

Choosing How to Cooperate: A Repeated Public-Goods Model of International Relations

RANDALL W. STONE

University of Rochester

BRANISLAV L. SLANTCHEV

University of California–San Diego

TAMAR R. LONDON

Pennsylvania State University

International relations theory has borrowed important intuitions from Olson's static public-goods model (hegemonic stability) and the repeated prisoners' dilemma (theories of international cooperation), and arguments often combine implications from both models. We develop a general, repeated public-goods model. We then allow the qualitative dimensions of cooperation to emerge endogenously: agreements can have broad or narrow membership and entail deep or shallow commitments; they can be multilateral or discriminatory; they can be ad hoc or institutionalized. We find that the relationship between the distribution of power and international cooperation is complex: a large leading state forms a narrow coalition of intensive contributors, and builds institutions, while a smaller leading state forms a broader coalition that makes shallow contributions, and is more inclined to multilateralism.

Theories of international cooperation generally make use of two key concepts drawn from formal theory: (1) the underprovision of public goods, and (2) the reputational benefits of repetition in cooperation games. However, most models that are used to derive these intuitions only incorporate one of these two features, and there is no reason to expect the intuitions to survive in a more general model. Most theories of international public goods are based implicitly or explicitly on a static models (Gowa 1989; Olson and Zeckhauser 1966; Oye 1992; Snidal 1985). On the other hand, theories that are based on repeated-game models, such as the repeated prisoners' dilemma, do not explicitly model cooperation as provision of a public good (Axelrod 1984; Keohane 1984; Krasner

Authors' note: Randall W. Stone is the corresponding author (randall.stone@rochester.edu). We thank Robert Keohane, Andy Kydd, Robert Pahre, and Duncal Snidal for useful comments on an earlier version. The replication package can be found at <http://polisci.ucsd.edu/slantchev> and the ISQ Dataverse at <http://dvn.iq.harvard.edu/dvn/dv/isq>. It includes Aptech Gauss programs for the simulations and the complete proofs.

1976). We propose to replace these partial models with a general model of international cooperation as a repeated game of provision of a public good, which subsumes many of the arguments in the existing literature as special cases.

The first payoff of such a model is that we are able to explore the implications of the distribution of power for international cooperation. The most serious contender for a theory of power and international cooperation has been the theory of “hegemonic stability”—a theory which, however, has been rather thoroughly discredited. There is little empirical evidence to support it, and even the scholars who did the most to popularize it argued from the start that its theoretical foundations were full of holes (Keohane 1984). This is unfortunate, because it is obvious that the distribution of power is a key variable, and there is an urgent need to develop credible theories of how it affects international cooperation. We continue to use the term *hegemon*, for the sake of continuity, but we allow the model to determine what the practical implications of a skewed distribution of power might be. The distribution of “sizes” of states is an exogenous parameter of the model, so we are able to derive comparative statics about the difficulty of sustaining cooperation under various degrees of hegemony—some of which confirm accepted wisdom, but some of which are surprising. For example, we find that participation in cooperative endeavors generally increases when a hegemon’s strength declines.

As we will show, the distribution of power not only affects the degree of cooperation, but redefines it qualitatively as well. The qualitative dimension of international cooperation is how cooperation takes place: are cooperative coalitions broad or narrow; are regimes discriminatory or nondiscriminatory; is cooperation informal or institutionalized? The model is designed to make each of these choices endogenous, so that the answers emerge as the result of strategic calculations. We address several key questions in the literature on international cooperation:

1. Why are some international cooperative arrangements inclusive, but shallow, while other more intensive cooperative regimes are limited in membership? What is the nature of the trade-off that has often been described between the breadth and depth of cooperation?
2. Why are some international regimes discriminatory, while others are multilateral? What is the relationship between discrimination and hegemony? Does the dominance of a single state lead to a more open or more closed regime?
3. Under what circumstances do the leading states in the system choose to build international institutions? Is the construction of institutions associated with the dominance of a single great power?

International Relations and Public Goods

The theory of hegemonic stability arose as a generalization from two prominent empirical cases in which the presence of unusually dominant states in the international economy—Britain in the late nineteenth century, and the United States in the middle of the twentieth—was associated with unusually high degrees of international openness in trade and foreign exchange policies (Gilpin 1972; Kindleberger 1973; Krasner 1976). Charles Kindleberger stated the argument in terms of Mancur Olson’s theory of collective action, claiming that international economic stability was a public good that was underprovided in the international system unless some dominant state had the interest and capacity to provide it. “For the world economy to be stabilized there has to be a stabilizer—one stabilizer” (Kindleberger 1973, 304).

Empirical tests of the hegemonic stability thesis in the area of international trade have had mixed results (Conybeare 1983; Cowhey and Long 1983; Mans-

field 1994; McKeown 1983, 1991). In international security, as well, the association between the dominance of a single power and international stability—the absence of major war—is contested, although supported by some quantitative results (Bueno de Mesquita and Lalman 1992). Nevertheless, the logic of hegemonic stability has gripped the imagination of the field. The discussion of hegemonic stability quickly moved beyond the original static model of collective action, with both critics and defenders using arguments about long-run considerations. Thus, Robert Keohane argued that long-run considerations could sustain cooperation after a hegemon declined, referring to the repeated prisoners' dilemma (Keohane 1984). When John Conybeare argued that hegemons may have the greatest incentives to exploit weaker nations, Joanne Gowa replied that long-run reputational considerations could keep them from doing so (Conybeare 1984; Gowa 1989). Several authors formalized the argument that hegemony was not necessary for collective action, or that it might even be counterproductive (Pahre 1998; Palmer 1990; Snidal 1985).

This article starts where the hegemonic stability debate left off. We model international cooperation as a repeated game in which a number of states make contributions to a public good. Our interest is to determine how the distribution of capabilities in the international system affects the qualitative dimensions of international cooperation: who contributes, how the burdens are shared, whether agreements are inclusive or exclusive, and whether institutions are constructed. Along the way, we connect our findings to the old hegemonic stability debate, rejecting the hypothesis at some points, confirming it at others, and qualifying it in several ways. The findings are more interesting than the footnotes, however. The distribution of capabilities has profound implications for how states choose to cooperate with each other.

Depth versus Breadth

A continuing refrain in international political economy is that a trade-off generally exists between deep, intensive cooperation and wide, broad-based cooperation. This trade-off has been noted in numerous empirical studies of international cooperation.¹ In environmental cooperation, for example, there is the celebrated “race to the bottom,” which is a metaphor for the debilitating effect upon cooperation of admitting less concerned states into agreements. These states tend to push for more permissive standards, which undermine the discipline of more advanced cooperators. Expanding the circle of cooperating states drives agreements toward a least common denominator that may be much less than a smaller group of states could have achieved.² The controversy concerning the interpretation of the Kyoto Protocol on global climate change at the Hague Summit in 2000 is a case in point: the United States pushed for wider adherence to shallow environmental norms, while the European Union took the lead on environmental issues and pushed for deeper changes in the practices of the advanced industrial countries and toleration of deviations by developing

¹ George Downs and David Rocke provide elegant models to explain why states might wish to limit the depth of cooperation (Downs and Rocke 1995, Ch. 5) or the size of coalitions (*ibid.*, Ch. 6) due to various kinds of uncertainty, and these could be considered competing hypotheses for our model. We regard it as an advantage that we are able to demonstrate this trade-off in the most general case of full information.

² Downs, Rocke and Barsoom (1998) argue that this problem explains why cooperative agreements often begin as narrow coalitions of intensive cooperators, and then gradually expand, taking advantage of sequential voting to ameliorate the diluting influence of including new members. Their evidence suggests that deep integration achieved by the EU and some environmental institutions is explained by the sequential growth strategies that they adopted. Another recent empirical study shows that the effectiveness of multilateral economic sanctions suffers when the cooperating coalition is large, pointing to a trade-off between wide adherence and the credibility of enforcement (Drezner 2000).

countries. Similarly, disputes surrounding the expansion of the European Union have often been cast in terms of the depth versus breadth trade-off: Great Britain has traditionally preferred wider membership and the less intense coordination of policies that would entail, while France and Germany were chary of expanding the club before completing the EMU project.

In order to investigate the relationship between the breadth and depth of cooperation, we construct a model in which these choices are made simultaneously. The leading state in the system, which we designate as the hegemon, chooses a level of cooperation to demand, taking into account the fact that some countries may not be willing to join the coalition if the demanded level of cooperation is too high. The level of cooperation can be thought of as a degree of effort, which has effects that vary depending upon the capabilities of the cooperating state. By the breadth of cooperation, we mean the number of states that actually cooperate. We do not posit that there is a trade-off between the breadth and depth of cooperation, but both the inclusiveness of the contributing coalition and the intensity of cooperation can vary, so it is possible for a trade-off to arise endogenously in the model, and it does. We can then investigate the relationship between the distribution of capabilities and the size and shape of cooperating coalitions.

Multilateralism and Discrimination

Theories of international public goods have long been criticized for the restrictive assumption that the products of international cooperation have the quality of public goods. Public goods are non-excludable and non-rival, like clean air. There is no practical way to prevent polluters from breathing what is, after all, a common resource; and it is difficult to argue that my consumption of air does anything to reduce the supply for my neighbors. Empirically, Joanne Gowa argues, many forms of international cooperation look much more like private goods (Gowa 1989). It is relatively easy to exclude non-contributors from the benefits of some international regimes, as China found, and Russia continues to find, in their long quests to join the WTO. Furthermore, it may not be the case that all the products of international collaboration are non-rival. Some international clubs become less attractive as their memberships grow. For example, Mexico stands to lose its privileged status as an exporter to the United States and Canada as the Free Trade Area of the Americas takes shape.

Kenneth Oye takes the argument a step further, arguing that many of the public goods in the international arena are public by design (Oye 1992). Most-favored-nation (MFN) status is not a necessary characteristic of the trade issue area, he points out, but a deliberately created institution that generates public goods by multilateralizing the concessions made in GATT trade rounds. This creates incentives for small countries to free-ride during negotiating rounds; the large countries tolerate this, however, because the cost of doing business bilaterally is intolerably high. Similarly, the Gold Standard was an exchange-rate regime that generated a demand for public goods, since the stability of the system depended upon the commitments of deficit countries to deflate their economies and ship gold when there was a run on their currency, and of surplus countries to extend emergency financing, hold troubled currencies as reserves, and refrain from sterilizing gold inflows (Simmons 1994). The literature about multilateralism has moved toward a similar consensus: multilateral institutions are a particular species that is non-discriminatory by design (Ruggie 1993). Some authors argued that multilateral institutions arise because they are more efficient than discriminatory regimes in some issue areas (Martin 1992), others that cooperation would be broader if discrimination were more widespread (Oye 1992), and still others that multilateralism generally works only in small groups of intensive cooperators (Kahler 1992).

Our model joins this discussion by allowing the choice of discriminatory or non-discriminatory (public-goods) regimes to be endogenous. States can choose whether to cooperate in a public-goods regime or to exclude non-cooperating countries from the enjoyment of those goods, at some cost. This modeling choice implicitly assumes that all of the products of international cooperation are really excludable, but that excluding free riders from benefiting from them is costly. The paradigmatic example of this argument, again, is the GATT/WTO trading system, which creates public externalities where they might otherwise not exist. This makes free riding attractive, but the cooperating states generally find that the benefits of generalized MFN status outweigh the costs. The parties could negotiate bilateral trade treaties with each other, but the number of transactions required increases geometrically with the number of parties, and bargaining is costly and time consuming. They balance the potential gain of additional cooperation from the current free-riders under the discriminatory regime against the cost of monitoring and cooperating in less efficient ways that minimize positive externalities. In fact, as Oye (1992) points out, the choice is not so clear cut: the trading system has always been a mixture of non-discriminatory and discriminatory forms, with free trade areas, Super 301 cases, and bilateral trade and investment treaties coexisting with the non-discriminatory rules of the GATT/WTO system. For the purposes of our model, we make the simplifying assumption that the regime must be either discriminatory or non-discriminatory. We will sometimes refer to discriminatory regimes as bilateral and non-discriminatory regimes as multilateral because this is a mechanism of limiting positive externalities that is familiar from the trade issue area, but the concept that is modeled is discrimination versus non-discrimination.

