DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

Alliances in Anarchic International Systems

Emerson M. S. Niou
Duke University

Peter C. Ordeshook
California Institute of Technology

SOCIAL SCIENCE WORKING PAPER 761

May 1991

Alliances in Anarchic International Systems *

Emerson M. S. Niou Peter C. Ordeshook

Abstract

Alliances play a central role in international relations theory. However, aside from applications of traditional cooperative game theory that ignore the issue of enforcement in anarchic systems, or interpretations of the repeated Prisoners' Dilemma in the attempt to understand the source of cooperation in such systems, we have little theory on which to base predictions about alliance formation. This essay, then, builds on an n-country, non-cooperative, game-theoretic model of conflict in anarchic systems in order to furnish a theoretical basis for such predictions. Defining an alliance as a collection of countries that jointly abide by "collective security strategies" with respect to each other but not with respect to members outside of the alliance, we establish the necessary and sufficient conditions for an alliance system to be stable. In addition, we show that not all winning or minimal winning coalitions can form alliances, that alliances among smaller states can be stable, that bipolar alliance structures do not exhaust the set of stable structures, and that only specific countries can play the role of balancer.

• This research was supported by the NSF Grant SES-8822308 to Duke University and the NSF Grant SES-8922262 to the California Institute of Technology.

Alliances in Anarchic International Systems

There is little disagreement over the proposition that the concept of ***alliance*** is central to international relations theory. In the realist view, "the historically most important manifestation of the balance of power ... is to be found ... in the relations between one nation or alliance and another alliance" (Morgenthau 1959:169) because "alliances and regional coalitions among the weak to defend themselves from the strong have been the typical method for preserving ... balance" (Wright 1965:773). And although neoliberals offers an alternative formula for stability, alliances in the form of regimes play an important role there as well to the extent that they facilitate the realization of mutually beneficial economic gains.

Insofar as our understanding of alliances is concerned, we are aided by the fact that definitions come within striking distance of acceptability by even rigorous theoretical standards. Consider Walt's (1987:12) definition: "an alliance is a formal or informal arrangement for security cooperation between two or more sovereign states;" or Snyder's (1990:104): "alliances ... are formal associations of states for the use (or non-use) of military force, intended for either the security or the aggrandizement of their members, against specific other states..." Although we might expand the domain of agreements to include economic cooperation, an alliance here matches the game-theorist's idea of a coalition in which people coordinate strategies to realize some outcome that cannot be realized through uncoordinated action.

However, definitions do not tell us how alliances are formed and maintained in anarchic systems nor do they provide predictions about their formation. To that end, we turn to a game-theoretic model that specifies the circumstances under which such systems are and are not stable. Of course, ours is not the first attempt to apply game theory to such matters. Beginning with Riker's (1962) "size principle," continuing through the analysis of public goods (Olson and Zeckhauser 1966, Oppenheimer 1979), and extending to the application of cooperative solution theory (Niou, et al 1989), many hypotheses have been offered about alliance formation. But these analyses do not directly confront the fact that alliances arise in environments in which there are no exogenous mechanisms for enforcing agreements. Applications of cooperative game theory and derivative hypotheses such as the size principle assume away the issue of enforceability and analyses that appeal to some feature of public goods focus on a different issue -- burden sharing -- than the one we wish to address here. And even that research that explicitly explores the sources of cooperation in anarchic environments relies on a particular game, the repeated Prisoners' dilemma, which cannot generally characterize international affairs whenever those affairs become purely conflictual (c.f., Taylor 1976, 1987; Bendor and Mookherjee 1987).

The question we try to answer here then is: If alliances arise and are sustained purely on the basis of individual self-interest in an otherwise anarchic world in which there is competition for scarce resources, what types of alliance structures are stable and what types are unstable? We begin in Section 1 by reviewing a model of anarchic systems developed previously (Niou and Ordeshook 1990, 1991) to formalize the notions of balance of power and collective security and to establish the possibility of systems that are stable in the sense that all countries can ensure their sovereignty against all threats. Here, however, we focus on the alliances that might form in anticipation of the necessity for conducting international politics in an otherwise anarchic world. In Section 2 we define alliances by equating them with limited collective security agreements, and in Section 3 we define stable alliance systems. In Section 4 we provide a necessary and sufficient condition for such systems to exist, and we offer some examples and subsidiary results that allow us to interpret those conditions. We also examine the following questions: Are profitable alliances restricted to winning or minimal winning coalitions? Can the largest (most militarily powerful) states form an alliance at the expense of smaller states, or will alliance structures necessarily divide the most powerful states into opposing camps -- is there any inherent tendency towards bipolarity? Must a collective security equilibrium encompass all states or can a subset of states enforce a system devoid of threats against sovereignty. Can alliances be purely defensive? Are offensive alliances more attractive than defensive ones? Finally, in Section 5 **we** examine the role in our model of balancing powers and in Section 6 we provide some concluding remarks. Appendix A offers some results under an alternative assumption about the actions of indifferent countries, and Appendix B contains the proofs of all results.

1. **Balance of Power versus Collective Security**

Much of the realist-neoliberal debate can be interpreted as an argument over whether a balance of power or a collective security equilibrium is more stable or is a more appropriate characterization of contemporary affairs (Niou and Ordeshook 1991). However, the resolution of this debate requires two things -- a model of anarchic systems and formulations of balance of power and collective security in terms of the strategies that countries employ.

A Model: Recognizing the heroic simplification implied by such an assumption, we begin with the supposition that, conditional on maintaining their sovereignty, countries pursue a single transferable resource in constant supply. Next, we suppose that the amount of that resource controlled by each country is the sole determinant of winning and losing coalitions, and that countries join coalitions because it is in their individual interest to do so. Finally, in accordance with Boulding's (1968:105) view that "threat systems are the basis of politics," we assume that threats and counter-threats are the mechanisms whereby countries secure resources from each other.

To formalize matters, we let $\mathbf{r}^o = (r^o_1, r^o_2, \ldots, r^o_n)$ be the initial distribution of resources across a set $S = \{1,2,\ldots,n\}$ of n countries, where $r^o_1 \geq r^o_2 \geq \ldots \geq r^o_n$. Next, we let $r(C)$ denote the total resources controlled by the subset of countries C, and $R = r(S)$ be the total resources in the system. Hence, C is a winning coalition if $r(C) > R/2$, it is losing if $r(C) < R/2$, and a winning coalition is minimal winning if the deletion of any member from it renders it losing. Countries who are in at least one minimal winning coalition are *essential*; otherwise they are *inessential*. If $r^o_i > R/2$, country i is *predominant* -- it is winning against all other countries and it can absorb their resources at will. Hence, every country has an incentive to avoid the predominance of any other country. If $r^o_i = R/2$, then i is *near-predominant*. Avoiding mathematical niceties, the game we use to model anarchic systems can then be described as follows:

1. A randomly chosen country, i, is given the opportunity to offer an initial threat or to "pass." An initial threat is a new resource distribution $\mathbf{r}$ and an implied threatening coalition C that corresponds to the countries who do not lose resources by moving from $\mathbf{r}^o$ to $\mathbf{r}$. Of course, $r(C) > r(S-C)$.
2. If i passes, we return to step 1.
3. If i threatens, its partners in C decide whether to accept participation in the threat. Only if all partners accept does i's threat call for a response by the threatened countries. If one or more members of C reject, we return to 1.
4. Responses by threatened countries are of two types. First, each threatened country, taken in some random sequence, can offer a counter-threat, which is a new threat. If this counter is accepted unanimously by the newly proposed coalition, it cancels the original threat by becoming the new current threat, and requires a response by the newly threatened countries.

5. A second type of response is a proposal by one or more threatened countries to surrender resources to one or more members of the threatening coalition. If a transfer is accepted by everyone involved, it determines a new status quo, and the game proceeds as before by returning to step 1.

6. Any threat that is not successfully countered is implemented -- the threatened resource distribution becomes the new status quo, and the game proceeds as before by returning to step 1.

This model does not consider the costs of conflict, uncertainty, and exogenous resource growth; and it characterizes anarchy in a highly structured way. Also, by supposing that countries maximize a resource that is in constant supply, it may be biased in favor of the realist argument. However, whatever cooperation emerges in it does not emerge merely because we have made cooperation sufficiently profitable via some assumption about the value of public goods. Also, it assumes that countries join and maintain coalitions because it is in their individual interest to do so and not because exogenously imposed constraints. And it matches Boulding's view of threat systems as the fundamental characteristic of anarchic politics.

