INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in “sectioning” the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.
DIXON, WILLIAM JAMES

MULTIPLE STRATIFICATION SYSTEMS AND INTERGOVERNMENTAL FOREIGN POLICY BEHAVIOR: A TEST OF TWO MODELS

The Ohio State University

Ph.D. 1980

University Microfilms International

300 N. Zeeb Road, Ann Arbor, MI 48106

18 Bedford Row, London WC1R 4EJ, England
MULTIPLE STRATIFICATION SYSTEMS
AND INTERGOVERNMENTAL FOREIGN POLICY BEHAVIOR:
A TEST OF TWO MODELS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

William James Dixon, A.B., M.A.

* * * * *

The Ohio State University
1980

Reading Committee: Approved By
Charles F. Hermann  
Chadwick F. Alger
Donald A. Sylvan  

Adviser
Department of Political Science
ACKNOWLEDGMENTS

I wish to express my gratitude to the members of my dissertation committee, Chadwick F. Alger, Donald A. Sylvan, and my principal mentor and adviser, Charles F. Hermann, for their patience and guidance.

This research would not have been possible without the generous support of the Mershon Center of Ohio State University. The data utilized in this dissertation were made available, in part, by the Inter-university Consortium for Political and Social Research and the Polimetrics Laboratory at Ohio State. Computing funds were provided by Ohio State's Instruction and Research Computer Center.

Finally, I want to thank my primary companion and source of support, Janice Love, who has helped me in far too many ways to mention here.
VITA

February 10, 1951 ........ Born - Warren, Pennsylvania

1973 ............... A.B., Bucknell University, Lewisburg, Pennsylvania

1974, 1977 ............. Graduate Teaching Associate, Department of Political Science, The Ohio State University, Columbus, Ohio

1974-1976 ............. Graduate Research Associate, Department of Political Science, The Ohio State University, Columbus, Ohio

1976 ............... M.A., The Ohio State University, Columbus, Ohio

1977-1980 ............. Graduate Research Associate, Mershon Center, The Ohio State University, Columbus, Ohio

PUBLICATIONS


FIELDS OF STUDY

Major Field: International Politics

Comparative Foreign Policy. Professor Charles F. Hermann

Teaching International Politics. Professor James E. Harf

Minor Field: Political Theory

Analytic Theory. Professor Stuart J. Thorson
### TABLE OF CONTENTS

ACKNOWLEDGMENTS ................................................ ii
VITA .................................................................. iii
LIST OF TABLES .................................................. vii
LIST OF FIGURES ................................................ ix

Chapter

I. A MULTIPLE SYSTEMS APPROACH TO INTERNATIONAL POLITICS ......................................... 1

   Introduction ......................................................... 1
   A Logic of Inquiry ................................................... 8
   Systems of International Politics ............................. 13
   Stratification and International Systems ................. 46
   Overview .............................................................. 55

II. INTERNATIONAL STRATIFICATION ....................... 62

   The Study of International Stratification ................. 64
   The Explanatory Logic of Stratification ................... 91
   The Behavioral Effects of Stratification ................. 119

III. DATA AND METHODS I: ASCERTAINING INTERNATIONAL STRUCTURE ................................. 148

   Measurement Strategy .......................................... 149
   Indicators of International Structure ..................... 160
   Convergent and Discriminant Validity ................. 231

IV. DATA AND METHODS II: ANALYSIS OF FOREIGN POLICY BEHAVIOR ................................. 249

   Measuring Patterns of Foreign Policy Behavior ........ 250
   Analysis Design .................................................. 295
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Spiegel's Index of International Hierarchy</td>
<td>172</td>
</tr>
<tr>
<td>2. Scales Used in Cox and Jacobson's Power Index</td>
<td>174</td>
</tr>
<tr>
<td>3. Cox and Jacobson's Power Index</td>
<td>176</td>
</tr>
<tr>
<td>4. Transformations of Capacity to Act Indicators</td>
<td>189</td>
</tr>
<tr>
<td>5. Principal Components Analysis of Size</td>
<td>192</td>
</tr>
<tr>
<td>6. Principal Components Analysis of Modernization</td>
<td>193</td>
</tr>
<tr>
<td>7. Principal Components Analysis of Societal Stress</td>
<td>195</td>
</tr>
<tr>
<td>8. Selected Nations Ranked on Capacity to Act</td>
<td>200</td>
</tr>
<tr>
<td>9. Intercorrelations Among Structure Measures</td>
<td>234</td>
</tr>
<tr>
<td>10. Average Convergence and Divergence of Structure Measures</td>
<td>236</td>
</tr>
<tr>
<td>11. Time Sample for Behavioral Data</td>
<td>260</td>
</tr>
<tr>
<td>12. Governmental Actors</td>
<td>261</td>
</tr>
<tr>
<td>13. Representativeness of &quot;Sampled&quot; Nations</td>
<td>266</td>
</tr>
<tr>
<td>14. Number of Events Comprising Monadic Patterns</td>
<td>282</td>
</tr>
<tr>
<td>15. Governmental Actors Representation in Dyadic Patterns</td>
<td>285</td>
</tr>
<tr>
<td>16. Effect Coding for Directional Hypotheses</td>
<td>320</td>
</tr>
<tr>
<td>17. Hypothesis 1: Structural Position and Amount of Foreign Activity (Linear Analysis)</td>
<td>330</td>
</tr>
</tbody>
</table>
18. Hypothesis 1: Structural Position and Amount of Foreign Activity (Nonlinear Analysis) ................................. 338

