is pertinent to realist security arguments, which often rely on worst-case analysis. Realists contend that in an uncertain, anarchic world, states must assume the worst, particularly about others' intentions, when making policy choices. Worst-case analysis

---

implies following policies that do not maximize expected utility for the sake of avoiding terrible outcomes. But if one can secure more information, it may be possible to follow policies that more nearly maximize utility.\textsuperscript{7} Realist writers from Kautilya on have stressed the significance of information (intelligence); if institutions can provide useful information, realists should see them as significant. The logic of institutionalist theory is directly applicable to security problems as realists define them.

Hence, if Mearsheimer meant to offer us a “loophole” through which to escape his criticism—that institutionalist theory is only applicable to non-security issues—we emphatically refuse to avail ourselves of his generosity. On the contrary, we hope that, to use Axelrod’s phrase, institutionalist theory will gradually “invade” the study of security issues, helping to explain variation in institutional form without denying the validity of many realist insights into power and interests.

RELATIVE GAINS AND INTERNATIONAL COOPERATION

The conclusions we draw from the “relative gains” debate are different from those of Professor Mearsheimer. It is true that when only two states exist and they have perfectly conflicting interests, institutions will not be significant, but this point is obvious. Two issues are more significant: 1) the conditions under which relative gains are important; and 2) the role of institutions when distributional issues are significant—that is, when relative gains are at stake.

It is important to understand the great variation in the extent to which relative gains matter. The major lesson of the recent debate on relative gains is that their importance is \textit{conditional} on factors such as the number of major actors in the system and whether military advantage favors offense or defense.\textsuperscript{8} Duncan Snidal has shown that relative gains are unlikely to have much impact on cooperation if the potential absolute gains from cooperation are substantial, or in any context involving more than two states.\textsuperscript{9} A valuable aspect of the relative gains debate is that it has made distributional and bargaining issues


more salient than they were in early neoliberal thinking, but if the debate becomes one of "whether" relative gains matter, that value will be dissipated. We need instead to ask under what conditions such distributional conflicts are severe.

What is the role of institutions when distributional issues are important? Contrary to the assertion that institutionalist theory is irrelevant to distributional issues, we argue that distributional conflict may render institutions more important. To understand this point, it is essential to distinguish between two problems that states face when they attempt to cooperate. They often worry about the potential for others to cheat, as in a Prisoners' Dilemma. But they also face the problem of coordinating their actions on a particular stable cooperative outcome (solving the problem of multiple equilibria, in game-theoretic terminology). Usually more than one cooperative outcome exists. The states involved may not agree on which of these outcomes is preferred, as each has different distributional implications. Disagreement about the specific form of cooperation is the principal barrier to cooperation in such coordination games. Unless some coordinating mechanism exists, states may fail to capture the potential gains from cooperation. Institutions do not provide the only possible coordinating mechanism. However, in complex situations involving many states, international institutions can step in to provide "constructed focal points" that make particular cooperative outcomes prominent.

Realists interpret the relative-gains logic as showing that states will not cooperate with one another if each suspects that its potential partners are gaining more from cooperation than it is. However, just as institutions can mitigate fears of cheating and so allow cooperation to emerge, so can they alleviate fears of unequal gains from cooperation. Liberal theory argues that institutions provide valuable information, and information about the distribution of gains from cooperation may be especially valuable if the relative-gains logic is correct. Institutions can facilitate cooperation by helping to settle distributional conflicts and by assuring states that gains are evenly divided over


time, for example by disclosing information about the military expenditures and capacities of alliance members.

In our view the successful functioning of institutions depends heavily on the operation of reciprocity, both specific and diffuse. In states using strategies of reciprocity are engaged in exchange with one another and so require information about the value of their exchanges. Institutionalized reciprocity and distributional concerns are simply two sides of the same coin, reflecting the difficulties of cooperating in a system lacking centralized enforcement and pointing to the need for reliable sources of information if states are to achieve gains from cooperation. Far from leading to the conclusion that institutions are not significant in world politics, the relative-gains debate has led us to understand yet another pathway through which they substantially influence the course of international relations. A crucial step in the institutionalist research program will be to understand the conditions under which institutions can provide the information necessary to serve as reliable solutions to distributional problems.