Strategies: The essential difference between a system governed by a balance of power and one governed by collective security concerns the strategies that countries employ. Briefly, a strategy is a plan of action that specifies three things. First, it tells the country what to do whenever there is no current threat and when nature gives it the opportunity to threaten or to pass. Second, for each threat in which it is included in the threatening coalition, a strategy tells it whether to accept or reject that threat. Finally, for each threat against the country, a strategy tells that country what counter to use and whether or not to accept a counter that is offered.

There are a great many alternative strategies for a game as complex as ours. However, we distinguish between balance of power and collective security systems by focusing on two particular types of strategies.

> **Definition:** In a "balance of power system," all countries use ***stationary*** strategies — strategies in which each country makes the same choices whenever it encounters the same threat, so that countries ignore who made a threat or who agreed to participate in it when fashioning their response.

Hence, in a balance of power system, "all states are potentially fit alliance partners; none is seen as much more evil than any other" (Jervis 1986:60). Alternatively,

Definition: In a "collective security system," all countries use ***punishment strategies,*** where punishments are directed against those who try to upset the status quo by making a threat or by agreeing to participate in one.

Strategies in a collective security system, then, are not stationary because they posit specific threats, depending on who defects from some prior agreement.

Equilibria: Throughout our analysis we employ the concept of an equilibrium most appropriate to our model -- subgame perfect equilibria. Thus, cooperation is enforced by individual self-interest, where that self-interest is defined by the strategies of other states. Briefly, now, if we find equilibria supported by stationary strategies and if we learn also that only countries controlling some critical relative resource level can ensure their sovereignty, then countries must be vigilant about relative gains and losses. On the other hand, if there is an equilibrium supported by punishment strategies in which no country offers an initial threat, then realization of this equilibrium renders the issue of sovereignty and relative position less salient and allows for greater flexibility in the design of cooperative arrangements. Moreover, if the benefits that accrue through free trade and the like require a non-conflictual world, and if these benefits disappear when agreements to achieve them are disrupted by competition over relative position, then the issue which bears directly on the realist-neoliberal debate is whether such an equilibrium is more or less attractive than the one supported by stationary strategies.

Skirting formalism, we need only review our two central conclusions.[1] First, if we further characterize stationary strategies by the statement "countries participate in threats if doing so does not lead to a reduction in their resources," then:[2]

Result 1: ***If all countries are essential, and if they all abide by stationary strategies, then there exists a strong equilibrium in which no country is eliminated. But if we allow sequential threats (i.e., i proposes that C threatens j, then k, etc.), then inessential countries and perhaps even "small" essential ones cannot assure their sovereignty*** (Niou and Ordeshook, 1990).

Thus, there is an equilibrium in which the sovereignty of "larger" states, but not of smaller ones, is assured, in which case countries must be vigilant about their relative share of resources. With respect to this equilibrium's attractiveness, in addition to being self-enforcing, it is strong in this sense: The "largest" countries, if allowed to make a threat, do so because they gain and thereby avoid the possibility of loss, whereas smaller countries, although unable to gain by participating in a threat, avoid the possibility of losses by doing so. So if a country believes that all others will abide by their equilibrium strategies, then it has a positive incentive to make or to agree to threats that include it in the threatening coalition, since not doing so diminishes its utility.

To model collective security, we let ***D*** denote the countries that are the potential targets of punishments, and we consider a punishment strategy that matches the simplicity of stationary strategies, because we do not want to confront the objection that one type of equilibrium is easier to compute and realize than another. Hence, we restrict our attention to the following characterization: (a) No country proposes an initial threat; (b) No country accepts an initial threat if one is offered; (c) Threats are directed against one or more defectors; (d) Countries accept threats that are punishments; (e) Whenever any threat that is not part of a punishment is accepted, all countries use stationary strategies thereafter. Assuming that players defecting from (a)-(d) are added to ***D*** and are thereafter subject to punishment, we have the following result:

> **Result 2: If** ***there are four or more essential countries, the strategy described in (a)-(e) yields a strong equilibrium such that no country makes an initial threat and the status quo is preserved; but if there are only three such countries or if countries must sequentially reject their participation in threats, then the collective security equilibrium is not strong*** (Niou and Ordeshook 1991).

Thus, punishment strategies support equilibria in which no one makes a threat, and no one is eliminated. But (and here we ought to keep in mind the bias in our model in favor of the realist conceptualization), such equilibria are vulnerable in that they are weak if there are only three essential countries or if countries must sequentially reveal their willingness to participate in threats. The particular problem occasioned by that weakness is that countries can "wander away" from the equilibrium. And because it is difficult to judge adherence to a punishment strategy (since commitment is revealed only after the fact), countries can not be certain beforehand that they are in fact in such an equilibrium. If all countries presume that all others have some chance of defecting from administering punishments -- if countries "play with a shaky hand" -- then the collective security equilibrium can break down. Thus, collective security requires "nurturing" by mechanisms that facilitate the realization of those mutual benefits that disappear when countries compete for relative position rather than pursue the pure objective of absolute resource maximization.

2. Threats and Alliances

The preceding analysis considers only two extremes -- either all states agree not to threaten anyone and to punish defectors or all states stand ready to threaten anyone. However, although we equate one idea with collective security and the other with balance of power, balance of power systems are in fact commonly associated with an intermediate possibility in which subsets of states establish alliances whereby members of an alliance agree not to threaten each other but to threaten or defend against those outside of the alliance. Thus, to fully understand the sources of stability and instability, we must consider the concept of an alliance.

In defining an alliance within the framework of our model, we can secure some guidance from recent history. Specifically, consider that throughout the Cold War the Warsaw Pact was held together not only by a set of economic relationships and by fear of invasion from the West, but also by military force. Defectors from the alliance -- Hungary, Czechoslovakia, and Romania -- were, until recently, punished by the alliance's remaining members. And although nothing as dramatic as tanks rolling through Paris, Tokyo, or Bonn cemented the Western alliance, the source of its durability was, in addition to the threat from Moscow, its economic profitability. The collective security arrangement among the United States, Japan, and Western Europe required the administration of no severe military punishment, but there has always existed the threat of economic reprisal and a withdrawal of military support in the event of any defection. Similarly, if we take liberties with the notion of a regime, then we see that alliances are also collective security arrangements held together by the promise of gain and the threat of punishment: "A hegemon may help to create shared interests by providing rewards for cooperation and punishments for defection, but where no hegemon exists, similar rewards and punishments can be provided if conditions are favorable" (Keohane 1984:78). Thus, regardless of whether we deem alliances as primarily defensive or offensive, the following definition seems appropriate:

> **Definition:** An ***alliance*** is a collective security arrangement among states in which all members of the alliance agree to not threaten each other, to punish defectors from this agreement whenever possible, and to threaten countries outside of the alliance whenever it is in their individual interest to do so.

Whether certain types of alliances act to preserve the status quo whereas others seek to upset it by making threats is an issue we address later. However, to ascertain whether this definition provides any leverage over our understanding of events, we must return to the method whereby we establish a balance of power and a collective security equilibrium. Briefly, the feature of our model that warrants emphasis is that it allows for the possibility that threats and counters, as well as resource reallocations, continue in sequence forever. Thus, Wagner (1986:551) outlines our game's correct treatment: "the basic question that concerns us is whether states will act so as to eliminate other states. If one state is eliminated from a four-actor game, for example, the result is to precipitate a three-actor game. If a value can be assigned to such a subgame for each player, it is possible to determine whether any players have an incentive to eliminate other players." We proceed, then, by pretending that the game is finite and that we know the consequences of all branches in its extensive form. After postulating these consequences, an equilibrium is characterized by strategies in which no one has an incentive to defect unilaterally to any choice not dictated by that strategy, and the postulated consequences are consistent in that they are "self-fulfilling prophesies" -- the choices they imply yield those consequences.