This generalizes the model beyond public goods, because it embraces private goods as well. For example, the case of pure private goods emerges when the cost of exclusion is very low. Furthermore, discrimination is an important complement to a theory about the trade-offs between breadth and depth, since a natural solution to the problem of non-contributing marginal members is to make the benefits of cooperation exclusive. For example, this is the intuition that underlay arguments for a "two-track" Europe. The benefits of Economic and Monetary Union accrue only to the participating states.

On the other hand, there are issue areas in which discrimination is not possible, or if possible, is not practical. An example is the global climate change regime: there is no feasible way of restricting the benefits of reducing CO₂ emissions to the cooperating countries. Similarly, one could argue that the benefits of the NATO alliance could not feasibly be withheld from countries that were not on the front line facing the Warsaw Pact (Olson and Zeckhauser 1966). As a result, France did not suffer a dramatic reduction in its security when De Gaulle withdrew it from formal participation in NATO military cooperation. There was no feasible way that the United States could threaten to abandon Monaco to the Soviet Union if it failed to contribute to the common defense. Our model can embrace cases like this as well, by assuming that the cost of exclusion is extraordinarily high. (It makes no difference in a formal model whether exclusion is technically impossible or simply very unattractive.) Thus, it is possible to interpret the parameter for the cost of exclusion in our model as a variable that captures the degree to which externalities are necessarily public in particular issue areas, and we will return to this interpretation below.

International Institutions

The origin of international institutions is one of the least well-understood issues in international political economy, although it has been a central theoretical concern since the emergence of the field (Keohane 1984; Keohane and

Nye 1977; Young 1994). We have a strong functionalist logic that holds that international institutions are constructed to facilitate collective action, yet the construction of these institutions is itself a feat of collective action that has to be explained. In the most persuasive surviving strand of hegemonic stability theory, Robert Keohane argues that international institutions are themselves the public good that hegemons provide gratis to the international community, and that they subsequently make cooperation—defined as mutual adjustment of policies—possible (Keohane 1984). In a darker vein, Robert Gilpin argues that international institutions are constructed by dominant powers and impregnated with their interests (Gilpin 1981).

We follow a rich tradition in international relations in assuming that institutions facilitate collective action by reducing transaction costs (Keohane 1984; Krasner 1983; Oye 1986). This notion, however, has never been formalized. We formalize institutions as investments, which, if made, will accelerate the pace of events in the game. This simulates the effect of reducing transaction costs: deviations from accepted norms are identified more rapidly, disputes are resolved more quickly, and coordinated efforts to impose sanctions can be organized more efficiently. Transaction costs arise from the searching, bargaining, and enforcement efforts that are necessary because incomplete information creates incentives for agents to misrepresent their interests or hide their actions (Coase 1960; Milgrom and Roberts 1992; Williamson 1985). In bargaining games, delay is the typical consequence (Rubinstein 1985). Institutions may accelerate the game by making norms explicit, so that deviations are more readily recognized; they may publish registers of deviant countries; or they may provide a judicial process for reaching authoritative rulings. The informal GATT (General Agreement on Tariffs and Trade) regime is an example of an institution that worked in precisely this way. GATT had no enforcement powers before the emergence of the WTO in 1995, but it facilitated international cooperation simply by clarifying expectations (Reinhardt 2001).

The Model

The international system consists of a hegemon, h , and n other countries of various sizes. Let $w_i > 0$ denote the size of country i and assume that $w_h > \max\{w_n\}$; that is, the hegemon is the largest country. Normalize the state sizes such that $w_h \leq 1$. The size of the group of potential contributors is $\mathcal{W} \leq \sum_n w_n$.³ Players have a common discount factor $\delta \in (0, 1)$ and act in discrete time with an infinite horizon and periods indexed by t ($t = 0, 1, 2, \dots$). In each period, the hegemon designates a contributing coalition of size $\mathcal{C} = \sum_i w_i \leq \mathcal{W}$ and chooses a level of contribution $a_t \geq 1$ ($a_t = 0$ if no contribution).⁴ This level represents the hegemon's demand of other contributing states, which then simultaneously choose levels of contribution $a_{i,t}$. This privileged position in the sequence of moves captures the notion that hegemons shape the rules of the international system (Pahre 1998).

In any period in which country i contributes at a_i , it produces an amount of public good proportional to its size, $q_{i,t} = w_i a_{i,t}^b$ with $0 < b < 1$, at a cost propor-

³ We allow for exogenous limitations on the size of the potential coalition, which may have to do with conflicts between states on other issues. Thus, for example, if the public good in question is the imposition of sanctions, the target country's allies are not considered to be potential contributors, but their capabilities are relevant to the effectiveness of sanctions.

⁴ This action space is designed to capture the substantively appealing notion that at very low levels, effort does not contribute much to the public good. As one of the referees suggested, this can be accomplished by redefining the production function to be S-shaped rather than concave, with the convex component between 0 and 1. Doing so will produce the same effect: zero contributions when nobody else is expected to contribute. It will be more convenient to keep the function as simple as possible mathematically for the analysis to follow.

tional to the demand, $a_{i,t}^c$ with $c > 1$, where the functional form captures the idea that there are diminishing production returns to higher contributions and increasing costs. The total amount of public good provided in that period is $Q_t = \sum_n q_{n,t} + q_{h,t}$. Let $Q_{-i,t} = Q_t - q_{i,t}$ denote the total contribution by countries other than i in period t . Country i 's per-period payoff from contributing is then:

$$\pi_i(a_{i,t}) = Q_{-i,t} + w_i a_{i,t}^b - a_{i,t}^c$$

States maximize the discounted stream of their per-period payoffs: $\sum_{t=0}^{\infty} \delta^t \pi_i(a_{i,t})$.⁵

We model the contribution to a public good as a non-cooperative, infinitely repeated game of perfect and complete information. Because Nash equilibrium may rely on incredible threats, the solution concept we use is subgame perfect equilibrium. Subgame perfection requires that player strategies are optimal at each point in the game whether or not that point is reached when players follow their equilibrium strategies. In other words, an equilibrium is subgame perfect if the strategies constitute a Nash equilibrium in every subgame.

Before presenting our equilibrium results, it is important to discuss equilibrium selection. The model presented here is infinitely repeated and the Nash equilibrium of the stage game is inefficient; consequently the Folk Theorem applies, and there exist an infinite number of subgame perfect equilibria that are Pareto superior to repeating the stage game Nash strategies. All of these equilibria have something of the same character: some (perhaps all) countries commit to contributing (at a level that may be efficient or inefficient), and if any of them defect they are subject to some punishment strategy. The question for the modeler is which of these equilibria are substantively interesting to study. Indeed, the question of equilibrium selection is just as important as the other key modeling choices, such as designating the actors' utility functions and choice sets. Our approach is to treat an equilibrium as a conjecture about a reasonable way to play the game (Kreps 1990). Thus, we prefer substantively realistic equilibria to unrealistic ones. Whether the restrictions we apply are reasonable depends upon their behavioral implications and what they require in the way of coordination. Do they generate behavior that is consistent with the world that we know? Can we imagine that rational actors might actually converge on playing the game in this particular way?

The most important choice that we make in this respect is that we restrict our attention to strategies that require all of the contributors to make proportional contributions, in a very special sense: Each state in the coalition contributes the same a . This can be interpreted as the amount of effort that states make to contribute to a public good, which is equal, although these efforts yield contributions of different sizes that depend upon each state's capabilities. Proportionality, in this sense, is very similar to what Olson and Zeckhauser (1966) describe as disproportionality, because they focus on the size of contributions rather than the amount of effort required to produce them. In their model, each state contributes at the point where the marginal benefit equals the marginal cost, and this leads larger states to make larger contributions. In our model, since the marginal benefits are identical for all states and only the marginal costs differ, this would be equivalent to requiring all states to exert the same amount of effort.

⁵ The results do not depend on choosing this particular functional form. In particular, we can derive the equilibrium constraints for arbitrary production, f , and cost, g , functions if we assume that $f' > 0$, $f'' < 0$, $g' > 0$, $g'' \geq 0$, and $w_n f(a) < g(a)$ for all a . These assumptions mean that (1) production is increasing in demand at a declining rate, (2) costs are increasing in demand at a non-declining rate, (3) unilateral production is costly even for the hegemon. If we also assume that $f/f' > g/g'$, then the constraints will be monotonic in a .

We argue that proportionality of effort is consistent with behavioral realism, because it is the almost universal practice of international agreements to set contributions according to general rules rather than country-specific quotas.⁶ Proportionality represents an intuitive conjecture about how expectations might converge, because it appeals to a norm of fairness that might form the basis for coordination. In addition, this restriction increases the realism of the model. We conjecture that, in a richer model with incomplete information, states would choose proportional contributions in order to reduce the inefficient bargaining over quotas.⁷ Consequently, restricting the equilibrium strategies allows us to incorporate a realistic element into the analysis that would make the model intractable if it were a built-in feature. It is not necessary for our qualitative results that the contributions be strictly proportional, but there must be some limits on the hegemon's ability to adjust the quotas to the circumstances of particular countries.⁸ We find strict proportionality attractive because it is simpler than any alternative rule.

Provision of a Public Good

As specified, the model is very flexible and admits a large variety of equilibria. However, since the environment is static (because state sizes do not change), it makes sense to focus on *stationary* equilibria; that is, equilibria in which the hegemon makes the same demand and the contributing coalition does not change across periods. The substantive idea behind such equilibria is that the hegemon and the contributing coalition form a "contract" initially, in effect agreeing on the level of contribution and the identity of contributors. Although in principle it is possible to change these parameters, in practice we should expect that doing so is both costly and detrimental to coordinating expectations around a consistent set of rules.

Arbitrary State Size Distributions

Without cooperation, no player conditions its behavior on any other player, so we can take the total amount contributed by players other than i as a given. This means that each player will rationally maximize the time-invariant stage game payoff: $\max_{a_i} \pi_i(a_i)$. Observe now that choosing to contribute implies

⁶ Contributions to international organizations such as the United Nations, the International Monetary Fund and the World Bank are based on formulas that roughly correspond to economic size. International trade agreements generally rely on proportional reductions in tariffs or application of across-the-board principles that impose proportional costs. Environmental treaties generally incorporate proportional reductions in emissions, proportional reductions in the size of fishing fleets, or proportional cost sharing. Debt restructuring agreements in the Paris Club and London Club rely upon additional contributions of financing that are proportional to the amount of principal already committed.

⁷ The Kyoto Protocol is an apparent exception, culminating in an agreement with diverse and apparently arbitrary emissions quotas. Upon closer examination, however, it fits our argument in two respects. First, the bargaining about the quotas took place through a meta-discussion about general rules: special exceptions were made to using 1990 as the reference year for CO_2 reductions for post-Communist countries, special credits were given for countries with extensive "carbon sinks" (i.e., forests and grazing land), etc. Second, the Kyoto Protocol is a case in point for our claim that bargaining over quotas is inefficient (McLean and Stone 2005).

⁸ This is not an innocent assumption, however: we impose this restriction because our substantive interest is to explain variation in the breadth of participation. There exist equilibria of this game in which the hegemon would receive a higher payoff by assessing country-specific contributions, and an optimal equilibrium with country-specific quotas would assure universal participation. We find these equilibria unrealistic because they involve universal participation, and substantively uninteresting because they do not shed any light on the variations in participation that we observe. Gilligan (2004) argues that uniform policies are required to generate a depth-versus-breadth trade-off. The proportional contributions that we use can be interpreted as uniform policies (i.e., 10% across-the-board emissions reductions, or banning the production of CFCs), but they can also be given a more general interpretation as proportional contributions. In any case, in our more general model, even strict proportionality is not necessary for the result.

$a_i \geq 1 \Rightarrow a_i^b \leq a_i^c$, and since $w_i < w_h \leq 1$, this means that $w_i a_i^b - a_i^c < 0$ for any $a \geq 1$. This means that contributing unconditionally even at the best possible level is worse than not contributing at all. Hence, the *unconditional optimum* for each state is to contribute nothing. This constitutes a Nash equilibrium of the repeated public-goods provision game as well.