**Empirical Work on the Impact of Institutions**

We agree with John Mearsheimer that “more empirical work is needed before a final judgment is rendered on the explanatory power of liberal institutionalism” (p. 26). The point of a new theory is to generate testable hypotheses: liberal institutionalism, like any other theory, only has value insofar as it generates propositions that can be tested against real evidence.

Institutionalist theory conceptualizes institutions both as independent and dependent variables: “institutions change as a result of human action, and the changes in expectations and process that result can exert profound effects on state behavior.” Institutional theory has a coherent account of both the creation of institutions and their effects: institutions are created by states because of their anticipated effects on patterns of behavior. Early research by institutionalists focused on institutions as dependent variables, examining the conditions under which they are created. Recent research has sought more systematically

---


to demonstrate that institutions are sometimes significant for political outcomes, and to determine the conditions under which this is the case.\textsuperscript{14}

In view of this research program, it should be clear that evidence that institutions change in response to underlying conditions is hardly a blow against institutionalist theory. That theory, after all, posits that international institutions are created in response to state interests, and that their character is structured by the prevailing distribution of capabilities. The real empirical issue is how to distinguish the effects of underlying conditions from those of the institutions themselves. One result of the interdependence between institutions and underlying forces is that research designed to isolate the impact of institutions is difficult to design and execute. Rarely, if ever, will institutions vary while the "rest of the world" is held constant. Thus finding the ideal quasi-experimental situation to test the impact of institutions is not possible.

However, these difficulties do not make it impossible to test the argument that institutions matter, since changes in underlying conditions and in institutions are not perfectly correlated. Hence it may be worthwhile to search for instances in which underlying conditions have changed rapidly while institutions have remained relatively constant, or where similar structural changes confront regions that have different institutional endowments. Another tactic may be to consider the level of institutional variation itself. The institutionalist perspective leads us to expect patterned variation in the types of institutions states construct, since they anticipate that institutions so constructed will constrain them. Analysis of institutional form, such as variations in the institutionalization of alliances or in the legalization of the international trading system, should therefore provide valuable evidence for evaluating institutionalist theory.

Realism's insistence that institutions have only marginal effects renders its account of institutional creation incomplete and logically unsound, and leaves it without a plausible account of the investments that states have made in such international institutions as the EU, NATO, GATT, and regional trading organizations. According to the precepts of realist theory, states act rationally when they construct institutions, although they know that these institutions will have

\textsuperscript{14} Since institutionalists do not claim that institutions always have a major impact on outcomes, finding weak institutions hardly constitutes a refutation of institutionalist theory. Hence the weakness of the International Energy Agency during the 1979 oil crisis, described by Keohane in After Hegemony: Cooperation and Discord in the World Political Economy (Princeton: Princeton University Press, 1984), is hardly the damning evidence that Mearsheimer claims.
no impact on patterns of cooperation. But what could be the rationale behind devoting resources to structures that will make no difference? Rather than asserting that institutions have no impact, realists must mean that institutions have some effect other than that assumed by liberal institutionalists. Perhaps institutions satisfy the ideological demands of statesmen, or help to pacify inattentive publics. Whatever the rationale, we challenge realists to construct an account of institutional variation and effects that can be tested against the institutionalist alternative. The difference between realism and liberal institutionalism does not lie in whether institutions are independent or dependent variables; it lies in contrasting understandings of why institutions are created and how they exert their effects.

A number of recent studies establish institutional effects through careful empirical research, guided by institutionalist theory and recognizing potential problems of endogeneity and omitted-variable bias. Ronald B. Mitchell shows that on three different issues involving oil pollution at sea, whether states complied with institutional regulations depended on the nature of the rules. “Clear causal links unambiguously demonstrate that treaty rules independently influenced behavior, with other plausible factors controlled for or absent.” New rules on the kinds of tanks that ships are allowed to use, for example, have had a dramatic impact on intentional discharge of oil into the oceans.

The European Court of Justice (ECJ) has also proven a fruitful ground for the study of institutional influence. Anne-Marie Slaughter Burley and Walter Mattli show how the ECJ has had an unexpectedly large impact on the politics of European integration, transforming political into legal issues with the aid of transnational networks of lawyers and judges. The ECJ has gone far to convert the Treaties of Rome into a constitution for the EU, with the result that EU law now reaches deeply into the domestic law of member states. Geoffrey Garrett and Barry Weingast, in another study of the ECJ, show how it resolved problems of multiple equilibria for EU member states by providing constructed

focal points in coordination problems. These studies show that institutions have the wide range of effects attributed to them by liberal institutionalists. They change the incentives for states to cheat; they also reduce transaction costs, link issues, and provide focal points for cooperation.