We associate consequences with actions by forming a 2-way classification of coalitions, where members of one class can make threats that yield one type of consequence (the largest country in the threatening coalition becomes near-predominant and no one in the coalition loses resources) and members of the other class can only make threats that can be countered by a coalition of the first type such that no threatening country is assured of gaining resources and some lose. Specifically,

Definition: The set of winning coalitions C^* is *advantaged* if

- i for every $C \in C^*$, the members of $S-C$ have sufficient resources to render the largest member of C near-predominant;
- ii for every $C \in C^*$ there is no other winning coalition C' such that the intersection of C' and C is the largest country in C' but not in C.
- iii. no two coalitions in C^* have a unique common largest member;[3]
- iv C^* is maximal in the sense that no additional coalitions satisfying conditions i and ii can be added without violating iii.

We then define a *primary threat* as a resource vector **r** proposed by a member C of C^* such that no member of C loses resources in **r** ($r_i \geq r_i^{\circ}$ for all $i \in C$), the members of

S-C are threatened with elimination ($r_i = 0$ for all $i \in S$-C), and the threat promises to make the largest member of C near-predominant ($r_i = R/2$ for $i = \max[C]$).

The critical components of our definition are ii and iii. Condition ii ensures that there cannot be a counter-threat that makes the same promise to a smaller country in C without simultaneously requiring the inclusion of some other members of C. Because these other members of C cannot simultaneously receive the same promise (it is always cheaper for a threatened coalition to transfer to resources to a single country if a transfer is their only way to disrupt a threat), these members will reject the counter in favor of the current threat. Conditions iii, in combination with ii, ensures that if a primary threat is made and accepted, there is no other threat that can serve as a counter to it and some or all of the threatened countries must cede resources to the largest threatening country. On the other hand, if a non-primary threat is made and accepted, then each (essential) threatened country can counter with a primary threat that is accepted. Hence,

> **Remark 1**: *Whether or not countries abide by punishment or stationary strategies, if everyone is essential, no one makes a non-primary threat. And if all countries abide by stationary strategies and if a country proposes an initial threat, then it is a primary threat, and the final outcome has the largest threatening country becoming near-predominant at the expense of the initially threatened countries.*

To illustrate C^* for specific initial resource distributions, consider the following 3, 4, and 5-country examples.

> **Example**: If $\mathbf{r}^\circ = (120,100,80)$, then $C^* = \{\{1,3\}, \{2,3\}\}$ and primary threats take the form (150,0,150) and (0,150,150). The winning coalition $\{1,2\}$ is not advantaged, because, in violation if condition ii, for $C' = \{2,3\}$, $C' \cap C = \{2\} = \max[C'] \neq C$. The set C^* is unique for any 3-country game in which every country has less than one half of the system's resources. Moreover, the final outcome has country 1 becoming near-predominant at the expense of country 2, or vise versa, with country 3 neither gaining nor losing resources.

> **Example:** If $\mathbf{r}^\circ = (100,95,75,30)$, country 4 is inessential, and C^* is the same as before. If countries 1 or 2 make the initial threat, then either 1 and 4 transfer to 2 or 2 and 4 transfer to 1; but if 3 makes the initial threat, then a threat such as (110,105,85,0) eliminates 4.

Example: If $\mathbf{r}^\circ$ = (110,80,60,50), everyone is essential and C^* = {{1,2,3}, {1,2,4}, {1,3,4}, {2,3,4}}.[4] Notice that {1,2} is not a member of C^* since, in violation of condition ii, {2,3,4} is winning; nor is {1,3} a member of C^* since, in violation of condition iii, {1,2,4} is winning and $r(\{2,3,4\}) > r(\{1,3\})$. Like the 3-country case, only countries 1 and 2 can gain resources; but unlike the 3-country case, countries 3 and 4 can lose resources if they are the target of an initial primary threat.

Example: If $\mathbf{r}^\circ$ = (70,65,60,55,50), then C^* = {{1,4,5}, {1,3,4}, {1,3,5}, {2,3,4}, {2,3,5}, {2,4,5}, {3,4,5}}. Moreover, since country 3 can be the largest member of a minimal winning coalition, it, in addition to 1 and 2, can gain resources if it participates in an appropriate initial threat.

Example: If $\mathbf{r}^\circ$ = (95,85,60,40,20), then C^* = {{1,2,3}, {1,2,4,5}, {1,3,4}, {1,3,5}, {1,3,4,5}, {2,3,4}, {2,3,5}, {2,3,4,5}}. C^* does not include {1,4,5} because $r(\{1,2,3\}) > r(\{1,4,5\})$; nor is {1,3} included because $r(\{1,2,4,5\}) > r(\{1,3\})$.

An important lesson of these examples is that C^* does not exhaust the set of minimal winning coalitions. For example, in the 3-country case the minimal winning coalition C = {1,2} is not included in C^*. However, to characterize advantaged coalitions further, let the set L_o denote the smallest countries that can never be the largest member of a minimal winning coalition and let $L = S - L_o$.[5] In our 4- and in our first 5-country examples L_o = {3,4}; in our last example, L_o = {3,4,5}. The following remark summarizes what we know about advantaged coalitions (parts 1 and 2 follow directly from definitions, whereas parts 3-5 are contained in the proofs we offer elsewhere of our previously stated results about equilibria):

Remark 2: *(1) Not all winning or minimal winning coalitions are advantaged; (2) Every C in C^* must contain at least one member of L and one member of L_o; (3) A coalition consisting of any one country in L and all of L_o is advantaged; (4) Every essential country is a member of some C in C^*; (5) If there are 3 or more essential countries, then only coalitions with 3 or more members are advantaged.*

To this list we can add one more fact. Notice that in our 3-country example, country 3 cannot be the target of a primary threat. On the other hand, every country is the target of such a threat in our other examples. Lemma 1 generalizes this fact:

Lemma 1: *If there are four or more essential countries, then every country is excluded from at least one coalition in* C^*.

3. Stable Alliance Structures: A Definition

Thus far we have merely identified the coalitions that assure a beneficial resource transfer, but we cannot determine what agreements can be reached in anticipation of playing our threat-counter-threat game. For example, is an all-encompassing collective security arrangement required forestall the possibility of threats; and are only alliances that correspond to advantaged coalitions profitable? Suppose, then, that prior to playing our threat-counter-threat game, countries bargain to partition themselves into exhaustive and disjoint alliances. We appreciate, of course, that such bargaining lies at the heart of the most interesting processes in international affairs. Indeed, we conjecture later that the period 1871-1914 consisted of just such an "out of equilibrium" process and, therefore, that "balance of power politics" does not correspond to the attainment of an equilibrium but rather to the process whereby a particular equilibrium is achieved.

However, rather than model bargaining itself, we proceed using a classical cooperative game theoretic approach -- by identifying the potential "sticking points" of bargains. In accordance with our view that international systems are anarchic, though, we continue to assume that countries participate in alliances because it is in their individual interest to do so, and not because of any exogenous enforcement. Thus, a partition is *a stable alliance structure* if no set of countries can gain by coordinating their actions so as to reform the partition into a different alliance structure, either by defecting from their current alliances or by joining two or more alliances into a single alliance.

To remove any ambiguity from this definition, let P be a partition of S. Next, define s_i^P as country i's *security value with respect to P*. Briefly,

> **Definition:** country i's security value with respect to the partition P, s_i^P, corresponds to i's minimum payoff if everyone subsequently abides by subgame perfect strategies, with the assumption that alliance partners play punishment strategies with respect to each other unless, in playing our threat-counter-threat game, they prefer to defect unilaterally from doing so.

This definition of security value is incomplete because we have some flexibility in the specification of certain actions when countries are indifferent. But before we examine alternatives, let us consider a situation that entails no ambiguity. If no alliances form in our 3-country example -- if $P = (\{1\},\{2\},\{3\})$ -- then $s_1^P = s_2^P = 70$, because neither 1 nor 2 can preclude the possibility that one or the other will join with 3. But $s_3^P = 80$, because there is no primary threat against 3 and it can always offer a primary threat as a counter to any threat against it. Indeed, since we already know that 3-country systems are unique in that country 3 cannot be threatened by a primary threat, its security value is always 80; and since it can never gain resources, it never has an incentive to admit or to expel a partner from an alliance. Countries 1 and 2, on the other hand, have a considerable incentive to ally with 3.