Let $\mathcal{C} \subseteq \mathcal{W}$ denote the size of the cooperating coalition that consists of states that contribute to the public good at the level demanded by the hegemon. We now look for a stationary equilibrium where the hegemon designates a coalition of size \mathcal{C} and demands a contribution a from each member. The first regime we examine is where the hegemon collaborates with other states to provide a pure public good. A public good has the characteristics that it is non-rival and non-excludable. In other words, one country's benefit from the good does not decrease another's, and it is not possible to exclude any country from enjoying it. Under this regime, the only way to punish a defector is by halting production of the good entirely. We define two phases of the game, a *cooperative* phase and a *punishment* phase, and define strategies in terms of these phases. Consider the following strategies:

- Hegemon: pick the level of contribution a , designate the contributing coalition \mathcal{C} , and begin the cooperative phase. In the cooperative phase, demand a ; if any member of the coalition contributes less than a , switch to the punishment phase in the next period. In the punishment phase, demand 0 for $T \geq 1$ periods, then return to the cooperative phase.
- Members of the coalition: in a cooperative phase contribute a , and in the punishment phase contribute nothing.
- Both hegemon and members: If the hegemon ever chooses $\hat{a} \notin \{0, a\}$ or ever readmits a deviating member to the coalition without punishing it for T periods, revert to the unconditional Nash equilibrium forever.
- Non-members: contribute nothing.

Note first that the strategy for non-members is optimal: not contributing is Nash and no other players condition their behavior on what non-members do. All non-members free-ride on the public-good provision by the coalition.

Turning now to the members' strategy, note that although each state can choose different levels of contribution, in equilibrium they either contribute a or nothing. This is because contributing is costly, so one would never contribute more than the minimum required, and since the hegemon punishes contributions smaller than a as defections, a state will never defect by making a positive contribution. Given a demand a , a coalition of size \mathcal{C} will provide $Q = (\mathcal{C} + w_h) a^b$ of the public good, so $\pi_i(a) = Q - a^c$.

We now examine the condition that would induce members to contribute. Failing to contribute at least a results in T periods of punishment with zero contributions by coalition members. The best possible deviation is to contribute nothing, so coalition member i would prefer to contribute if, and only if,

$$\frac{Q - a^c}{1 - \delta} \geq Q - w_i a^b + \frac{\delta^{T+1}(Q - a^c)}{1 - \delta}.$$

After simplifying and rearranging terms, this yields the *marginal contributor's constraint*:

$$w_i \geq \left(\frac{1}{1 - \delta}\right) [(1 - \delta^{T+1}) a^{c-b} - (\delta - \delta^{T+1})(\mathcal{C} + w_h)] \equiv \text{MC}(a, \mathcal{C}). \tag{MC}$$

That is, a state must be at least as large as $MC(a, C)$ to contribute to the public good if designated a member of the coalition. Observe now that

$$\frac{\partial MC(a, C)}{\partial a} = \frac{(1 - \delta^{T+1})(c - b)a^{c-b-1}}{1 - \delta} > 0.$$

That is, the more the hegemon demands, the higher the size of the marginal contributor has to be. This leads to our first important result, which is that there is a general trade-off between the breadth and depth of cooperation. The marginal contributor decides whether to participate in the coalition by comparing the benefits it gets from contributing with the benefits it gets from free riding for one period, and then being punished.

The trade-off exists in our model because the benefit from contributing depends upon the size of the contribution the country is able to make for a given level of effort, but the benefit of free riding does not. If its own contribution makes very little difference to its utility, the cost of contributing looms large in its calculations, and it does better by defecting. Conversely, since the benefits of contributing increase with size and the benefits of free riding do not, larger states break even by contributing at higher levels of contributions than smaller states. For any given level of contribution required to participate in the coalition, therefore, countries rank themselves according to size, and every country above a certain size participates, while all below that size defect. Because marginal costs rise faster than marginal benefits as the size of the required contribution increases, the minimum size of the marginal contributor increases along with the size of the contribution demanded.

We now turn to the punishment strategy. Because everyone is expected to contribute nothing while punishment lasts, no country has an incentive to contribute anything unilaterally. Furthermore, if the hegemon ever demands some $\hat{a} \notin \{0, a\}$ or attempts to readmit a deviating member without punishment, then any country i expects the coalition to fall apart because nobody else will contribute, so it has an incentive to stop contributing as well. Only the hegemon may potentially have a profitable alternative strategy: instead of demanding no contributions (and starting a punishment phase), it can continue to demand a but exclude the deviating contributor from the coalition forever. Given the strategy of the contributors, this will not trigger the breakup of the remaining coalition unless the hegemon readmits the deviating member without punishment.⁹ Clearly, there is no incentive to delay punishment, so the hegemon will not deviate by excepting the defector from the coalition until some time t' and then punishing it for T periods to restore cooperation. This is because in all periods prior to t' the contributing group will be getting strictly worse payoffs than under full cooperation that punishment would restore. This means that the hegemon is better off imposing the punishment sooner rather than later. Hence, the only potentially profitable deviation is for the hegemon to exclude the deviating member forever and continue demanding a . In other words, the hegemon must be willing to halt production when faced with free-riding. The threat to do

⁹ We have in mind a type of "renegotiation" in which a violation of the implicit agreement is ignored but since the rules require costly punishment, the other members can avoid having to suffer if they simply exclude the offender thereby obviating the need to punish. This is a more demanding requirement than the usual trigger strategy which dissolves cooperation upon deviation. Substantively, it is also more appealing to require that institutions should be immune to some attempt to manipulate them. On the other hand, they should not be infinitely flexible: if the hegemon tries to undermine the rules twice (once by excluding a deviating member and again by trying to readmit it without punishment), the trigger strategies come into play and the institution is dissolved. Arguably, an institution that can be manipulated with such impunity is not worth its name. Although in some sense our solution will be immune to the aforementioned renegotiation, it is not fully renegotiation proof. We are not sure that requiring full renegotiation-proofness is substantively plausible as we do not believe that institutions can always be adjusted to survive all changes in the environment in which they arose.

so is credible only if it prefers to punish the deviating contributor for T periods instead of continuing production and allowing it to drop out of the coalition. Observe that if this threat is not credible for some member of the coalition, then this member will drop out and free ride on the efforts of others because it knows that production will continue. Since production must continue when the hegemon excludes the deviating member, the following condition must hold for any member k when some other member i drops out:

$$\delta^T \pi_k(a) \geq \pi_k(a) - w_i a^b,$$

which is satisfied if, and only if, $w_i \geq (1 - \delta^T)(\mathcal{C} + w_h - a^{c-b})$. This means that if the hegemon can credibly threaten to punish the deviations of the marginal contributor, then it can credibly threaten all other members of the coalition. This yields the hegemon's *credibility constraint*:

$$w_i \geq (1 - \delta^T)(\mathcal{C} + w_h - a^{c-b}) \equiv \text{CR}(a, \mathcal{C}). \tag{CR}$$

Observe that $\partial \text{CR}(a, \mathcal{C}) / \partial a = -(c - b)(1 - \delta^T)a^{c-b-1} < 0$. That is, the more the hegemon demands, the smaller the size of the smallest state it can credibly threaten with punishment. In other words, the higher the demand, the easier it is to maintain the coalition. We now look for conditions that will establish the existence of a subgame-perfect equilibrium (SPE) in the strategies specified in this section.

Let \underline{w} be some existing state size. Define $\mathcal{C} = \sum_i w_i$, where $w_i \geq \underline{w}$, is the size of the coalition consisting of all states at least as large as \underline{w} . Let \mathcal{S} denote the set of all such coalitions. For any such coalition $\mathcal{C} \in \mathcal{S}$, $\underline{w}(\mathcal{C})$ denotes its smallest member. A demand-coalition pair, (a, \mathcal{C}) , is a *profile*. A profile is *feasible* if $\underline{w}(\mathcal{C}) \geq \text{MC}(a, \mathcal{C})$, that is, if all members of the coalition satisfy the contributor's constraint at the level demanded. A feasible profile that also satisfies the credibility constraint at that level, that is, $\underline{w}(\mathcal{C}) \geq \text{CR}(a, \mathcal{C})$, is *admissible*. The hegemon chooses among the admissible profiles the one that maximizes its payoff. Let \mathcal{A} denote the set of admissible profiles. The following lemma shows the conditions under which \mathcal{A} is non-empty (all proofs are in Appendix A).

Lemma 1: *For any $\mathcal{C} \in \mathcal{S}$, define the minimum required contributor size as:*

$$\tilde{w}(\mathcal{C}) = \begin{cases} (1 - \delta)\text{X}(\mathcal{C}) & \text{if } \text{X}(\mathcal{C}) \geq 1 \\ \text{MC}(1, \mathcal{C}) & \text{otherwise,} \end{cases}$$

where $\text{X}(\mathcal{C}) = (1 - \delta^T)(\mathcal{C} + w_h) / (2 - \delta - \delta^T)$. The set of admissible profiles, \mathcal{A} , is non-empty if, and only if, $\underline{w}(\mathcal{C}) \geq \tilde{w}(\mathcal{C})$ for some $\mathcal{C} \in \mathcal{S}$.

An important consequence of Lemma 1 is that if there exists some coalition such that $\mathcal{C} + w_h > (T + 1)/T$, then there always exist admissible coalitions for high enough δ . To see this, note that $\lim_{\delta \rightarrow 1} \text{X}(\mathcal{C}) = T(\mathcal{C} + w_h) / (T + 1)$, and so $\mathcal{C} + w_h > (T + 1)/T$ means that for high enough δ , the two constraints will intersect. But then $\lim_{\delta \rightarrow 1} \tilde{w}(\mathcal{C}) = 0$, and so any such coalition will be admissible. We are now ready to state our first result formally.

Proposition 1: *Suppose $\mathcal{A} \neq \emptyset$ and let $(a^*, \mathcal{C}^*) = \text{argmax}\{\pi_h(a, \mathcal{C}) : (a, \mathcal{C}) \in \mathcal{A}\}$ be the admissible profile that maximizes the hegemon's payoff. The contribution game has a stationary SPE in which the hegemon chooses (a^*, \mathcal{C}^*) and the players use the strategies specified in this section.*

It is worth working through an example that shows how Lemma 1 determines whether $\mathcal{A} \neq \emptyset$ and how Proposition 1 establishes the SPE solution to the contribution game.

Example 1: Parameters: $b = 0.80$, $c = 1.05$, $T = 2$, $\delta = 0.99$, and $w_h = 0.86$. We shall examine three state systems, each with $\mathcal{W} = 2.04$ and consisting of five states in addition to the hegemon. Let $a^*(\mathcal{C})$ denote the demand that maximizes π_h for the admissible profile $(a^*, \mathcal{C}) \in \mathcal{A}$. Table 1 shows all possible coalitions that consist of all states larger than the smallest member. The third column shows $\tilde{w}(\mathcal{C})$ using the definition from Lemma 1. This establishes the minimum required contributor size for a coalition to be admissible. The highlighted row is the SPE solution for a particular distribution.

In this example, we hold the size of the world constant and vary the distribution of state sizes. Because both constraints are harder to satisfy for the smaller potential contributors, we concentrate on variations among small and medium-sized states. The purpose is to demonstrate how feasible and admissible coalitions vary in the distribution of state sizes and how the constraints limit what the hegemon can demand from the optimal coalition. Even this simple example shows why it is impossible to make general statements about the dynamics of the SPE solution. We have selected three possible state size distributions, each with the same \mathcal{W} . The first case exhibits the broader-deeper trade-off persistently: to maintain the larger coalition, the hegemon must decrease its demand. The second case exhibits the trade-off only intermittently: increasing the coalition size from, say 0.85 to 1.60 involves a corresponding jump in optimal demand from 2.88 to 12.34, while increasing the coalition size from 2.02 to 2.04 involves a decline in optimal demand from 14.98 to 13.89. Finally, the third case does not exhibit the trade-off anywhere: increasing the coalition size always involves a corresponding increase in the optimal demand. (Note also that it is not necessarily true that the hegemon will always choose the widest possible coalition.)