19. Hypothesis 2: Structural Position and Scope of Action (Linear Analysis) .................................................. 341

20. Hypothesis 2: Structural Position and Scope of Action (Nonlinear Analysis) ............................................. 342

21. Hypothesis 3: Structural Position and Initiatory Behavior .................................................................................. 347

22. Hypothesis 4: Structural Position and Unilateral Behavior ................................................................................. 349

23. Hypothesis 5: Structural Proximity and Communal Problems (4 Event Threshold) ........................................... 352

24. Hypothesis 5: Structural Proximity and Communal Problems (5 Event Threshold) ........................................... 354

25. Hypothesis 6: Structural Proximity and Amount of Foreign Activity ................................................................. 356

26. Hypothesis 7: Structural Proximity and Collaborative Behavior ........................................................................ 360

27. Hypothesis 8: Structural Direction and Internal/Communal Problems .............................................................. 365

28. Hypothesis 9: Structural Direction and Reactive Behavior (4 Event Threshold) .................................................. 368

29. Hypothesis 9: Structural Direction and Reactive Behavior (5 Event Threshold) .................................................. 370

30. Hypothesis 10: Structural Direction and Multilateral Behavior (4 Event Threshold) ........................................... 372

31. Hypothesis 10: Structural Direction and Multilateral Behavior (5 Event Threshold) ........................................... 374

32. Summary of Comparison Outcomes Differentiated by Problem Area ................................................................. 383
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>The Stratification Process</td>
<td>65</td>
</tr>
<tr>
<td>2.</td>
<td>Structural Position and International Involvement</td>
<td>114</td>
</tr>
<tr>
<td>3.</td>
<td>Stages of Events Research</td>
<td>253</td>
</tr>
<tr>
<td>4.</td>
<td>Problem Area Decision Tree</td>
<td>274</td>
</tr>
<tr>
<td>5.</td>
<td>Event Composition of Problem Specific Monads</td>
<td>280</td>
</tr>
<tr>
<td>6.</td>
<td>Event Composition of Problem Specific Dyads</td>
<td>283</td>
</tr>
<tr>
<td>7.</td>
<td>Event Composition of Overall System Dyads</td>
<td>287</td>
</tr>
<tr>
<td>8.</td>
<td>Design for Evaluating Systemic Construct</td>
<td>304</td>
</tr>
<tr>
<td>9.</td>
<td>Hypothetical Relationship Between Multilateral Behavior and Signed Structural Distance (Hypothesis 10)</td>
<td>316</td>
</tr>
<tr>
<td>10.</td>
<td>Scatter Plot of Capacity to Act and Amount of Overall Foreign Activity</td>
<td>335</td>
</tr>
<tr>
<td>11.</td>
<td>Scatter Plot of Military Manpower and Amount of Foreign Military Activity</td>
<td>336</td>
</tr>
<tr>
<td>12.</td>
<td>Scatter Plot of Geographically-Weighted Share and Scope of (Trade) Action</td>
<td>343</td>
</tr>
<tr>
<td>13.</td>
<td>Scatter Plot of Capacity to Act and Scope of (Trade) Action</td>
<td>344</td>
</tr>
<tr>
<td>14.</td>
<td>Scatter Plot of Military Manpower Proximity and Amount of Dyadic Military Activity</td>
<td>358</td>
</tr>
</tbody>
</table>
15. Scatter Plot of Commodity-Weighted Share Proximity and Amount of Dyadic Trade Activity ........................................ 359

16. Summary of Results (4 Event Threshold) .................. 377

17. Summary of Results (5 Event Threshold) .................. 378
CHAPTER I

A MULTIPLE SYSTEMS APPROACH TO INTERNATIONAL POLITICS

Of all the advances that have occurred in the study of international phenomena, perhaps none is more important than the ever-growing tendency to regard the world as the international system.

—James N. Rosenau (1961: 77)

Introduction

Nearly two decades have passed since Rosenau's sanguine assessment of the role to be played by systems approaches to international politics. The source of Rosenau's, and others', optimism was the belief that a systems perspective would provide a new vantage point for thinking about the subject matter of international relations, a vantage point that promised to open new avenues of research, uncover new insights, and ultimately lead to new explanations. Have these early promises been fulfilled? Has the systems idea advanced our knowledge by generating new theories and explanations of international phenomena?