To generalize our discussion to larger systems, we say that the partition P is stable if there is no alternative partition P' such that the security value of all members of C' in P' is greater than their security value in P. Formally,

> **Definition:** The partition P is stable if there does not exist a P' such that for any C' in P', $s_i^{P'} > s_i^P$ for all i in C'.

A stable alliance system, then, looks like an element of the core of a cooperative game -- an outcome in which no coalition has the ability and a unanimous incentive to upset. An alliance partition is stable if no collection of countries has a unanimous positive incentive to establish a different partition. But unlike the usual applications of the core that presuppose the exogenous enforcement of agreements, alliances here are enforced by the mutual self-interest that arises from subsequently playing the threat-counter-threat game we use to model anarchic systems.

4. **Stable Alliance Systems: Existence**

To characterize stable and unstable alliance systems, we offer the following general theorem: Letting $G(P) = \{ i \mid s_i^P < r^\circ_i\}$ denote the set of countries with security values less than their current resource distribution, then,

> **Theorem 1**: *The alliance structure P is stable if and only if* $S \not\subseteq G(P)$.

Theorem 1 by itself does not determine the stability of particular structures, because security values depend on the actions we ascribe to countries that are indifferent between abiding by and defection from an equilibrium strategy. For example, if $\mathbf{r}^\circ = (70,65,60,55,50)$, if $P = (\{1,3,4\},\{2,5\})$, and if 1, 3, and 4 abide by their equilibrium strategy, then, because $\{1,3,4\} \in C^*$, $G(P) = \{2,5\}$ -- that is, countries 2 and 5 can be the targets of a primary threat by $\{1,3,4\}$. However, $\{2,4,5\}$ and $\{2,3,5\}$ also are in C^*, and 3 and 4 are both indifferent as to which primary threat they participate in. So if, in the case of indifference, countries have some probability of unilaterally defecting from an alliance's collective security arrangement, then everyone's security value in P is less then their current endowments. On the other hand, the partition ({3,4,5},{1,2}) is immune from such instability because no individual defection from {3,4,5} yields an alternative coalition with a primary threat.

Taking account of such uncertainty is important. In matters of survival, national leaders are unlikely to be comforted by technical arguments about equilibria and should plan for worst-case scenarios, which includes the defection of indifferent allies. Assuming, then, that indifferent countries "play with a shaky hand," Theorem 2 provides our first specific result about stable alliances (Appendix A describes stable alliances if indifferent countries abide by equilibrium strategies with certainty):

> **Theorem 2**: *If "countries play with a shaky hand," and if there are more than three essential countries, then $P = (C,S-C)$ is stable if and only if:*
>
> *(1) $S-C$ is empty; or*
>
> *(2) C has a primary threat and for no $i \in C$ does $S-C + \{i\}$ have a primary threat; or*
>
> *(3) C hasn't a primary threat and for no $i \in S-C$ does $C + \{i\}$ have a primary threat.*

Before illustrating this result, it is useful to distinguish between profitable and unprofitable alliances, because such a distinction removes some of the ambiguity between bipolar and multipolar systems and reveals that Theorem 2 applies to certain types of "multi-polar" systems. If $\mathbf{r} = (70,65,60,55,50)$, then $P = (\{3,4,5\},\{1\},\{2\})$ and $P' = (\{3,4,5\},\{1,2\})$, in addition to being stable, are equivalent in the sense that it matters little whether 1 and 2 ally to transform P into the bipolar system P'. Allied or not, both countries are certain to be targets subsequently of a primary threat by {3,4,5}. Hence, to sort through these equivalences, we offer the following definition:

> **Definition**: Suppose $A_k = \{k,i,...,m\}$ and $A_k \in P$. If we substitute $\{k\},\{i\},...,\{m\}$ for A_k in P to form P', then P and P' are *equivalent* if $s_j^P = s_j^{P'} < r_j^o$ for all $j \in A_k$, and if $s_j^P = s_j^{P'}$ for all other $j \in S$. In this event A_k is *unprofitable*.

Thus, $P = (C,A_1,...A_k)$ is equivalent to $(C,S-C)$ if $S-C = A_1 \cup ... \cup A_k$ is unprofitable.

Letting Table 1 summarize our discussion, we now illustrate Theorems 1 and 2 (Table 1 also summarizes the results in Appendix A).

Example: Let $\mathbf{r}^\circ = (120,80,60,40)$ and $P = (\{1,2,3\},\{4\})$. Since $\{1,2,3\} \in C^*$, $s_4^P < r_4^\circ$. The only potential defectors from $\{1,2,3\}$ are 2 and 3, because neither gains by threatening 4, but neither has an incentive to defect unilaterally since both can be punished. Hence, P is stable. Now let $P = (\{1,2\},\{3,4\})$, in which case 1 and 2 must each be concerned that its partner will defect since both $\{1,3,4\}$ and $\{2,3,4\}$ are advantaged. Thus, $s_j^P < r_j^\circ$, $j = 1,2$. And although neither 3 nor 4 has a positive incentive to defect, each is indifferent between maintaining the alliance and defecting, so if both play with a shaky hand, $s_j^P < r_j^\circ$, j = 3,4. Hence, $S = G(P)$, and from Theorem 1, $(\{1,2\},\{3,4\})$ is unstable. For similar reasons, the partitions $(\{1,3\},\{2,4\})$ and $(\{1,4\},\{2,3\})$ are unstable.

Example: If $\mathbf{r}^\circ = (70,65,60,55,50)$, then $L = \{1,2,3\}$. Now consider $P = (\{3,4,5\},\{1\},\{2\})$. Notice that $\{3,4,5\} \in C^*$, but no unilateral defection yields a coalition in C^*. Thus, any defection can be punished and P is stable. In contrast, let $P = (\{1,4,5\},\{2,3\})$. Although $\{1,4,5\} \in C^*$, 4 and 5 can each unilaterally defect to $\{2,3\}$ to form an advantaged coalition, which is a possibility we cannot preclude if countries play with a shaky hand whenever indifferent. Indeed, this argument applies to all bipolar systems in which one country has a primary threat except $(\{3,4,5\},\{1\},\{2\})$. Now let $P = (\{1,2,3,4\},\{5\})$. If $\{1,2,3,4\}$ proposes to eliminate 5, the result is a 4-country game in which no country is immune from being the target of a primary threat. But if 1 moves first in our threat-counter threat game and proposes that $\{1,4,5\}$ threaten $\{2,3\}$, 5 accepts, but 4 is indifferent between accepting and rejecting and so there is some probability that 4 rejects and 1 and 5 are targets of a subsequent threat. Thus, no threats are made and no one has an incentive to defect from their alliance, so P is stable. This situation reveals that not all states need agree to a collective security arrangement for that arrangement to yield the collective security outcome.

Example: If $\mathbf{r}^\circ = (100,80,60,40,20)$, then $L = \{1,2\}$, and 4-country coalitions against 1, 2 or 3 are advantaged as is $\{1,2,3\}$. And, since no unilateral defection from any of these coalitions generates a coalition with a competing alliance, the four alliance structures portrayed in Table 1 are stable. On the other hand, $(\{1,3,4\},\{2,5\})$ is not stable even though $\{1,3,4\} \in C^*$, because if 3 defects, $\{2,3,5\}$ has a primary threat.

Example: Let $\mathbf{r}^{\circ} = (100,80,20,20,20,20,20,20)$. Notice first that any $(C,S-C)$ such that $C \in C^{*}$ is stable, because no unilateral defection can generate a counter-coalition in C^{*}. This example differs from our previous ones, however, in that previously, at least one alliance in the stable alliance structure is a winning coalition. In contrast, $(\{1\},\{2\},\{3,4,5,6,7,8\}=L_o)$ is stable here, because no unilateral defection from L_o yields a coalition in C^{*}. Because they are never the recipients of a transfer, no $i \in L_o$ has no preference for a different alliance system. For the same reason $(\{1\},\{2\},\{3\},\{4,5,6,7,8\})$ is also stable. But if we delete too many from L_o, as in $P = (\{1\},\{2\},\{3\},\{4\},\{5,6,7,8\})$, the alliance structure is unstable. First, the defection of anyone from $\{5,6,7,8\}$ renders $\{1,2,3,4,j\}$ advantaged, so $s_j^P < r_j^{\circ}$, $j = 5,6,7,8$. Second, $\{1,5,6,7,8\} \in C^{*}$, so $s_j^P < r_j^{\circ}$, $j = 2,3,4$. Third, given their security values, countries 2,3,...,8 have an incentive to threaten 1 (that such a threat exists follows from Lemma 1), so $s_1^P < r_1^{\circ}$. The instability of P follows from Theorem 1. Finally, $(\{1,2,...,7\},\{8\})$ and $(\{1,2,...,6\},\{7,8\})$ are stable for the same reason that $(\{1,2,3,4\},\{5\})$ is stable in our first 5-country example and that each of these alliance systems yields the same outcome as an all-encompassing collective security arrangement -- the absence of any threat.