Monte Carlo Simulations

If we examine the distribution of states in Table 1, it is readily noted that the trade-off “kicks in” whenever the hegemon wants to include a relatively small marginal contributor in the coalition—it is these states that are most difficult to keep in cooperative mode. Correspondingly, our focus would be on simulating

TABLE 1. Admissible Demand-Coalitions for Three Distributions

<i>Smallest Member, $\underline{w}(\mathcal{C})$</i>	<i>Coalition Size, \mathcal{C}</i>	<i>Minimum Contributor Size, $\tilde{w}(\mathcal{C})$</i>	<i>Optimal Demand, $a^*(\mathcal{C})$</i>	<i>$MC(a^*)$</i>	<i>$CR(a^*)$</i>	<i>Hegemon Payoff, $\pi_h(a^*)$</i>	<i>Total Quantity, $Q(a^*)$</i>
World: 0.02, 0.07, 0.25, 0.85, 0.85							
0.02	2.04	0.0193	13.89	0.02	0.0193	7.955	23.79
0.07	2.02	0.0192	13.99	0.07	0.0188	7.808	23.77
0.25	1.95	0.0187	14.40	0.25	0.0172	7.281	23.74
0.85	1.70	0.0170	14.47	0.75	0.0121	5.169	21.71
World: 0.02, 0.17, 0.25, 0.75, 0.85							
0.02	2.04	0.0193	13.89	0.02	0.0193	7.955	23.79
0.17	2.02	0.0192	14.98	0.17	0.0182	7.958	25.12
0.25	1.85	0.0180	12.53	0.25	0.0165	6.263	20.49
0.75	1.60	0.0164	12.34	0.72	0.0117	4.373	18.37
0.85	0.85	0.0114	2.88	0.50	0.0081	0.949	3.99
World: 0.12, 0.17, 0.25, 0.75, 0.75							
0.12	2.04	0.0193	14.88	0.12	0.0186	8.116	25.15
0.17	1.92	0.0185	13.07	0.17	0.0175	6.867	21.73
0.25	1.75	0.0174	10.87	0.25	0.0158	5.356	17.59
0.75	1.50	0.0157	10.45	0.69	0.0112	3.673	15.43

various distributions in which small states either dominate or are about as frequent as others.

The beta distribution, with its flexible (α, β) parameters, allows us to simulate international systems with various degrees of small state presence. By keeping $\alpha = 1$, we can obtain distributions ranging from the uniform ($\beta = 1$), where all state sizes occur with equal frequency, to the fragmented ($\beta = 3$), where there are many small states and very few large ones. We shall take $M = 25$ values of β evenly distributed between 3 and 1, which gives us M state system types.

Because we wish to isolate the effect of increasing the size of the hegemon from overall wealth, we will keep \mathcal{W} constant throughout, just as we did in Example 1. This now means that if we keep the number of states per system constant as well, the variation in sizes between simulations will be too small (because the same number of sizes have to sum up to a constant). To overcome this problem, we model fragmented systems with $N = 18$ states and steadily decrease that number as systems become more consolidated until we reach $N = 10$ for uniform systems.¹⁰

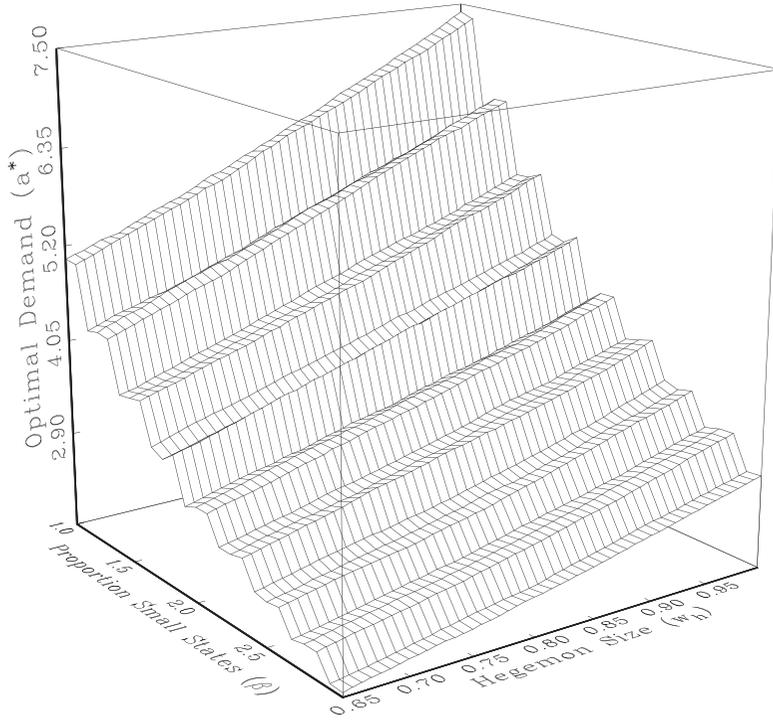
For each system type, we randomly draw $K = 2,000$ state systems, and for each of these, we compute the SPE (a^*, C^*) for $L = 50$ hegemon sizes that vary between 0.65 and 1. Since we have fixed the wealth of potential contributors at $\mathcal{W} = 2$, this means that the hegemon's size is between a third and a half of it. This gives us K estimates per hegemon size that we can use to compute the corresponding mean optimal demand-coalition and confidence intervals. To avoid cluttering the figures, we plot the means only—the other quantities can be easily computed with the replication package.

Figure 1 shows the SPE demand-coalition results for highly fragmented systems in which small states predominate to relatively consolidated systems in which fewer larger states are evenly distributed. The parameters are $b = 0.90$, $c = 1.15$, $T = 5$, and $\delta = 0.85$. The most important conclusion from these simulations is that *overall, as the hegemon grows, the depth of the contribution (amount demanded by the hegemon) increases, but the breadth of cooperation (size of contributing coalition) decreases because the size of the marginal contributor increases. That is, larger hegemons demand more but control smaller coalitions.*

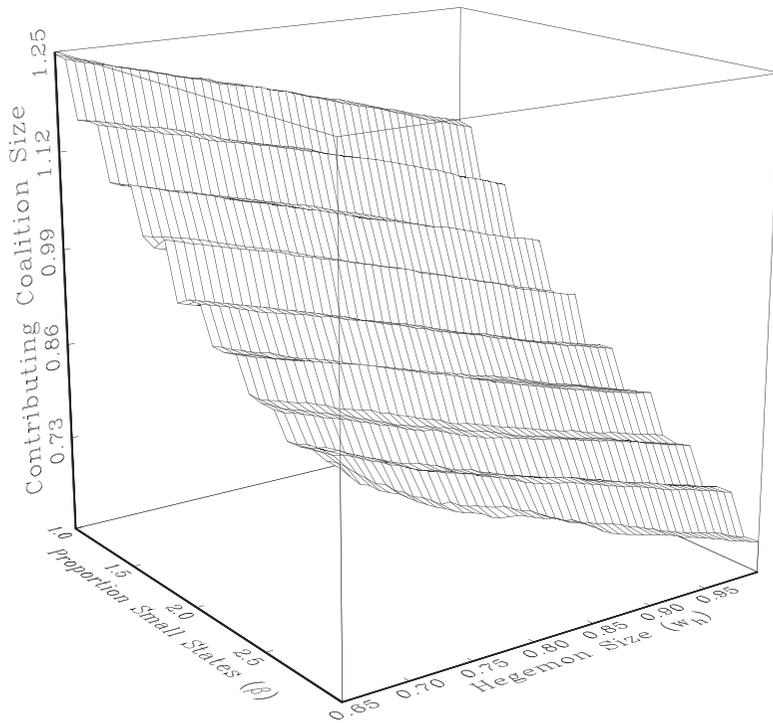
Observe now that in highly fragmented systems, the optimal demand is much smaller than the corresponding demand in consolidated systems of same wealth (this is easily seen in Example 1 by comparing $a^* = 13.89$ for the first, most fragmented system, with $a^* = 14.88$ for the third, the most consolidated one). The reason is that in the fragmented systems, the hegemon must keep small marginal contributors in the coalition, which means that it cannot demand too much. This is no longer a problem where all contributors are roughly similar in size and there are few of them (which means they are larger as well), and so the hegemon can demand much more in consolidated systems.

This effect is present regardless of the size of the hegemon, although, it does weaken as the hegemon grows: the ratio of a^* in a fragmented to a consolidated system is approximately 1:2.80 for a small hegemon and 1:2.59 for a large hegemon. In other words, the “penalty” the hegemon must pay for living in a fragmented world is lower if the hegemon is larger. In absolute terms, the escalation of demands by larger hegemons can be quite dramatic: from 5.04 to 7.40 in consolidated systems and 1.80 to 2.86 in fragmented ones. However, even though the increase is more substantial in the former in absolute terms, it is more pronounced in the latter in relative terms. The ratio of a^* by a small to a large hegemon is 1:1.47 when the system is consolidated, and 1:1.59 when the system is fragmented. That is, as a percent of the small hegemon's demand, the larger hegemon's demand is higher in fragmented systems.

¹⁰ Technical details can be found in the replication package.



(a) Optimal Demand.



(b) Optimal Coalition.

FIG. 1. Simulation Results for Optimal Demand and Coalition Size

This now implies that although the size of the hegemon dramatically affects both optimal demand and coalition choice, it is relatively more important in fragmented systems: that is, larger hegemon's escalate their demands more substantially when they face many small contributors and the decline in coalition size will be more pronounced. Note, for example, that the coalition size drops from 0.88 to 0.64 in fragmented systems (a ratio of 1.37:1), and from 1.25 to 1.09 in consolidated ones (a ratio of 1.15:1). It is worth noting, however, that as a percentage of potential contributors' total wealth, the cooperating coalition is significantly larger in consolidated systems (between 54% and 62%) than in fragmented ones (32% and 44%). The fact that the hegemon does not have to accommodate very small marginal contributors in consolidated systems enables it to extract more from fewer states that account for a larger share of world wealth.

The general conclusion is that as the hegemon's size increases, the level of contribution increases as well, but the size of the contributing group decreases. The exact relationship between size and optimal contribution is nonlinear and depends on the distribution of potential contributor sizes.

Analytical Results

The simulations demonstrate why a general analytical result that holds for arbitrary distributions of state sizes is impossible. However, we are interested in investigating comparative statics that require analytical derivations. To this end, we introduce a simplifying assumption that captures the dimension of variation in the elasticity of coalitions while preserving tractability:

Assumption 1: Let $\mathcal{C} = \mathcal{W} - \epsilon \underline{w}$, where $0 < \epsilon < \bar{\epsilon} = (1 - \delta)/[\delta(1 - \delta^T)]$.

In this specification, ϵ measures how strongly \mathcal{C} responds to changes in \underline{w} . This is the simplest functional form that captures the fact that \mathcal{C} decreases as \underline{w} increases, and it has the mathematical virtue that \mathcal{C} is a continuous function of \underline{w} . The ceiling on ϵ is necessary to ensure that the marginal contributor constraint derived under this assumption behaves analogously to the general form in (MC). The hegemon's choice can now be expressed as (a, \underline{w}) , which is equivalent to the demand-coalition pair $(a, \mathcal{W} - \epsilon \underline{w})$.

In equilibrium, \underline{w} must satisfy both (MC) and (CR). Rewriting the constraints under Assumption 1 yields:

$$\underline{w} \geq \frac{(1 - \delta^{T+1})a^{c-b} - (\delta - \delta^{T+1})(\mathcal{W} + w_h)}{1 - \delta[1 + \epsilon(1 - \delta^T)]} \equiv \text{MC}_1(a),$$

where we used $\epsilon < \bar{\epsilon}$, and:

$$\underline{w} \geq \frac{(1 - \delta^T)(\mathcal{W} + w_h - a^{c-b})}{1 + \epsilon(1 - \delta^T)} \equiv \text{CR}_1(a).$$

We use the subscript on $\text{MC}_1(\cdot)$ and $\text{CR}_1(\cdot)$ to denote the idea that these constraints are specific to the functional form implied by Assumption 1. Observe now that this functional form has preserved the properties of the constraints from the original specification: $\text{MC}_1(a)$ is strictly increasing in a while $\text{CR}_1(a)$ is strictly decreasing.¹¹

¹¹ This is where the ceiling on ϵ comes into play: if ϵ is too large, then the denominator of $\text{MC}_1(a)$ will be negative, and so $\text{MC}_1(a)$ will decrease in a , contradicting the property of the original specification.