If there is a consensus of opinion on these questions—and it may yet be too early to draw any definite conclusions—it appears to be one of emerging skepticism. Certainly the initial enthusiasm
has faded. The sharpest criticism has charged that, at best, systems approaches have delivered little of value beyond a moderately instructive heuristic, while at worst they have seriously impeded the study of world politics by confusing flimsy analogies and obfuscating language with scientific rigor and theoretical insight (Lieber, 1972; Stephens, 1972; Haas, 1975). Even one as genuinely sympathetic to the systems idea as Kenneth Waltz (1975) has seen fit to severely question the utility of works by the discipline's most prominent systems theorists.

The reasons for this growing discontent are varied and have precipitated numerous prescriptions ranging from wholesale abandonment of systems research to more moderate corrective actions designed to rehabilitate the most problematic aspects of specific approaches. One of the most intriguing suggestions to emerge in recent years calls for reevaluation of the widely held assumption that world politics may be conceived as a single, comprehensive system. An alternative assumption that is considered by its proponents to be both more plausible and more productive claims that "The global political arena is actually the scene of many distinctive systems, many of which overlap and are interdependent, and all of which vary in terms of the actors that participate in them and the issues around which these actors interact" (Lampert and Mansbach, 1976: 1-2). This summary statement of what shall be designated the "multiple systems approach" projects a strikingly different and exceedingly more
complex image of world politics than that entailed by the conventional, single system perspective. What advantages are derived from such a radical adjustment to the traditional way of thinking about international systems?

One potential benefit of the multiple systems approach is the analytical tractability it can bring to the study of international systems. Of course, dissecting international relations into substantively homogeneous sets of phenomena will necessarily involve an overall loss of parsimony; but, by the same token, the narrowed focus and enhanced discriminatory power of the multiple systems perspective may ultimately facilitate inquiry by easing the strain on over extended concepts and generalizations. For example, most systems analysts prefer to conceptualize social activity—including international relations—in terms of an open system that is set apart from an environment but not impervious to it (for example, McClelland, 1966; Kaplan, 1957). If the openness idea is to retain any meaning it must be possible to distinguish systemic interrelatedness from environmental entropy, a task that is highly problematic when all of world politics is folded into a single, comprehensive system. The multiple systems approach, on the other hand, at least begins to alleviate this difficulty by identifying any one system's environment as other systems rather than merely as some amorphous nonpolitical or noninternational residue (Lampert, Falkowski, and Mansbach, 1978).
Adoption of a multiple systems approach might also prove to be an effective means for dealing with apparent anomalies in the behavior of national governments and other global actors under the single system assumption. In the 1960s, for instance, the enmity between the two superpowers was nowhere more evident than in Vietnam where the United States, as an active combatant, and the Soviet Union, as a principal supplier of weapons and equipment, were poised on opposite sides of a decade-long war. Yet during much of this period representatives of these two governments met regularly in a series of discussions that eventually produced the arms limitation agreements known as SALT I. To take another example, consider the curiously ambivalent relationship between Zambia, a consistent supporter of African liberation movements, and its white-ruled neighbors in southern Africa. Because of Zambia's acute dependence on Southern Rhodesian rail facilities for transport of its foreign trade goods, its commercial relations with the white minority regime continued long after Rhodesia's unilateral declaration of independence in 1965. Moreover, when Zambia finally resolved to end its reliance on the Rhodesian route in 1973, it found itself in the no less awkward position of having to increase its commercial ties to the colonial regimes in Angola and Mozambique (Sklar, 1974).

Both illustrations reveal the extent to which foreign policies can vary by issue area. Although such cross-issue differences are by no means incompatible with the single system assumption, neither are they made any more comprehensible by fitting them into a single,
overarching system. The multiple systems approach, on the other hand, might actually anticipate such differences by organizing each discrete system around a coherent substantive issue that reflects the values, purposes, and particular capabilities of its participant units. Of course, it remains to be seen whether these advantages are real or illusory, whether the multiple systems interpretation advances the cause of systems inquiry or mires it in unneeded complexity.