Thus, using $\mathbf{r}^{\circ} = (75,65,60,55,50)$ to illustrate matters, the following conclusions summarize our results:

1. not all winning or advantaged coalitions can establish a stable alliance structure -- $(\{1,2,3\},\{4\},\{5\})$ is not stable;
2. the consequence of an all encompassing collective security system is achieved by a collective security agreement that encompasses "nearly all" states -- $(\{1,2,3,4\},\{5\})$ is stable and no country offers a threat in the subsequent play of our threat-counter threat game;
3. stable alliance systems need not be bipolar -- $(\{3,4,5\},\{1\},\{2\})$ is stable;

And, using $\mathbf{r}^{\circ} = (100,80,20,...,20)$ as our example,

4. a stable alliance system need not contain any winning alliances -- $(\{1\},\{2\},\{3,...,8\})$ is stable.

To see these conclusions differently, notice that three types of alliances are the centerpiece of a stable alliance system:

A. an advantaged alliance such as {3,4,5} in ({1,2},{3,4,5}), which is thereby offensive;

B. a "large" alliance such as ({1,2,3,...,7},{8}) which supports a collective security outcome; and

C. an alliance like {3,4,...,8} in ({1},{2},{3,4,...,8}), which is not winning and which is both defensive (because it can block any primary threat) and offensive (because it can join with others to form such a threat).

Although our model cannot specify precisely the events following the formation of one type of alliance structure as against another, we can speculate. Briefly, type A alliances -- advantaged coalitions that can make primary threats -- are offensive. They are not formed to preserve any balance and they are likely to be short lived, because they are designed to upset the status quo in favor of a reallocation of resources. In contrast, type B alliance structures model the abortive League of Nations and more recently, perhaps, Bush's "New World Order." Historical evidence suggests that such alliances are short-lived as well, and indeed, we already know from Remark 2 that they require a special form of "nurturing" if they are to compete against alliances of the first type. Finally, type C alliances are perhaps most congruent with classical balance of power notions in that they play the role of balancer and can either prevent profitable threats or determine which threat is eventually made and accepted. We cannot say, however, whether such alliances have any advantage over other types in terms of durability.

5. Balancers

Alliances that can block the formation of a primary threat are necessarily pivotal between coalitions that can make such a threat. Although ({1},{4},{5},{2,3}) is stable if $r^o = (100,80,60,40,20)$ and if countries play with a sure hand (see Appendix A), {2,3} can join with either 4 or 5 to form a profitable threat; and the alliance {3,4,5} plays a similar role in the structure ({1},{2},{3,4,5}). Thus, both {2,3} and {3,4,5} appear to satisfy the following definition of a balancer:

> **Definition:** a country or an alliance is a *balancer* in a given alliance structure if it is pivotal between any two coalitions of alliances with primary threats.

The examples of Britain in the 19th century and China in the 20th, however, raise several questions. Owing ostensibly to its geographical isolation and its desire to preserve the status quo so as to maintain profitable trading relationships, Britain is credited with playing the role of balancer in the 19th century, and thus we can ask: Does our analysis predict Britain's role and does it rationalize the relevance of geography. China in this century sought a similar role for itself, although, unlike Britain, it sought to form an alliance of third world states in order to offset the American and Soviet-led blocks. Thus we can ask: Why must China, as opposed to the US or USSR, play this role, and is there any reason to suppose that China would find a balancing alliance of third world states especially valuable?

Looking at the issue of whether any country can be a balancer, given an appropriate partition, two remarks help us answer such a query. First,

> **Remark 3**: *The two largest countries can never individually be balancers in any alliance structure.*

This remark appears to be contradicted by Britain's 19th century role. By most measures, Britain's military capability exceed that of any continental power despite the fact that Germany was closing fast at the end of the century. However, it is here that Britain's geographical position with respect to the continent becomes relevant. Although her navy ensured far greater force projection than Germany with respect to Africa, India, the Far East, and even, perhaps, the Balkans, the events of World War I confirm that Britain was severely handicapped in any military engagement close to Germany. Britain's potential on the continent as compared to Russia or France is less clear, but if we assume that one or the other exceeded Britain's continental capability, then Britain is no longer precluded from playing the role historians and diplomats assigned to her. Indeed, Remark 3 supplies a formal basis for rationalizing geography as an important determinant of that role.

Whether Britain was uniquely positioned to play this role depends in part on whether the issue of Alsace-Lorraine precluded any effective Franco-German alliance and whether language and culture rendered Germany and Austria "natural" allies. If we simplify matters by ignoring Italy, if we assume that military capability on the continent was ordered G(ermany) + Austria > R(ussia) > B(ritain) > F(rance), and if we assume, for purposes of a specific numerical representation, that $r^{\circ} = (120,80,60,40)$, then there are two offensive alliance structures: ({R,B,F},{G}) and ({G,R,B},{F}), plus one "blocking" structure, ({G},{F},{R,B}). In all three instances, then, Britain and Russia pivot together. Although Russia sought the role of balancer, it attempted to play this role using one alliance that was not winning -- {R,F} -- and one that was winning but not advantaged -- {G,R}. In addition, its efforts were hampered by its ineptitude and perceived perfidity -- certainly, Britain's leaders played their role with considerably greater skill.

However, more problematical from the perspective of asserting that Britain's role was preordained, is this result:

> **Remark 4**: *If the number of essential countries equals three, then the smallest state is a balancer; but if the number of such countries exceeds three, then no individual country can be a balancer in any stable alliance structure.*

Reassessing Britain's role, recall that the period 1870 - 1914 was marked by considerable alliance instability. The short life of the League of Three Emperors, the formation of the Triple Alliance, Germany's courting of Britain, Russia's vacillation between alliance with Germany and alliance with France, British overtures to Austria, and Italy's uncertain role contrast sharply with 40+ years of stability exhibited by NATO and the Warsaw Pact. Thus, although Britain may have been assisted in its role of balancer by geography, playing that role unilaterally also required a fluidity of alignments that disappeared when Germany threatened continental predominance.

In contrast, Remark 4 suggests that to the extent that it perceived a 3-state system dominated by America and the USSR, China alone could seek to play the role of balancer. To the extent, though, that other states become relevant and potentially independent actors, China cannot play that role alone and to extract resources from a stable international system, intermediate states must forge an alliance with smaller states. Admittedly, though, the balancing role we have outlined for the states in L_o is not altogether supported by historical evidence. As Fox (1959:185) observes "attempts to add to the power of the small states by combining with other small and presumably disinterested small states regularly failed ..." On the other hand, Fox also reveals that a balancing role was not precluded as a possibility: "none of the small states ... dared go so far in using the strength of one side to oppose another ...[but] the possibility of such a move was frequently in the minds of the great-power leaders."

6. Conclusions

The preceding discussion does not explain why countries seek the role of balancer, which points to one of the limitations of our analysis. For Britain, as an industrialized trading state with significant overseas investments whose benefits came from maintaining the status quo of loosely formed alliances, the implementation of a threat jeopardized economic gains. Thus, Britain may have preferred a system in which no stable, threatening alliance formed. In contrast, China, as a consumer of technology and investment, sought to extract the benefits that arose when competing alliances felt compelled to bid for her "services." However, neither the gains sought by Britain nor the transfers sought by China are part of our model.

Our model also fails to consider the process whereby alliances form and dissolve, and focuses instead on the outcomes that end bargaining. But if we accept Claude's (1962:145) view that in balance of power politics the system "divides [countries] into antagonistic groups, jockeying for position against each other," then such politics concerns process rather than final outcomes, and we cannot be certain that we have not bypassed the part of international affairs that occupies the greatest share of the diplomat's time. Indeed, our evaluation of Britain's role in the 19th century rests on the presumption that this period did not correspond to an equilibrium outcome, but rather to an period in which one country (Britain) deliberately sought to forestall any ultimate resolution of bargaining. The suggestion here, then, is that the next step in modeling anarchic systems is a model that pays special attention to the bargaining process whereby alliances are negotiated.