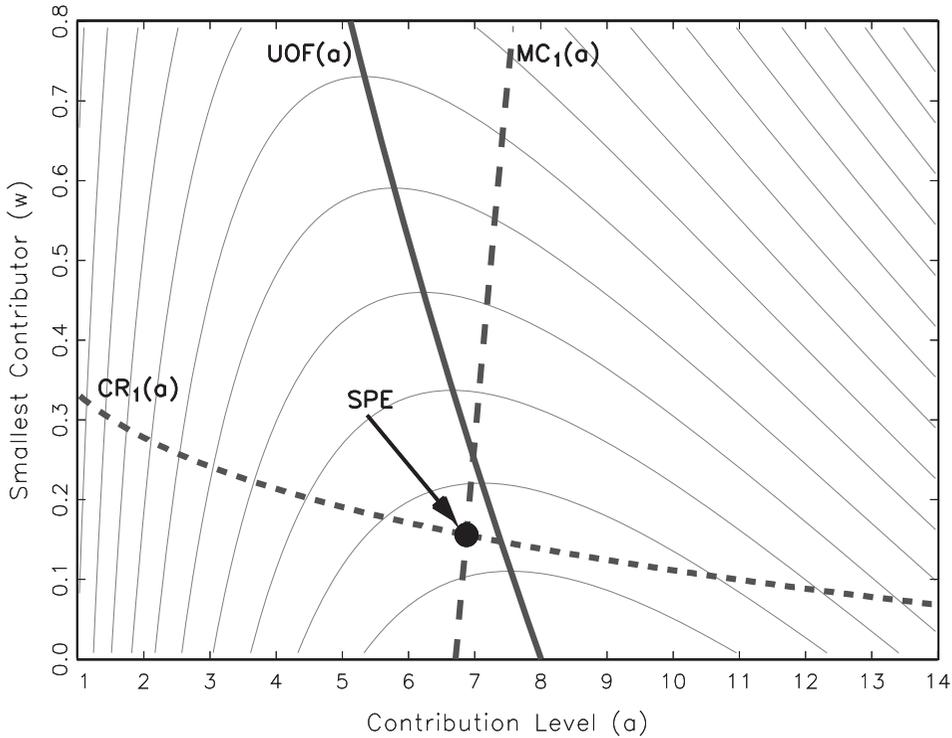


FIG. 2. Both Constraints Binding at the Solution

Proposition 2: *The restricted version of the game has a stationary SPE in which the players use the trigger strategies and the hegemon chooses (a^*, \underline{w}^*) such that $\underline{w}^* = MC_1(a^*)$ if c/b is sufficiently small, $\underline{w}^* = CR_1(a^*)$ if c/b is sufficiently large, and $\underline{w}^* = MC_1(a^*) = CR_1(a^*)$ if c/b is between the two extremes. Both equilibrium demand and smallest contributor size are strictly increasing in the size of the hegemon.*

If the hegemon were not bound by the constraints, it would maximize $\pi_h(a, \underline{w})$ to obtain $\hat{a}^{c-b}(\underline{w}) = b(\mathcal{W} - \epsilon \underline{w} + w_h)/c$. Solving this for \underline{w} gives us the smallest contributor choice from the *unconstrained optimum function* (UOF) of the hegemon’s demand: $\underline{w}^*(a) = (\mathcal{W} + w_h - ca^{c-b}/b)/\epsilon$. It is worth noting that the UOF is maximized at $\underline{w} = 0 \Rightarrow \hat{a}^{c-b} = b(\mathcal{W} + w_h)/c$, which implies that in equilibrium no demand will exceed that level. Figure 2 shows one example where both constraints are binding at the solution and its relationship to the UOF.¹² The hegemon’s indifference curves are convex because it is trying to minimize \underline{w} in order to maximize the size of the contributing coalition. Hence, higher payoffs are found at lower indifference curves, and the best unconstrained payoff is at the intersection of the UOF and the horizontal axis. Naturally, SPE payoffs will be smaller because the constraints do not permit the hegemon to minimize \underline{w} all the way down to zero.

Consistent with the trends from the simulations, as the hegemon grows, the depth of the contribution (amount demanded by the hegemon) increases, but the minimum size of the marginal contributor increases as well. This implies that the participation rate falls as hegemony increases in size, and that the participation rate increases as the hegemon declines. The close correspondence between

¹² The parameters for this case are: $b = 0.85$, $c = 1.08$, $T = 4$, $\mathcal{W} = 1.25$, $\epsilon = 0.25$, $w_h = 0.8$, and $\delta = 0.90$. Since $K_1 \approx 1.05 < c/b \approx 1.27 < K_2 \approx 1.32$, both constraints must be binding. The solution is $(a^*, \underline{w}^*) \approx (6.85, 0.16)$.

these analytical results and the general ones obtained from the simulations gives us confidence that the simplification is not too distorting.

The analytical solution allows us to characterize more precisely the situations in which the hegemon's credibility constrains cooperation. Since which case pertains depends only on the cost and benefit parameters, the three cases represent different kinds of issues, which are characterized by different relationships between the cost and benefit of international cooperation. In some issue areas, the ratio of benefits to costs is so favorable that the optimal degree of cooperation from the hegemon's point of view is extremely high. These issues represent the first case of the proposition: Since the hegemon desires very intense cooperation, its credibility constraint is not binding. When the benefits of cooperation are so great that the hegemon is willing to make an extremely large contribution, the issue of credibility simply does not arise. For example, the United States was more successful in organizing sanctions against the USSR after the invasion of Afghanistan (1979), when it imposed a politically costly grain embargo, than after the repression of Solidarity in Poland (1981), when it was unwilling to do so. In the second case, the United States threatened to impose sanctions on European subsidiaries of U.S. firms that did business with the Soviet Union, but backed down when the Europeans called its bluff (Martin 1992).¹³

In other issue areas, where the cost of cooperation is high relative to its benefits, both the contributors' willingness to participate and the hegemon's ability to commit to punishing deviations represent binding constraints (the second case in the proposition). We believe that this represents the majority of cases of international cooperation, where the costs are neither prohibitive nor trivial. Finally, when the costs of cooperation become exceedingly high relative to the benefits, the credibility of the punishment threat becomes the focal issue, and the states that the hegemon can credibly keep in the contributing coalition are necessarily much larger, which means that their contributor constraint is slack. Cooperation is expected to be quite limited in this case, if it emerges at all. An example is exchange-rate coordination. The United States has been unwilling to make significant changes in macroeconomic policy in order to influence the dollar's exchange value ever since the demise of the Bretton Woods fixed-rate system. As a result, agreements such as the Plaza Accord in the 1980s simply ratified what countries already intended to do (Oye 1992). Similarly, current efforts to convince U.S. trade partners to revalue their currencies are unconvincing when the United States is unwilling to change its own policies in order to devalue.

We feel that it is the second set of issues, with cost-benefit ratios in the middle range, that are of most interest in international relations—issues where cooperation is costly, and therefore difficult to achieve, but nevertheless valuable. In these cases, both the marginal contributor's willingness to cooperate and the hegemon's willingness to punish are in question, and both of these constraints are routinely tested in international relations. Consequently, the rest of this article will focus on the results for the case where both constraints are binding.¹⁴

¹³ Martin argues that willingness to bear costs is associated with credible threats to enforce cooperation, but the reason is that bearing costs sends a signal about the organizer's "type," which is an argument about incomplete information. We make the same prediction using a full information model, and our logic is different: hegemon's whose coalitions are not constrained by considerations of credibility are those that want to bear high costs. One way to distinguish between the predictions of the two models would be to ask why hegemon's bear high costs: in order to signal resolve, or because they find high contributions to be optimal? An interpretation of the case consistent with our argument is that the Carter administration chose to impose the embargo because it categorically preferred a vigorous response, not because it was attempting to encourage U.S. allies to cooperate in imposing sanctions.

¹⁴ Almost all of our results generalize to all three cases. The results about institution building, however, require that the marginal contributor constraint be binding (the credibility constraint may be binding or slack).

Multilateralism and Discrimination

We now extend our basic model into the area of private goods, which will allow us to use the model to explain discrimination. We focus on a regime in which, instead of halting cooperation as a punishment for free-riders, the contributors pay an exclusion cost and thus ensure that only other contributors can enjoy the benefits of cooperation. Like the public-goods case analyzed above, this is an ideal type. Actual cooperative arrangements combine elements of discrimination and externalities to non-participants that cannot be controlled. However, we focus on the extreme cases as a strategy for casting light on the ways in which the design of empirical schemes for cooperation makes trade-offs between incentive compatibility and efficiency. One of the hidden costs of free riding is that cooperation is constrained in ways that make it less attractive than it could otherwise be.

In our model, there is no collective action problem under the discriminatory regime. Non-contributors get zero utility under the discriminatory regime, and contributors get strictly positive utility, so every state contributes under the discriminatory regime. Furthermore, the credibility constraint is also satisfied regardless of the deviating state's size: excluding that state for T periods and then resuming cooperation is strictly better than excluding it forever, so the hegemon is never tempted to be lenient. The hegemon can now maximize its utility without being constrained by the contributors.

Let $m > 0$ be the cost of exclusion. The hegemon's optimization problem is

$$\max_a \{ (\mathcal{W} + w_h) a^b - a^c - m \},$$

which has a solution

$$a_D^* = \sqrt[c-b]{\frac{b(\mathcal{W} + w_h)}{c}},$$

where the subscript on the optimal demand indicates the discriminatory regime. We can now establish the following result:

Proposition 3: *If the exclusion cost is sufficiently low, the hegemon strictly prefers the discriminatory regime to the public-goods regime. Furthermore, for any given exclusion cost, there exists a size threshold such that hegemons above that size will strictly prefer to discriminate.*

Our basic finding about multilateralism and discrimination, then, is that larger hegemons prefer more discriminatory regimes. This happens for two reasons. First, larger hegemons prefer more intensive cooperation, which leads to narrower coalitions and heavier losses from free riding. Second, as the hegemon grows, its threats of punishment become less credible, because the contribution of marginal states to its utility becomes less and less important relative to its own policy. This means that the coalition of contributors shrinks, and the benefits that the hegemon can gain from international cooperation under the public-goods regime decline. If the hegemon imposes a discriminatory regime, however, only contributors can enjoy the benefits of cooperation, so cooperation becomes more widespread; under the simplifying assumptions of our model, it becomes universal. As the hegemon grows, the point at which the benefits of broader cooperation outweigh the efficiency losses associated with discrimination draws closer. The largest hegemons prefer the private-goods regime.

Again, this finding contradicts a basic contention of the hegemonic stability theory: that hegemony is associated with openness in the international system (Keohane 1984; Krasner 1976). It is much more consistent with the empirical record, however, which shows that the high point of British power in the nine-

teenth century coincides with increasing bilateralism rather than unconditional free trade (Ruggie 1993). It explains the apparent anomaly of the American policy in the 1930s, when the newly dominant industrial power pursued a policy of bilateralism in trade and finance. It helps to explain why the gradual erosion of U.S. power after World War II was accompanied by a broadening of international cooperation in trade, finance, and a variety of other issue areas. The anomaly to be explained, from this perspective, is not the persistence of cooperation after hegemony, but the burst of multilateralism in the immediate post-war years, which set the stage for the expansion of cooperation that followed.

The explanation for multilateralism after the Second World War, from the perspective of this model, is that the U.S. chose a multilateral strategy because the cost of discrimination was prohibitive. Indeed, the policymakers who were “present at the creation” told us clearly why they acted as they did: to balance the emerging power of the Soviet Union (Acheson [1969] 1987; Kennan 1951, 1969). The United States was willing to forgo remarkable opportunities to exploit the spoils of war and bear extraordinary costs to rebuild Europe and Japan because it faced a threatening adversary. Multilateral regimes to manage trade, international finance, and the exchange-rate system were seen as the natural complements—indeed, as the essential economic infrastructure—of the system of alliances that the United States constructed after the war to counter the ambitions of its erstwhile ally and new rival. In a competition that was conceived from the beginning as a test of alternative economic models, ultimate success was expected to rest more upon superior economic performance than upon military force, so economic institutions were called for that would maximize the economic performance of U.S. allies. In the furor of the early Cold War, questions of economic policy were naturally subordinated to those of sound security strategy. In short, discrimination was rejected—as tempting as it was in economic terms—because the security costs were prohibitive.