The institutionalist perspective has also been applied with success to the analysis of security regimes. John Duffield has considered NATO as a regional security regime. He finds that NATO made an independent contribution to the “Long Peace” in Europe by drawing boundaries, demonstrating U.S. commitments and making them credible, and facilitating the augmentation of NATO allies’ military capabilities. He also finds that the stable norms and rules of NATO led to stability in levels of conventional forces within the regime that cannot be explained by structural theories.

In Coercive Cooperation, Lisa Martin showed that the involvement of international organizations in economic sanctions is strongly correlated with high levels of cooperation. Since such a correlation does not establish causality, she also did qualitative work on several cases involving sanctions, including EC sanctions against Argentina during the Falklands War. Mearsheimer considers the Falklands case in isolation from the rest of this research, and dismisses it as “less than a ringing endorsement for liberal institutionalism” on the grounds that concerns about cheating were not involved (p. 25). In fact, Martin does find evidence that states used the EC framework to reduce fears of cheating, in the form of taking advantage of the situation to profit from trade with Argentina. However, the major effect of institutions came through institutionalized linkages that would otherwise have been nonexistent: a linkage between EC budget contributions and the sanctions issue. Prevention of cheating is not the only mechanism by which institutions facilitate cooperation. By creating issue linkages, they allow for more effective retaliation against cheaters and also create scope for mutually-beneficial exchanges. Further evidence for the

22. Ibid., p. 143.
EC's role in coordinating sanctions comes from the fact that outside the EC, the only other significant support Britain received came from Commonwealth nations and the United States. In the U.S. case, support was delayed until after the outbreak of war, in distinct contrast to the behavior of EC members. Mearsheimer's dismissal of international institutions implies that linkages are easy to forge when a state desires cooperation, and that cooperation is easy to coordinate even without institutions, yet Britain did not find either to be the case. Even in isolation from the robust statistical results and other case studies reported in *Coercive Cooperation*, the Falklands case illustrates the central role of formal international institutions in enabling states to cooperate to impose multilateral economic sanctions.

Institutions sometimes matter for state policy, but we do not adequately understand in what domains they matter most, under what conditions, and how their effects are exerted. More research on this subject, by students of world politics critical of institutionalist theory as well as by those working from it, is essential, and will be most welcome.

**Conclusion**

Far from demonstrating the irrelevance of international institutions, Mearsheimer's characterization of conflict in world politics makes institutions appear essential if states are to have any hope of sustained cooperation, and of reaping its benefits. This necessity for institutions does not mean that they are always valuable, much less that they operate without respect to power and interests, constitute a panacea for violent conflict, or always reduce the likelihood of war. Claiming too much for international institutions would indeed be a "false promise." But in a world politics constrained by state power and divergent interests, and unlikely to experience effective hierarchical governance, international institutions operating on the basis of reciprocity will be components of any lasting peace.

23. Japan initially refused British pleas to impose sanctions, and took only minor steps following U.S. imposition of sanctions, much later than EC members.
24. The Falklands case cannot be dismissed on grounds that, as Mearsheimer claims, striking a deal was "not difficult." The historical record shows intense conflict, including public protests in some countries and challenges to the sitting government in others. The Thatcher government believed that its survival was at stake in the Falklands War. While perhaps not a "core interest" by realist standards, government survival is surely a fundamental concern of policymakers that could impede cooperation.
The institutionalist research program in international relations is a promising one. The logic of institutionalist theory, with its focus on the informational role of institutions, appears solid. Institutionalists should respond to Mearsheimer's criticisms by better integrating distributional considerations into their models, further specifying the causal mechanisms by which institutions exercise influence, and building on existing empirical work to provide more convincing evidence of institutional effects. Both the questions raised and the provisional answers given by institutionalists, during the relatively short life of this research program, indicate that these tasks may be rewarding. In comparison with the extant alternatives, the promise of institutionalist theory seems bright.