There are other matters that we have ignored. First, we do not allow countries to invest resources so that relative resource shares (power) change over time. Second, aside from the assumption that countries prefer to have resources ceded to them over securing them by implementing threats, we do not fully accommodated the costs of war. A final matter to be confronted is that of uncertainty. Although we allow indifferent countries to choose probabilistically, this is not the only way in which uncertainty can effect our analysis. A second and perhaps more important way is bypassed by our assumption that every state knows the point at which a state becomes predominant. Hence, there is no risk to allowing a state to become near-predominant, which is not an assumption that we can comfortably assert characterizes reality.

Despite these limitations we can provide conclusions that are more powerful than those offered by previous research. First, once the issue of exogenous enforcement is confronted, hypotheses such as the size principle must be modified — indeed, stable alliance structures need not even contain winning alliances. Second, the realization of a universal collective security arrangement does not require the acquiescence of all states. Something other than a coalition-of-the-whole can enforce such an equilibrium if states fear the uncertainty that prevails after excluded states are "eliminated." Only systems in which states suffer the hubris of believing that they can enforce something that was unenforceable before they took aggressive action requires an all-encompassing agreement to ensure a threat-free international system. Third, although alliances can be both offensive and balancing, balancing alliances are most easily formed by collections of smaller states that cannot aspire to near-predominance. Thus, our analysis is consistent with the hypothesis that there is a fundamental divergence in the foreign policy objectives of "small" versus "large" states. In summary, there is a great variety of stable alliance systems. In particular, we can conjecture that there is no necessity for choosing between the "state of nature of pure balance of power" and the seemingly unrealizable **Utopia** of an all encompassing collective security system.

Our final conclusion concerns the tendency of alliance systems to move towards bipolarity. Our approach has been to identify systems that have core-like stability in the sense that there are not countries with a positive incentive to reform a preexisting system prior to the play of our threat-counter-threat game. Thus, in looking at the process whereby alliances "build up" from some initial state of nature, we cannot say whether we are more likely to move towards competing alliances, blocking alliances, or "large" alliances that ensure the collective security outcome. On the other hand, notice that even a stable partition such as ({1 },{2},{3},{4,5}) can eventually result in bipolarity. Countries 4 and 5 are indifferent between blocking any profitable threat and eventually forming such a threat with 1, 2, or 3, and in fact making either choice is an equilibrium choice. Thus, whether or not a system eventually results in bipolarity (in the form of a primary threat by C against ***S-C)*** in the play of our threat-counter threat game depends on factors that we do not fully consider such as the costs of conflict. If there are mechanisms whereby 4 and 5 are assured of sharing in the spoils of "victory," and if conflict is not too costly, then bipolarity results; but if there are no such mechanisms or if conflict destroys any potential gains from implementing threats, then the eventual outcome is indeterminate. Any tendency towards bipolarity, then, must originate from considerations in addition to the desire of nations to survive in anarchic environments.

Table 1: Stable Alliance Structures with profitable alliances

r°	with "shaky hand"	without "shaky hand"
(120,100,80)	({1,3},{2}) ({2,3},{1}) ({1},{2},{3})	same as with shaky hand + ({1,2,3})
(120,80,60,40)	({1,2,3,4}) ({1,2,3},{4}) ({1,2,4},{3}) ({1,3,4},{2}) ({2,3,4},{1})	same as with shaky hand + ({1},{2},{3,4}) ({1},{3},{2,4}) ({1},{4},{2,3})
(70,65,60,55,50)	({1,2,3,4,5}) ({3,4,5},{1},{2}) ({1,2,3,4},{5}) ({1,2,3,5},{4}) ({1,2,4,5},{3}) ({1,3,4,5},{2}) ({2,3,4,5},{1})	same as with shaky hand + ({1,4,5},{2},{3}) ({2,4,5},{1},{3}) ({1,3,4},{2},{5}) ({1,3,5},{2},{4}) ({2,3,5},{1},{4}) ({1},{2},{3},{4,5})
(100,80,60,40,20)	({1,2,3,4,5}) ({1,3,4,5},{2}) ({2,3,4,5},{1}) ({1,2,4,5},{3}) ({1,2,3},{4},{5})	same as with shaky hand + ({1,3,4},{2},{5}) ({1,3,5},{2},{4}) ({2,3,4},{1},{5}) ({2,3,5},{1},{4}) ({1},{2},{3,4,5}) ({1},{4},{5},{2,3})

Appendix A

If indifferent countries "play with a sure hand" and always make equilibrium choices, then the set of stable alliance structures necessarily expands and ascertaining the extent of this expansion allows us to evaluate the effect of the form of uncertainty that our analysis admits. Our central result in this circumstance is this:

Theorem 3: *If countries defect from alliances only if they gain from doing so, then any bipolar alliance system is stable.*

This result appears to admit too much, but recall the three types of stable systems identified previously. For example, if $\mathbf{r}^{\circ}$ = (70,65,60,55,50), then P = ({1,2,4},{3,5}), P' = ({1,2,3,4},{5}), and P'' = ({1,2,3},{4,5}) are stable. In P the alliance {1,2,4} has a primary threat, P' is equivalent to an all encompassing collective security arrangement, and P'' establishes {3,4} as a potential blocking alliance. Thus, if we again ignore unprofitable alliances, P illustrates Remark 5, P' illustrates Remark 6, and P'' illustrates Remark 7:

Remark 5: *If countries defect from alliances only if they gain from doing so, then any alliance system in which one alliance has a primary threat is stable.*

Remark 6: *If countries defect from alliances only if they gain from doing so, and if alliances must be renegotiated after any reallocation of resources, then the bipolar alliance system (C,S-C) is stable and is equivalent to an all-encompassing collective security arrangement if* $r(S-C) + r_{\max[C]} < R/2$.

Remark 7: *As long the members of S - A, r(S-A) > r(A) and A ∈ P, require some member of A to form a primary threat, then P is stable.*

These remarks do not imply that the only interesting non-bipolar stable alliances structures are those that entail coalitions of countries in L_o, or that only the smallest countries can form defensive alliances. Our next theorem allows us to establish circumstances under which other types of alliance systems are stable, including systems in which members of L join a defensive alliance.

Theorem 4: *If countries defect from alliances only if they gain from doing so, then P is stable if there is a* $C \in P$ *such that there does not exist a* $K \subseteq S-C$ *in which* $K \in C^*$ *and for no* $i \in C$ *is* $\{i\} + K \in C^*$ *with* $i = \max[K,i]$.

Example: If $\mathbf{r}^{\circ}$ = (70,65,60,50,30,7,6,6,6), then L_o = {4,5,6,7,8,9}, and from Remark 6, ({1},{2},{3},{4,5,6,7,8,9}) is stable. But let P = ({1,6,7},{2,8,9},{3,4,5}).

Because no combination of 1, 2, 6, 7, 8, and 9 is advantaged, {3,4,5} cannot be threatened and it has no incentive to accept another country into it. Hence, {3,4,5} is a purely defensive alliance, and P is stable.

Example: If $\mathbf{r}^{\circ}$ = (100,80,60,40,20), then L_{o} = {3,4,5}, and, from Remark 6, ({1},{2},{3,4,5}) is stable. But if P = ({1},{4},{2,3},{5}), then {2,3} cannot be threatened by a primary threat -- indeed, as long as {2,3} maintains itself, only {2,3,5} and {2,3,4} are advantaged. However, {2,3} is not purely defensive since it can join other countries to form an advantaged coalition.

Thus, if countries play with a sure hand, we can observe blocking alliances that include members of L -- {2,3} when $\mathbf{r}^{\circ}$ = (100,80,60,40,20) and {3,4,5} when $\mathbf{r}^{\circ}$ = (70,65,60,50,30,7,6,6,6). But the stability of such alliances rests on a precarious assumption, and although "great powers" (those in L) might try to construct an alliance that blocks threats, they are likely to be thwarted by any uncertainty of commitment.