International Institutions

In this section, we consider the possibility of building institutions. International institutions can explain cooperation once they have been established, but how are we to overcome the collective action problem posed by the need to create these institutions in the first place? A fundamental claim of the theory of hegemonic stability was that hegemons are responsible for building institutions (Keohane 1984). Consequently, we formalize the claim that institutions facilitate collective action by reducing transaction costs, and we make it possible for hegemons to choose to build these institutions, at some cost.¹⁵ We limit our attention to institutions for the public-goods regime, because in our model monitoring and enforcement are only necessary in the public-goods regime. We do not mean to imply that international institutions are in fact non-discriminatory, but we do claim that institutionalization and discrimination represent alternative strategies to deal with free riding. Thus, for example, the international trading system faced a crossroads as it completed the Uruguay Round, which could have led to an intensification of bilateral discrimination had it not led to the creation of the WTO. The current stalemate in the Doha Round appears to be leading to greater emphasis on free trade areas and bilateral trade and investment treaties.

¹⁵ For simplicity, we assume that the hegemon bears the full cost of building institutions, but none of the results depend upon this. Regardless of how the cost is paid, the results follow from the assumptions that building institutions implies paying a fixed cost and reduces the difficulty of collective action.

We modify the model so that players are allowed to move (i.e., change their actions) only every other period, but utility continues to accrue every period. If a state chooses to produce a^* , for example, its choice takes effect for two periods. We ask under what circumstances a hegemon will be willing to pay for the creation of an institution that will allow players to interact every period instead of every other period.¹⁶ In this section, we assume that the hegemon chooses the public-goods regime regardless of whether it chooses to build an institution. Consequently, our conclusions hold when the cost of exclusion is prohibitively high, or, for the set of hegemon, that are too small to prefer discrimination when the exclusion cost is low.

Proposition 4: *The optimal demand without the institution is smaller than the optimal demand with the institution. Moreover, the minimum size of the marginal contributor without the institution is larger than the corresponding size with the institution.*

This result leads to our first conclusion about institutions: creating institutions broadens and deepens cooperation. The effect of institutions, in our model, is to reduce the number of periods in which a non-contributor can free ride before being identified and punished. Since potential free riders compare their net gains from cheating with their net gains from being good citizens, this decreases the temptation to defect and broadens the coalition of contributors to include more marginal states. At the same time, by relaxing the contributor's constraint, institutions allow the hegemon to select a higher level of contributions. This sets the stage for our evaluation of the decision to build institutions.

An institution is always beneficial, but it is costly to create. Indeed, if institutions are to explain anything, it is critical that they be costly to construct. As Stephen Krasner has pointed out, if institutions are too inexpensive, they become epiphenomenal; they emerge and fade away in response to changing interests and conjunctural factors, and never really explain anything independently of these shifting conditions (Krasner 1985). Furthermore, if institutions are cheap, they ought to be ubiquitous, in which case the supply of institutions cannot explain patterns of variation in outcomes. We argue that it is reasonable to regard institutions as costly. It may be true that states find the expenses of letterhead, office buildings, and salaries trivial, but they cannot regard as trivial the investment of high-level time and talent in negotiations. Bargaining is a very costly human activity, and as long as the design of institutions affects interests, it will be fraught with hard bargaining.¹⁷ Furthermore, the construction of institutions has long-term distributional consequences, so the would-be builders have to be prepared to make side payments, and these provide an additional incentive to stall the negotiations (Fearon 1998).

We now ask which hegemon will be willing to pay the cost of constructing institutions. Let $k > 0$ denote the cost of creating the institution. The hegemon will be willing to pay for an institution when the benefit of moving to the institutional setting outweighs the cost of having to create the institution. The following proposition establishes that only sufficiently large hegemon will be willing to create an institution.

Proposition 5: *For any given institution-building cost there exists a size threshold, such that only hegemon above that size will build the institution.*

¹⁶ Lohmann (1997) uses a similar approach to capture the dynamics of linkage in a repeated prisoners' dilemma.

¹⁷ Such bargaining does not occur in our model, because the introduction of the institution is Pareto optimal for the preexisting members of the coalition, the interests of the coalition members do not diverge regarding the optimal level of contribution, and we do not discuss alternative possible institutions. In a richer model where cooperation is possible along multiple dimensions and members are more heterogeneous, conflicts would arise over institutional design.

Our second finding about institutions, then, is that within the public-goods regime, it is larger hegemony that chooses to invest in institutions. This is because larger hegemony is more demanding and is more disadvantaged by their size; they find it impossible to sustain large coalitions of contributors. By building institutions, however, they reduce the incentives for small contributors to defect, and thereby expand their coalitions. This finding confirms an important strand of hegemonic stability theory, but does so by standing it on its head. It is indeed the case that hegemony chooses to build institutions; and in issue areas where discrimination is impractical, larger hegemony will be more likely to make the investment. However, it is not because hegemony benefits more from cooperation that they build institutions, nor because they can better afford to pay the costs, nor because the costs are lower as a consequence of the inordinate influence that they wield in the international system. None of these conventional arguments are present in our model. Instead, in a very spare model, we find that larger hegemony benefits more from building institutions because they are uniquely disadvantaged at promoting collective action. Large hegemony is institution builders because building institutions compensates for the disadvantages of hegemony.

Large hegemony is disadvantaged in our model because they have a credibility problem. From our perspective, the larger the hegemon, the more difficult it becomes to credibly threaten to refrain from producing public goods, because larger hegemony can produce them at lower unit cost. This formalizes the conjecture of Olson and Zeckhauser (1966, 274) that large allies would have credibility problems if they extended their model: “the large country loses more from withholding an alliance contribution than a small country does...it may be deterred by the very importance to itself of its own alliance contribution from carrying out any threat to end that contribution.” In our model, this logic leads other countries to free ride, although they would be willing to contribute if the hegemon were able to credibly commit to punishing free riding. The regime for international aid is a case in point. In the early years of the Cold War the United States provided aid virtually unilaterally, and it was only when its allies’ economies became large relative to its own that it became credible to condition U.S. contributions on theirs. In order to promote multilateral cooperation in providing development assistance, among other goals, the United States pushed to create the Organization for Economic Cooperation and Development (OECD) in 1960.¹⁸

Institutions and Discrimination

Does the opportunity to build institutions allow some large hegemony to switch from discrimination to the public-goods regime? In this section, we determine whether it is possible for a large hegemon to use institutions to overcome the

¹⁸ We do not intend to enter the debate about the effectiveness of foreign aid or the motivations for providing it here; it is sufficient for the illustration that U.S. allies perceived that there could be some positive externalities. Consider the U.S. position in 1947, when Britain announced that it was no longer going to provide economic aid to Greece. In announcing the Truman Doctrine, the president said: “There is no other country to which democratic Greece can turn. No other nation is willing and able to provide the necessary support for a democratic Greek government. The British Government, which has been helping Greece, can give no further financial or economic aid after March 31. I believe that it must be the policy of the United States to support free peoples who are resisting attempted subjugation by armed minorities or by outside pressures” (Truman 1947).

The situation was quite different by 1961, when the Organization for European Economic Cooperation (OEEC) was transformed into the OECD, which expanded its portfolio to include monitoring the delivery of bilateral development assistance. The OECD charter embodied a common commitment to contributing to development assistance: “Believing that the economically more advanced nations should co-operate in assisting to the best of their ability the countries in process of economic development the Members agree that they will contribute to the economic development of both Member and non-member countries in the process of economic development by appropriate means and, in particular, by the flow of capital to those countries, having regard to the importance to their economies of receiving technical assistance and of securing expanding export markets” (OECD 1960).

disadvantages of size, and switch to the public-goods regime. We have previously shown that larger hegemony prefer private goods to public goods (discriminatory to multilateral regimes) and that larger hegemony prefer public goods with institutions to public goods without institutions. We now ask whether there exist hegemony that are large enough to prefer public goods with institutions to public goods without institutions, but are still sufficiently small to prefer public goods with institutions to private goods.

In order to make this comparison, we revisit our previous finding that larger hegemony prefer the private-goods regime, but we slow down actions in the game. Formally, payoffs continue to accrue every period, but actions are taken only every other period. This has the consequence that free riding is more attractive, because defectors benefit from their defection for two periods before punishment begins. We find the condition for hegemony to prefer private goods to public goods under these new conditions ("slow time"), which corresponds to a situation where there are no institutions. We then allow hegemony to accelerate moves in the game by investing in an institution, and we find the condition for the hegemon to prefer the private-goods regime under the accelerated game, if it must make an investment in order to choose the public-goods regime.

By comparing the conditions with and without institutions, we are able to show that some hegemony are large enough to choose the private- over the public-goods regime, and yet small enough to be willing to switch to the public-goods regime if they have the option of constructing institutions.

Proposition 6: *For a range of costs of building institutions, there will be hegemony that would choose discrimination over multilateral cooperation if institution building were not an option, but choose to build institutions rather than discriminate.*

Our third finding about institutions, therefore, is that some hegemony that are so large that they prefer the discriminatory regime when institutions are not an available option are willing to build institutions and shift to the public-goods regime when they are. Institutions can make multilateralism attractive enough to justify the cost of their construction and forgoing discrimination. When discrimination is a viable alternative, however, the largest hegemony will exercise it rather than building institutions. Institutions will only be built when the hegemon is both large enough to need them and the cost of discrimination is high enough to make them attractive. Very large hegemony, or those facing low costs of discrimination, continue to prefer the discriminatory regime because they do not attract large coalitions of contributors even if they build institutions.

Conclusion

We present a general formal model of international cooperation. The model allows for the possibility that cooperation takes the form of providing a public good, but also allows the cooperating countries to choose, at some cost, to exclude non-contributors from enjoying the benefits of cooperation. It allows the capabilities of countries to vary, and explores the implications of varying the size of the leading state in the system. It allows countries to choose forms of cooperation that are more or less intense, and allows coalitions of various sizes to be constructed. It formalizes the notion, which has remained informal in the international relations literature until now, that international institutions can facilitate cooperation by reducing transaction costs. Finally, for the first time it determines under what conditions states will choose to create international institutions in a context that allows for infinite repetition, varying state capabilities, and both public goods and discrimination.

We are not surprised that the results of the model reject the traditional interpretation of the hegemonic stability thesis, which held that the presence of a dominant state in the system facilitates broad international cooperation. Under very general conditions we find that the stronger the leading state, the narrower the coalition of contributors that can be supported in equilibrium. This result is actually more consistent with the intuition of Olson's theory of collective action than the hegemonic theory that was derived from it: larger hegemons are more vulnerable to free riding.

Our most important results concern the qualitative impact of the distribution of power on cooperation: its breadth and depth, its discriminatory or nondiscriminatory nature, and its institutionalization. We find that there is a general trade-off between the depth and breadth of cooperation, which emerges because larger contributors are willing to make larger contributions. The marginal contributor differs from inframarginal contributors only in the effect of its contributions. The larger that effect, the sooner the discounted benefits of future cooperation balance the current cost of contributing. Consequently, as the size of the contribution demanded—and therefore the current cost—increases, the minimum size of the marginal contributor increases.

The more intense international cooperation, the narrower it must be. While other scholars have found limits to the optimal breadth or depth of cooperation in models with incomplete or imperfect information, our model establishes a more general result: all that is required to impose a trade-off between breadth and depth is that cooperation have the character of a public good and that different countries make different size contributions for a given level of effort. Further, casting cooperative strategies in terms of simultaneous decisions about breadth and depth reveals an important qualification to our finding that a skewed distribution of power leads to less participation in cooperative enterprises. We also find that larger hegemons are associated with more intensive cooperation within their smaller coalitions. This represents a new testable proposition, and we are not yet in a position to verify it.

We find that when the choice of multilateralism or discrimination is endogenous, larger hegemons choose discrimination. Again, this runs counter to the strand of international relations theory that has assumed that the presence of a dominant state is associated with openness in the international economy. This conclusion follows in a straightforward way from the conclusion that larger hegemons are disadvantaged in the public-goods regime because they are not able to credibly commit to punishing small states for defecting. They are not similarly disadvantaged if they discriminate, so discrimination allows them to compensate for the disadvantages of size.

We are able to confirm one of the important arguments about hegemonic stability. We find that when they engage in multilateral cooperation, larger hegemons are the ones who prefer to construct international institutions. This is not because larger hegemons can construct institutions at lower cost (they cannot in our model); instead, it follows from the fact that large hegemons lose more from free riding. Institutions allow large states to compensate for the disadvantages of their size, so they are a substitute for discrimination. Multilateral cooperation is harder to sustain than discrimination, and is particularly hard for large countries to sustain; however, institutions make multilateralism feasible.

The conclusion that large states are institution builders only holds when they choose the multilateral regime, however. In our model institutions play no role in the discriminatory regime, so countries that choose bilateralism—such as Britain in the nineteenth century—have no need to invest in building them. However, very large states may be institution builders if the cost of exclusion is prohibitively high, as might be the case in some issue areas. We ask whether an institutional strategy can be attractive enough to large hegemons that would

otherwise choose to create discriminatory regimes to motivate them to build institutions and foster multilateral cooperation instead, and we find that it can be, if the cost of discrimination is sufficiently high.

We conclude that for the last 20 years, international relations theorists have been trying to explain the wrong puzzle. The problem is not how to explain cooperation “after hegemony,” but rather how to explain cooperation “during hegemony”—for after a hegemon declines, cooperation becomes much easier to sustain and spreads to a wider circle of participants. The puzzle is how to explain the remarkable burst of institution building and multilateral cooperation inaugurated by the United States after the Second World War, when U.S. power surpassed anything that the modern international system had ever seen. The temptation of bilateral discrimination should have been overwhelming. The obvious answer from the perspective of our model is that the cost of discrimination was seen as prohibitive; given that discrimination was not a viable option, our large, disadvantaged hegemon had no choice but to construct institutions.

Appendix A. Proofs

The full proofs are available in the replication package.

Proof of Lemma 1: Take some $\mathcal{C} \in S$ and note that:

$$\frac{\partial \text{MC}}{\partial a} = \left(\frac{1 - \delta^{T+1}}{1 - \delta} \right) f'(a) > 0 \quad \text{and} \quad \frac{\partial \text{CR}}{\partial a} = -(1 - \delta^T) f'(a) < 0$$

where $f(a) = a^{c-b} > 0$ and $f'(a) = (c - b)a^{c-b-1} > 0$. Hence, the constraints either intersect once or not at all. The necessary and sufficient condition for an intersection is $\text{MC}(1, \mathcal{C}) \leq \text{CR}(1, \mathcal{C}) \Leftrightarrow (1 - \delta)/(1 - \delta^T) \leq \mathcal{C} + w_h - 1$, or $\text{X}(\mathcal{C}) = [(1 - \delta^T)(\mathcal{C} + w_h)]/(2 - \delta - \delta^T) \geq 1$. If $\text{X}(\mathcal{C})$ satisfies this, the intersection is at $f(\tilde{a}) = \text{X}(\mathcal{C})$, where $\tilde{a} = c^{-b} \sqrt[\mathcal{C}]{\text{X}(\mathcal{C})}$. Substituting into (MC) and (CR) yields $\text{MC}(\tilde{a}, \mathcal{C}) = \text{CR}(\tilde{a}, \mathcal{C}) = (1 - \delta)\text{X}(\mathcal{C})$, which is the value of $\tilde{w}(\mathcal{C})$ for the case of the intersection stated in the lemma. If $\underline{w}(\mathcal{C}) \geq (1 - \delta)\text{X}(\mathcal{C})$, then there exist values of a such that (a, \mathcal{C}) is admissible. Otherwise, there is no solution and this coalition does not permit any admissible profiles. If the constraints do not intersect, then $\text{MC}(a, \mathcal{C}) > \text{CR}(a, \mathcal{C})$ for all values of a . Therefore, it suffices to check whether $\text{MC}(1, \mathcal{C})$ admits a solution. That is, if $\underline{w}(\mathcal{C}) \geq \text{MC}(1, \mathcal{C})$, then there exist values of a such that (a, \mathcal{C}) is admissible. \square

Proof of Proposition 1: Given that there exist admissible profiles, the hegemon picks the one that maximizes its payoff. Doing so is optimal for the hegemon because any alternative admissible profile will yield a worse payoff and any inadmissible profile fails at least one of the constraints, which implies that cooperation will unravel and the hegemon’s payoff will be zero. Since the profile the hegemon chooses is admissible, both (MC) and (CR) are satisfied, and all members prefer to contribute while non-members free-ride on their efforts. Since the constraints are derived from the strategies specified in this section, these strategies form a stationary SPE of the game. In fact, these strategies will support any admissible profile in a stationary SPE of the continuation game after the initial choice by the hegemon, so the hegemon is effectively choosing which stationary SPE to play. \square

Proof of Proposition 2: Define the optimal choices as follows:

- (i) if $c/b < K_1$, then $\underline{w}^* = \text{MC}_1(a_1^*) > \text{CR}_1(a_1^*)$, where $a_1^* = c^{-b} \sqrt[\mathcal{C}]{\frac{b(\mathcal{W} + w_h)}{c(1 + \epsilon)}}$,

- (ii) if $c/b > K_2$, then $\underline{w}^* = CR_1(a_2^*) > MC_1(a_2^*)$, where $a_2^* = c^{-b} \sqrt[b]{\frac{b(\mathcal{W}+w_h)}{c}}$,
- (iii) if $c/b \in [K_1, K_2]$, then $\underline{w}^* = MC_1(a_3^*) = CR_1(a_3^*)$, where $a_3^* = c^{-b} \sqrt[b]{\frac{\mathcal{W}+w_h}{K_2}}$,

where $K_2 = 1+(1-\delta)(1+\epsilon(1-\delta^T))/(1-\delta^T) > K_2/(1+\epsilon) = K_1 > 1$. The hegemon’s optimization problem is:

$$\max_{a, \underline{w}} \pi_h(a, \underline{w}) \text{ subject to } \underline{w} \geq \max\{MC_1(a), CR_1(a)\} \ \& \ \underline{w} \geq 0 \ \& \ a \geq 1.$$

The Lagrangian is:

$$\mathcal{L} = \pi_h(a, \underline{w}) - \lambda_1(MC_1(a) - \underline{w}) - \lambda_2(CR_1(a) - \underline{w}) - \lambda_3(1 - a) - \lambda_4(-\underline{w}),$$

with the Kuhn–Tucker conditions:

$$\frac{\partial \mathcal{L}}{\partial a} = \frac{\partial \pi_h(a, \underline{w})}{\partial a} - \lambda_1 \frac{\partial MC_1(a)}{\partial a} - \lambda_2 \frac{\partial CR_1(a)}{\partial a} + \lambda_3 = 0; \tag{1}$$

$$\frac{\partial \mathcal{L}}{\partial \underline{w}} = -\epsilon a^b + \lambda_1 + \lambda_2 + \lambda_4 = 0; \tag{2}$$

$$\lambda_1 \geq 0, \quad MC_1(a) \leq \underline{w} \quad \& \quad \lambda_1(MC_1(a) - \underline{w}) = 0; \tag{3}$$

$$\lambda_2 \geq 0, \quad CR_1(a) \leq \underline{w} \quad \& \quad \lambda_2(CR_1(a) - \underline{w}) = 0; \tag{4}$$

$$\lambda_3 \geq 0, \quad -a \leq -1 \quad \& \quad \lambda_3(1 - a) = 0;$$

$$\lambda_4 \geq 0, \quad -\underline{w} \leq 0 \quad \& \quad \lambda_4(-\underline{w}) = 0. \tag{5}$$

Since we are looking for an interior solution, we set $\lambda_3 = \lambda_4 = 0$. Observe now that at least one of the constraints must be binding at a solution. To see that this must be the case, suppose that it is not. Since both constraints are slack, $MC_1(a) - \underline{w} > 0$ and $CR_1(a) - \underline{w} > 0$ at the solution, and so equations (3) and (4) imply that $\lambda_1 = \lambda_2 = 0$. But then equation (2) requires that $\epsilon a^b + \lambda_4 = 0$. Since $\epsilon a^b > 0$, this inequality cannot be satisfied for any $a \geq 1$ because $\lambda_4 \geq 0$ must hold by equation (5). Therefore, at least one of the constraints must be binding at a solution.

Suppose first that $CR_1(a)$ is slack, and so $\lambda_2 = 0$. In this case, equation (2) implies that $\lambda_1 = \epsilon a^b > 0$, which means that equation (3) requires that $\underline{w} = MC_1(a)$, whose unique solution is a_1^* . We now need to ensure that $CR_1(a_1^*) - \underline{w}_1 < 0$. This inequality reduces to: $c/b < K_1$, which can be satisfied because $K_1 > 1 \Leftrightarrow \epsilon < \bar{\epsilon}$, which holds by Assumption 1. It is easy to show that $c/b < K_1 \Rightarrow \underline{w}_1 > 0$, as required. Therefore, (a_1^*, \underline{w}_1) is a valid solution if, and only if, $b/c < K_1$, which yields the first case in the proposition.

Suppose now that $MC_1(a)$ is slack, and so $\lambda_1 = 0$. In this case, equation (2) implies that $\lambda_2 = \epsilon a^b > 0$, which means that equation (4) requires that $\underline{w} = CR_1(a)$, whose unique solution is a_2^* . Since $\underline{w}_2 = CR_1(a_2^*) > 0$ is always satisfied, we only require $MC_1(a_2^*) - \underline{w}_2 < 0 \Leftrightarrow c/b > K_2$ to ensure that this is a solution. Hence, (a_2^*, \underline{w}_2) is a valid solution if, and only if, $c/b > K_2$. This yields the second case in the proposition.

Finally, suppose that both constraints are binding. Solving $MC_1(a) = CR_1(a)$ yields a_3^* , and so $\underline{w}_3 = MC_1(a_3^*) = CR_1(a_3^*)$. The derivation ensures that $MC_1(a_3^*) - \underline{w}_3 = 0$ and $CR_1(a_3^*) - \underline{w}_3 = 0$, as required. To satisfy equation (2), we need $\lambda_1 > 0$, $\lambda_2 > 0$, and $\lambda_1 + \lambda_2 = \epsilon a_3^b > 0$, which can certainly be satisfied provided $c/b \in [K_1, K_2]$. This yields the third case in the proposition.

To establish the claim about the comparative statics of the optimal demand, we need to show that $\partial a^*/\partial w_h > 0$, which is clearly true from inspection. To see that $\partial \underline{w}^*/\partial w_h > 0$ as well, set \underline{w}^* as specified above and note that the restrictions on c/b imply that the derivative is strictly positive. \square

Proof of Proposition 3: First we show that if the exclusion cost is sufficiently low, the discriminatory regime is always preferred to the public-goods regime. Let the subscripts D and P denote the discriminatory and public-goods regimes, respectively. Since a_D^* is exactly the unconstrained optimum when $\underline{w} = 0$, it follows that the hegemon's payoff under the discriminatory regime is precisely the maximum unconstrained payoff net the discrimination cost: $\pi_h(a_D^*, 0) - m$. But $\pi_h(a_D^*, 0) > \pi_h(a^*, \underline{w}^*)$ implies that there exists $\bar{m} > 0$ such that $\pi_h(a_D^*, 0) - \bar{m} = \pi_h(a^*, \underline{w}^*)$. Hence, for any $m < \bar{m}$, the hegemon will strictly prefer the discriminatory regime to the public-goods regime.