Appendix B: Proofs

Proof of Lemma 1: The lemma is clearly true $\forall\, k \in L$ since $|L| \geq 2$ and since, for any $k \in L$, $\exists\, C \in C^*$ that excludes k, namely $C = L_o + \{j\}$, $j \neq k$, $j \in L$. For any $j \in L_o$, if $r_j + r_1 \geq R/2$, then clearly S-$\{i\} \in C^*$. So suppose $r_j + r_1 < R/2$. Then construct the coalition C' by adding members of S-$\{j\}$ to j, beginning with the largest (country 1), then the next largest, and so on to m+1, until $r(C') + r_m \geq R/2$. Then S-$C' \in C^*$ -- by construction, condition i is satisfied; condition ii cannot be violated since $r_1 + r_j < R/2$ implies that $r_m + r < R/2$; and condition iii cannot be violated since $1 \notin S$-C'.

We proceed now to establish three additional lemmas.

> **Lemma 2:** If $C \notin C^*$ because condition iii is violated, then both S-C + $\{\max[C]\}$ and S-C + K are advantaged, where $K \subseteq C$ - $\{\max[C]\}$.

Proof: That S-C+$\{\max[C]\} \in C^*$ follows from the assumption that condition iii is violated. Next, notice that $r(C$-$\{\max[C]\}) + r(S$-$C) > R/2$, otherwise $\max[C] > R/2$. So add members from C-$\{\max[C]\}$ to S-C, beginning with the smallest members of C-$\{\max[C]\}$, until the resulting coalition is winning. This coalition is advantaged -- neither condition ii nor iii can be violated.

> **Lemma 3:** If $C \notin C^*$ because condition ii is violated, then $\exists\, j \in C$, $j \neq \max[C]$, such that S-C+$\{j\} \in C^*$, with $j = \max[S$-$C + \{j\}]$.

Proof: If condition ii is violated, then $\exists\, j \neq \max[C]$ in C such that S-$C + \{j\} \in W$. Let j be the smallest country in C for which this is true. S-$C + \{j\} \notin C^*$ either because it fails to satisfy condition ii or iii. It cannot violate condition ii, though, since $\nexists\, C'$ $\in W$ that excludes S-C, has j as its largest member, and is advantaged. Nor can condition iii be violated; otherwise, C-$\{\max[C],j\} + \{h\} \in W$, where $h \neq \max[S$-$C]$. However, S-$C + \{j\} \in W$ by construction, so C-$\{j\}$ and C - $\{\max[C],j\} \notin W$. And since $r^{\circ}_{\max[S-C]} < r_j^{\circ}$ by construction, C - $\{\max[C],j\} + \{h\} \notin W$.

> **Lemma 4**: If all $i \in S$ are essential and $r_{\max[C]} + r(S$-$C) < R/2$, then $|C| \geq 4$.

Proof: The lemma is clearly true if $|S| = 4$, otherwise country 4 is inessential. By the same token, if $|S| \geq 5$ and if $r_{\max[C]} + r(S$-$C) < R/2$, then $|C| \geq 4$; otherwise, members of S-C are inessential.

Proof of Theorem 1: To prove sufficiency, notice first that if $G(P) = \emptyset$, then no $i \in L_o$ gains by defecting to some other alliance or by admitting someone to their alliance. Since threats in which members of L gain require the participation of members of L_o, P is stable. Second, suppose that $G(P) \neq \emptyset$ and that $S \neq G(P)$. Thus, $j \in G(P)$ can improve its security value if $G' \subsetneq G(P)$ is advantaged or if G' can form an advantaged coalition with $i \in S\text{-}G(P)$ such that $i = \max[G',i]$. But then $s_j^P < r_j^o$ for $j \in S - G(P) - \{i\}$, which is a contradiction. To prove necessity, we must show that if $S = G(P)$, then P is not stable, which follows from the fact that in this instance, every country can improve its security value by reforming P so that it is the member of an alliance with a primary threat.

Proof of Theorem 2: We already know that an all-encompassing collective security system is stable (Result 2). So suppose $S\text{-}C$ is not empty.

(Sufficiency, part 2): To see that no $i \in C$ has an incentive to defect in the threat-counter-threat game, suppose i switches from C to $S\text{-}C$ to form $P' = (C\text{-}\{i\}, S\text{-}C+\{i\})$. If $r(C\text{-}\{i\}) \geq r(S\text{-}C+\{i\})$, then since i may tremble back to $C\text{-}\{i\}$, $s_j^{P'} < r_j^o\ \forall\ j \in S\text{-}C$. And by Lemmas 2 and 3, $s_j^{P'} < r_j^o\ \forall\ j \in C\text{-}\{j\}$. So $\forall\ j \in S\text{-}\{i\}$, $s_j^{P'} < r_j^o$. By Lemma 1, $\exists\ C' \in C^*$ with $i \notin C'$, so $s_i^{P'} < r_i^o$. Thus, i will not defect. Alternatively, if $r(C\text{-}\{i\}) < r(S\text{-}C+\{i\})$, then given the conditions of the Theorem, $S\text{-}C+\{i\} \notin C^*$, and by Lemmas 2 and 3, $s_i^{P'} < r_i^o\ \forall\ i \in S\text{-}C+\{i\}$. Thus, i does not defect from C.

(Sufficiency, part 3): Suppose $C \notin C^*$. We have two cases. First. if $r(S\text{-}C) + r_{\max[C]} \geq R/2$, then $C \notin C^*$ because condition ii or iii is violated. Lemma 2 implies that all $j \in C$ are vulnerable to a primary threat. Lemma 3 implies that all members of $C\text{-}\{j\}$, $j \neq \max[C]$, are vulnerable to a primary threat -- but then j is vulnerable as well since $S\text{-}C + \{\max[C]\} \in C^*$. Hence, $s_j^P < r_j^o\ \forall\ j \in C$. Since, by the assumption of the theorem, $C + \{i\} \notin C^*\ \forall\ i \in S\text{-}C$, any defection from $S\text{-}C$ can be punished and no $i \in S\text{-}C$ has an incentive to defect. Hence, $s_i^o = r_i^o\ \forall\ i \in S\text{-}C$. By theorem 1, $(C,S\text{-}C)$ is stable. Alternatively, let $r(S\text{-}C) + r_{\max[C]} < R/2$, so unilateral defection from C can form an advantaged coalition with $S\text{-}C$. After the elimination of $S\text{-}C$, by Lemmas 4 and 1, everyone is the target of some primary threat if alliances are renegotiated. Thus, P is stable (by Theorem 1).

(Necessity, part 2): If $C \in C^*$, then $s_j^P < r_j^o\ \forall\ j \in S\text{-}C$. However, if there is an $i \in C$ such that $S\text{-}C \cup \{i\} \in C^*$, then by the "shaky hand assumption," $s_j^P < r_j^o\ \forall\ C - \{i\}$, in which case all members of $C - \{i\}$ are willing to join a primary threat against i in the play of our threat-counter threat game. And since i can be threatened by a primary threat (Lemma 1), $s_i^P < r_i^P$, so P is not stable (by Theorem 1).

(Necessity, part 3): If $C \notin C^*$ but if $\exists\, i \in S\text{-}C$ such that i can form an advantaged coalition with C, then if there is some probability that i will join in a primary threat against $S\text{-}C\text{-}\{i\}$ whenever i is indifferent, $s_j^P < r_j^o\ \forall\, j \in S\text{-}C\text{-}\{i\}$. So as before, $s_j^P < r_j^P\ \forall\, j \in S\text{-}\{i\}$, in which case $\exists\, C \in C^*$ such that $i \notin C$ (Lemma 1) and P is unstable (by Theorem 1).

Proof of Remark 3: Country 1 cannot be a balancer, otherwise, condition iii is violated. But if we attempt to make 2 a balancer, then condition ii is violated.

Proof of Remark 4: Let $P = (A_1, A_2, \{j\})$, where $A_1 + \{j\}, A_2 + \{j\} \in C^*$. Then $s_j^P < r_j^o$ $\forall\, j \in A_1$ and A_2, in which case, from Lemma 1, $s_j^P < r_j^o\ \forall\, j$, and from Theorem 1, P is unstable.

Proof of Remark 5: If $C \in P$ and $C \in C^*$, then for no $i \in C$ and $K \subseteq S\text{-}C$ is $K + \{i\} \in C^*$ and $i = \max[K + \{i\}]$. So no $i \in C$ has a positive incentive to defect, and $s_{iP} = r^o\ \forall$ $i \in C$. By Theorem 1, P is stable.

Proof of Remark 6: Members of C have two choices: eliminate or not eliminate $S\text{-}C$. If C eliminates $S\text{-}C$, then since alliances must be renegotiated, by Lemma 1, $s_j^P < r_j^o$ $\forall\, j \in C$ in the new system provided that $|C| \geq 4$ (since from Remark 2, every $i \notin C'$ for some $C' \in C^*$), which is what Lemma 4 establishes. If C does not eliminate $S\text{-}C$, then since no $i \in C$ can form an advantaged coalition with members of $S\text{-}C$, $s_i^P = r^o$ $\forall\, i \in C$. By Theorem 1, P is stable.

Proof of Remark 7: No subset of $S\text{-}A$ can coalesce to form an advantaged coalition, so the members of A cannot be threatened with a primary threat. Clearly now, no $i \in A \cap L_o$. For $i \in A \cap L$, if i is the largest country in the newly formed advantaged coalition, then $r(A\text{-}\{i\}) + r_i > R/2$, which contradicts the assumption that $r(S\text{-}A) > r_A$.

Proof of Theorem 3: Let $P = (C, S\text{-}C)$. If $C \in C^*$, then Remark 5 establishes that P is stable. If $C \notin C^*$, but if $r(S\text{-}C) + r_{\max[C]} > R/2$, then Remark 7 establishes that P is stable. And if $r(S\text{-}C) + r_{\max[C]} > R/2$, then P's stability follows from Remark 6.

Proof of Theorem 4: A direct corollary of Theorem 1.

Footnotes

1. Our results require four assumptions, the last of which pertains to the payoffs we associate with alternative outcomes. **First**, a near-predominant country can take advantage of conflicts among other countries to become predominant. Hence, outcomes in which one country is near-predominant are terminal since no one makes a new threat for fear that the near-predominant country will become predominant. **Second**, if i can become near-predominant by implementing a threat or by a resource transfer, i prefers the transfer. So if $i = \max[C]$, then the system is frozen if $S-C$ offers to render i near-predominant. Clearly, if $S-C$ prefers freezing the system, it should transfer to i, since this choice minimizes the resources that it must surrender, and i accepts the offer, because $R/2$ is i's most preferred feasible outcome. **Third**, when countering a threat and whenever it is possible to do so, i chooses a counter that includes all jointly threatened countries in the newly proposed coalition (with the rationale that the threat against $S-C$ makes the formation of $S-C$ less costly). **Finally**, in the case of terminal nodes (when some country is near-predominant), the payoff to country i, $u_i(\mathbf{r})$, equals r_i. For non-terminal nodes, rather than becoming concerned with complex expected value calculations, assume that countries are risk-averse in this sense: if $\mathbf{R}(\mathbf{r})$ is the set of terminal and non-terminal resource distributions that might be reached from $\mathbf{r}$, given the assumed strategies of the players, then $u_i(\mathbf{r}) = \min[R_i(\mathbf{r})]$.

2. Although the alternative "countries make or participate in threats only if doing so promises them a gain" yields an equilibrium, this equilibrium is unstable in that countries have an incentive to defect if there is any chance that others will defect. That is, the equilibrium is not perfect.

3. The actual formal statement of this condition is: For no $C \in C^*$ is there a

winning coalition C' such that $C \cap C' = \{1\}$ and $r(C') > r(C)$;

4. Notice that the threat (150,0,80,70) by {1,3,4} is, like (150,0,75,75), also a primary threat. But if two threats by the same coalition satisfy our requirements, then those threats are strategically equivalent.

5. Recalling that the subscript i on r_i orders the countries from largest to smallest, L_o consists of countries n, n-1, ..., k, such that L_o is losing but L_o plus country k-1 (the next largest country) is winning. In 3-country systems, L_o is always the smallest country.

References

Bendor, **Jonathan** and Dilip Mookherjee. 1987. "Institutional Structure and the **Logic** of Ongoing Collective Action," **American Political Science Review,** 81(1): 129-54.

Boulding, Kenneth. 1968. **Beyond Economics: Essays on Society, Religion and** Ethics. **Ann Arbor: U. of** Michigan Press.

Claude, Innis L. 1962. **Power and International Relations.** New York: Random **House.**

Fox, Annette B. 1959. **The Power of Small States.** Chicago: U. of Chicago **Press.**

Jervis, Robert. (1986). "From Balance to Concert: A Study of International Security Cooperation," in K.A. Oye, ed., **Cooperation Under Anarchy,** Princeton: Princeton Univ. Press

Keohane, Robert O. 1984. **After Hegemony.** Princeton: Princeton U. Press.

Morgenthau, Hans, **J.** 1959. **Politics Among Nations.** New York: Knopf.

Niou, Emerson M.S., Peter C. Ordeshook and Gregory Rose, 1989. **The Balance of Power,** New York: Cambridge U. Press.

Niou, Emerson M.S. and Peter C. Ordeshook. 1990. "Stability in Anarchic International Systems," **American Political Science Review,** 84(4): 1208-34.

______________________________. 1991. "Realism versus Neoliberalism: **A** Formulation," **American Journal of Political Science,** May forthcoming

Olson, Mancur and Richard Zeckhauser. 1966. "An Economic Theory of Alliances," **Review of Economics and Statistics,** 48(3).

Oppenheimer, **Joe.** 1979. "Collective Goods and Alliances: A Reassessment," **Journal of Conflict Resolution,** 23:387-407.

Riker, William H. 1962. **Theory of Political Coalitions,** New Haven: Yale U. Press.

Snyder, Glenn H. 1990. "Alliance Theory: A Neorealist First Cut," **Journal of International Affairs,** Winter, 103-23.

Taylor, Michael. 1976. **Anarchy and Cooperation.** N.Y.: Wiley

Taylor, Michael. 1987. **The Possibility of Cooperation.** Cambridge: Cambridge University Press

Wagner, R. Harrison. 1986. "The Theory of **Games and the Balance of** Power," **World Politics,** 38:546-76.

Walt, **Stephen** M. 1987. **The Origins of Alliances. Ithaca:** Cornell **U.** Press.

Wright, Quincy. 1965. **A Study of War.** Chicago: U. of Chicago Press.

RECENT SOCIAL SCIENCE WORKING PAPERS

748. Boylan, Richard T. "Laws of Large Numbers for Dynamical Systems with Randomly Matched Individuals." August 1990.

749. Kousser, Morgan J. "The Voting Rights Act and the Two Reconstructions." September 1990.

750. Boswell, Laird. "The French Rural Communist Electorate in the 1920s and 1930s." April 1991.

751. Bossaerts, Peter and Hillion, Pierre. "Arbitrage Restrictions Across Financial Markets: Theory, Methodology and Tests." Revised May 1991.

752. Hoffman, Philip T. "Land Rents and Agricultural Productivity: The Paris Basin, 1450-1789." April 1991.

753. Dubin, Jeffrey A. and Wilde, Louis L. "Welfare Economics for Tobit Models." January 1991.

754. Jackson, Matthew O.; Palfrey, Thomas R.; and Srivastava, Sanjay. "Undominated Nash Implementation in Bounded Mechanisms." November 1990.

755. Hoffman, Elizabeth and Spitzer, Matthew L. "The Divergence Between Willingness-To-Pay and Willingness-To-Accept Measures of Value. November 1990.

756. To be announced.

757. El-Gamal, Mahmoud A.; McKelvey, Richard D. and Palfrey, Thomas R. "A Bayesian Sequential Experimental Study of Learning in Games." February 1991.

758. Plott, Charles R. "Will Economics Become an Experimental Science?" December 1990.

759. Dubin, Jeffrey A. and Zeng, Langche. "The Heterogenous Logit Model." February 1991.

760. Palfrey, Thomas R. "Implementation in Bayesian Equilibrium: The Multiple Equilibrium Problem in Mechanism Design." December 1990.

761. Niou, Emerson M. S. and Ordeshook, Peter C. "Alliances in Anarchic International Systems." May 1991.

762. Ordeshook, Peter C. "The Development of Contemporary Political Theory." May 1991.