We now show that for any exclusion cost there exists a size threshold above which hegemony prefers to discriminate. The payoff from the public-goods regime is: $\pi_h(a, \underline{w}) = a^b[\mathcal{W} + w_h - \epsilon \underline{w} - a^{c-b}]$, and the payoff from the discriminatory regime is:

$$\pi_h(a_D^*, 0) - m = (\mathcal{W} + w_h)^{\frac{c}{c-b}} \left(\frac{c-b}{c}\right) \left(\frac{b}{c}\right)^{\frac{b}{c-b}} - m.$$

The hegemon will prefer to pay the exclusion cost and discriminate rather than produce public goods whenever $\pi_h(a_D^*, 0) - \pi_h(a^*, \underline{w}^*) > m$, or whenever $\Pi \cdot (\mathcal{W} + w_h)^{\frac{c}{c-b}} > m$, where

$$\Pi = \left(\frac{c-b}{c}\right) \left(\frac{b}{c}\right)^{\frac{b}{c-b}} - (1-\delta)(P-\epsilon) \left(\frac{1}{K_2}\right)^{\frac{c}{c-b}} > 0, \tag{6}$$

and $P = [1 + \epsilon(1 - \delta^T)](1 - \delta^T)^{-1}$. Define $\underline{w}_h(m) = (m/\Pi)^{\frac{c-b}{c}} - \mathcal{W}$, and note that for any $m > 0$, every hegemon with $\underline{w}_h > \underline{w}_h(m)$ will strictly prefer to discriminate. \square

Proof of Proposition 4: Let the subscript S denote slow-time (without an institution). We first derive the optimal profile without institution where players interact every other period only using the usual trigger strategies. For the contributor to be willing to contribute it must be the case that:

$$\frac{\pi_i(a, \mathcal{C})}{1-\delta} \geq (1+\delta)[(\mathcal{C} + w_h)a^b - w_i a^b] + \left(\frac{\delta^{T+2}}{1-\delta}\right) \pi_i(a, \mathcal{C}),$$

or $\underline{w} \geq [(1-\delta^{T+2})a^{c-b} - \delta^2(1-\delta^T)(\mathcal{W} + w_h)]/[1 - \delta^2 - \epsilon\delta^2(1-\delta^T)] \equiv MC_S(a)$. The credibility constraint is the same as before, $CR_1(a)$, because the hegemon decides whether to punish by looking at present and future payoffs, which have not changed. Since both constraints are binding, we obtain $\underline{w}_S^* = CR_1(a_S^*) = MC_S(a_S^*)$, where

$$a_2^* = \sqrt[c-b]{\frac{\mathcal{W} + w_h}{K_2}} \quad \text{where} \quad \hat{K}_2 = 1 + \frac{(1-\delta^2)[1 + \epsilon(1-\delta^T)]}{1-\delta^T}.$$

Let (a_3^*, \underline{w}^*) be the optimal (fast-time, with institution present) profile from Proposition 2 when both constraints are binding. Note now that: $a_3^* > a_2^* \Leftrightarrow \hat{K}_2 > K_2 \Leftrightarrow 1 > \delta$, so the optimal demand with the institution is strictly larger than without it. Since $\underline{w}^* < \underline{w}_S^* \Leftrightarrow \hat{K}_2 > K_2$, the size of the smallest contributor is strictly smaller with the institution than without it. \square

Proof of Proposition 5: We need to compare the expected payoffs with an institution, $\pi_h(a^*, \underline{w}^*)$ from Proposition 2 with both constraints binding, and without an institution, $\pi_h(a_S^*, \underline{w}_S^*)$ from the proof of Proposition 4. Letting $k > 0$ represent the cost of building the institution, the payoff from doing so exceeds the payoff of having no institution whenever $\pi_h(a^*, \underline{w}^*) - k > \pi_h(a_S^*, \underline{w}_S^*)$. Since $\pi_h(a_S^*, \underline{w}_S^*) = (\mathcal{W} + w_h)^{\frac{c-b}{c}}(1 - \delta^2)(P - \epsilon)(1/\hat{K}_2)^{\frac{c-b}{c}}$, we can rewrite the condition as: $\hat{\Pi} \cdot (\mathcal{W} + w_h)^{\frac{c-b}{c}} > k$, with

$$\hat{\Pi} = \left[(1 - \delta) \left(\frac{1}{\hat{K}_2} \right)^{\frac{c-b}{c}} - (1 - \delta^2) \left(\frac{1}{\hat{K}_2} \right)^{\frac{c-b}{c}} \right] (P - \epsilon) > 0, \tag{7}$$

where the inequality follows $\hat{K}_2 > K_2$ implying that the bracketed term is positive. Define $\hat{w}_h(k) = (k/\hat{\Pi})^{\frac{c-b}{c}} - \mathcal{W}$, and note that for any $k > 0$, every hegemon with $w_h > \hat{w}_h(k)$ strictly prefers to build an institution. \square

Proof of Proposition 6: The hegemon will prefer the private-goods regime to “slow time” whenever $\pi_h(a_D^*, 0) - m > \pi_h(a_S^*, \underline{w}_S^*)$, which reduces to $\Pi_S \cdot (\mathcal{W} + w_h)^{\frac{c-b}{c}} > m$, where

$$\Pi_S = \left(\frac{c - b}{c} \right)^{\frac{b}{c-b}} - (1 - \delta^2)(P - \epsilon) \left(\frac{1}{\hat{K}_2} \right)^{\frac{c-b}{c}}.$$

Hence, any hegemon with size $w_h > (m/\Pi_S)^{\frac{c-b}{c}} - \mathcal{W} \equiv \underline{h}$ will strictly prefer paying the exclusion cost to continuing without an institution. The hegemon will prefer to pay the exclusion cost whenever $\pi_h(a_D^*, 0) - m > \pi_h(a^*, \underline{w}^*) - k$, or: $\Pi \cdot (\mathcal{W} + w_h)^{\frac{c-b}{c}} > m - k$, with Π as defined in equation (6). Hence, any hegemon with size $w_h > [(m - k)/\Pi]^{\frac{c-b}{c}} - \mathcal{W} \equiv \bar{h}$ will strictly prefer paying the exclusion cost to building an institution. Note now that: $\underline{h} < \bar{h} \Leftrightarrow k < m(\Pi_S - \Pi)/\Pi$, where $\Pi_S - \Pi = \hat{\Pi} > 0$ and $\hat{\Pi}$ is defined in equation (7). In this case, all $w_h \in [\underline{h}, \bar{h}]$ prefer building an institution to discriminating, and prefer discriminating to living without an institution. \square

References

ACHESON, DEAN. [1969] 1987. *Present at the Creation: My Years at the State Department*. New York: W. W. Norton & Company.

AXELROD, ROBERT. 1984. *The Evolution of Cooperation*. New York: Basic Books.

BUENO DE MESQUITA, BRUCE, AND DAVID LALMAN. 1992. *War and Reason*. New Haven: Yale University Press.

COASE, RONALD. 1960. “The Problem of Social Cost.” *Journal of Law and Economics* 3: 1–44.

CONYBEARE, JOHN. 1983. “Tariff Protection in Developed and Developing Countries: A Cross-sectional and Longitudinal Analysis.” *International Organization* 37(3): 441–467.

CONYBEARE, JOHN. 1984. “Public Goods, Prisoners’ Dilemmas and the International Political Economy.” *International Studies Quarterly* 28(1): 5–22.

COWHEY, PETER, AND EDWARD LONG. 1983. “Testing Theories of Regime Change: Hegemonic Decline or Surplus Capacity?” *International Organization* 37(2): 157–188.

DOWNS, GEORGE W., AND DAVID M. ROCKE. 1995. *Optimal Imperfection? Domestic Uncertainty and Institutions in International Relations*. Princeton: Princeton University Press.

DOWNS, GEORGE W., DAVID M. ROCKE, AND PETER N. BARSOOM. 1998. “Managing the Evolution of Multilateralism.” *International Organization* 52(2): 397–419.

DREZNER, DANIEL W. 2000. “Bargaining, Enforcement, and Sanctions: When Is Cooperation Counterproductive?” *International Organization* 54(1): 73–102.

FEARON, JAMES D. 1998. “Bargaining, Enforcement, and International Cooperation.” *International Organization* 52(2): 269–305.

- GILLIGAN, MICHAEL J. 2004. "Is There a Broader-Deeper Trade-off in International Multilateral Agreements?" *International Organization* 58(Summer): 459–484.
- GILPIN, ROBERT. 1972. "The Politics of Transnational Economic Relations." In *Transnational Relations and World Politics*, ed. Robert O. Keohane and Joseph S. Nye. Cambridge: Harvard University Press.
- GILPIN, ROBERT. 1981. *War and Change in World Politics*. Princeton: Princeton University Press.
- GOWA, JOANNE. 1989. "Rational Hegemons, Excludable Goods, and Small Groups: An Epitaph for Hegemonic Stability Theory?" *World Politics* 41(3): 307–324.
- KAHLER, MILES. 1992. "Multilateralism with Small and Large Numbers." *International Organization* 46(3): 681–708.
- KENNAN, GEORGE F. 1951. *American Diplomacy, 1900–1950*. Chicago: University of Chicago Press.
- KENNAN, GEORGE F. 1969. *Memoirs, 1925–1950*. New York: Bantam Books.
- KEOHANE, ROBERT O. 1984. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton: Princeton University Press.
- KEOHANE, ROBERT O., AND JOSEPH S. NYE. 1977. *Power and Interdependence: World Politics in Transition*. Boston: Little Brown & Co.
- KINDLEBERGER, CHARLES P. 1973. *The World in Depression, 1929–39*. Berkeley: University of California Press.
- KRASNER, STEPHEN. 1976. "State Power and the Structure of International Trade." *World Politics* 28(3): 317–347.
- KRASNER, STEPHEN D. 1985. *Structural Conflict: The Third World Against Global Liberalism*. Berkeley: University of California Press.
- KRASNER, STEPHEN D., ED. 1983. *International Regimes*. Ithaca: Cornell University Press.
- KREPS, DAVID M. 1990. *Game Theory and Economic Modelling*. Oxford: Clarendon Press.
- LOHMANN, SUSANNE. 1997. "Linkage Politics." *Journal of Conflict Resolution* 41(1): 38–67.
- MANSFIELD, EDWARD D. 1994. *Power, Trade, and War*. Princeton: Princeton University Press.
- MARTIN, LISA M. 1992. *Coercive Cooperation*. Princeton: Princeton University Press.
- MCKEOWN, TIMOTHY. 1983. "Hegemonic Stability Theory and Nineteenth Century Tariff Levels in Europe." *International Organization* 37(1): 73–91.
- MCKEOWN, TIMOTHY. 1991. "A Liberal Trade Order? The Long-run Pattern of Imports to the Advanced Capitalist States." *International Studies Quarterly* 35(2): 151–171.
- MCLEAN, ELENA V., AND RANDALL W. STONE. 2005. "Two-Level Bargaining and the Kyoto Protocol." Presented at the 101st Annual Meeting of the American Political Science Association, Washington, D.C., September 1–4, 2005.
- MILGROM, PAUL, AND JOHN ROBERTS. 1992. *Economics, Organization, and Management*. Englewood Cliffs, NJ: Prentice Hall.
- OECD. 1960. "Convention on the Organization for Economic Co-operation and Development." Paris: OECD December 14.
- OLSON, MANCUR, AND RICHARD ZECKHAUSER. 1966. "An Economic Theory of Alliances." *The Review of Economics and Statistics* 48(3): 266–279.
- OYE, KENNETH A. 1992. *Economic Discrimination and Political Exchange: World Political Economy in the 1930s and 1980s*. Princeton: Princeton University Press.
- OYE, KENNETH A., ED. 1986. *Cooperation Under Anarchy*. Princeton: Princeton University Press.
- PAHRE, ROBERT. 1998. *Leading Questions: How Hegemony Affects the International Political Economy*. Ann Arbor: The University of Michigan Press.
- PALMER, GLENN. 1990. "Corralling the Free Rider: Deterrence and the Western Alliance." *International Studies Quarterly* 34(2): 147–164.
- REINHARDT, ERIC R. 2001. "Adjudication without Enforcement in GATT Disputes." *Journal of Conflict Resolution* 45(2): 174–195.
- RUBINSTEIN, ARIEL. 1985. "Choice of Conjectures in a Bargaining Game with Incomplete Information." In *Game-Theoretic Models of Bargaining*, ed. Alvin E. Roth. New York: Cambridge University Press.
- RUGGIE, JOHN GERARD. 1993. *Multilateralism Matters: The Theory and Praxis of an Institutional Form*. New York: Columbia University Press.
- SIMMONS, BETH A. 1994. *Who Adjusts? Domestic Sources of Foreign Economic Policy During the Interwar Years*. Princeton: Princeton University Press.
- SNIDAL, DUNCAN. 1985. "The Limits of Hegemonic Stability Theory." *International Organization* 39(4): 579–614.
- TRUMAN, HARRY S. 1947. "Address Before a Joint Session of Congress." March 12.
- WILLIAMSON, OLIVER. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.
- YOUNG, ORAN R. 1994. *International Governance: Protecting the Environment in a Stateless Society*. Ithaca: Cornell University Press.