A rational appraisal of the multiple systems approach hinges on two interrelated questions. The first concerns the methodological status of its fundamental assumption that world politics is better envisaged as a congeries of distinct and distinguishable systems. Does this assumption represent a factual claim about reality or is it an analytical construct devoid of any empirical content? Unfortunately the designers of the multiple systems model furnish no transparent answer to this question; indeed, at times they seem to vacillate from one position to the other. For example, we are told that "systems do not exist 'out there' awaiting discovery...systems do not exist apart from the perceptions of those who create them." Yet on the preceding page the authors implied just the opposite, declaring that "the existence of a single global political system is at least logically possible..." and furthermore that "...some analyst might eventually demonstrate the existence of such a system..." (Lampert, Falkowski and Mansbach, 1978: 145, 146). Despite the authors' ambiguity, an answer to our first question is easily obtained
once we understand the distinction between what we shall call the system construct and the system referent. The former is a mental image intended as an allegory of the latter, which refers to that portion of the real world under investigation. Allegories, or models, are always approximations of empirical phenomena (Sutherland, 1973). The multiple systems assumption is a construct having no physical existence itself but it approximates a systems referent which does exist and in that sense it represents a factual claim about some aspect of the real world.

The second of our two questions may already be evident for it ensues directly from the first. Once it is established that the assumption of multiple systems is an empirical statement we are in a position to ask a far more demanding question: How well— that is, how faithfully— does the multiple system model approximate reality? The research reported here represents an initial attempt to investigate this question in the hope that such an exercise will lead to greater understanding of the form and function of what is commonly designated the international system.

In order to insure a rational basis for our inquiry we adopt the methodological posture espoused by Imre Lakatos (1970), the late philosopher and historian of science. In brief, Lakatos maintained that scientific knowledge advances only when competing theories are set against observational data and one is shown to have excess empirical content over the other. According to this
view, which is examined in greater detail later in this chapter, it
is not enough for a single theory to be confronted with empirical
facts. To better reflect this methodological stance we can reformu­
late our principal question: Does the multiple systems construct
advance our knowledge by more closely approximately reality than
the single system model, its chief competitor?

The main body of this chapter addresses a number of preliminary
issues and develops a research strategy to be followed in subsequent
chapters. To preview that development, we will argue that successful
completion of our task is contingent on the delineation of conceptually
clear and empirically verifiable criteria for delimiting one system
from another. Once such criteria are established it becomes possible
to distinguish a world organized as a single, composite system from
a world of multiple systems. Our criteria for system change will
be based on changes in system structure, a departure from more common
techniques that rely on interaction densities. To apply these
criteria effectively we will need a relatively explicit, a priori
image of system structure that is compatible with both the single
system and multiple systems approaches. Accordingly, we will assume
that systems are structured in a hierarchical fashion by an unequal
distribution of socially desirable characteristics, such as power,
wealth or prestige; in short, we will assume that international
systems are stratification systems. This assumption is especially
appropriate because studies of international stratification have
traditionally followed the single system approach by focusing on one or more indicators of status or rank to define a single, overarching stratification structure. In contrast, a multiple systems version would posit an overlapping or layered arrangement of several issue-based stratification systems bounded by discontinuities in membership and member ranks as well as patterns of behavior. We will then propose that these two contrasting views of international structure by comparatively evaluated against observational data measuring various properties of governmental foreign policy behavior.

A Logic of Inquiry

Let us state at the outset our commitment to the methods of science as the most persuasive, systematic and rational means for accumulating knowledge about the world. One implication of this commitment is that science must provide some procedure whereby conjectures are adjudged according to certain criteria before they are admitted, even provisionally, into the realm of scientific knowledge. Certainly one of the most significant of these criteria concerns the extent to which our putative knowledge squares with observable fact. Our models and theories must survive rigorous empirical tests if they are to enhance our comprehension of the world. The set of procedures and standards that constitute any such test, and hence that preserve the integrity of the knowledge base they serve to define, must ultimately rest upon the logic of the normative research
methodology that guides scientific inquiry. Philosophers of science have faced few issues as perplexing as those surrounding the discovery of a rational standard for establishing the empirical basis of scientific knowledge.

Since the breakdown of "justificationism"—the thesis that scientific propositions can be proven true, either in an absolute sense or within some range of probability—the prevailing methodological position has involved some variant of "falsificationism."

The simplest brand of falsificationism concedes that theoretical propositions can never be confirmed, but claims that once confronted with contradictory facts they can be ultimately and conclusively refuted. This view, styled "dogmatic falsificationism" by Lakatos (1970), can be criticized on several grounds. In the first place, it mistakenly assumes that observational propositions can be wholly and unambiguously distinguished from theoretical propositions. Secondly, it implies that propositions can be disproved by facts, although the rules of elementary logic dictate that propositions can be disproved only by other propositions. Finally, it is incongruent with the history of scientific practices where the strategic use of a ceteris parabus clause often renders theories incapable of specifying falsifying observations.

Considerably less objectionable is a "methodological falsificationism" of a "naive" sort which holds that observation is pursuant to intersubjective background knowledge taken to be unproblematic by
the scientific community. Theories, according to this view, are not
disproven in the face of contradictory facts; rather, they are re-
jected as inconsistent with currently accepted knowledge. But this
position, like its dogmatic counterpart, is at variance with the
methods that historically have guided scientific inquiry on at least
two counts. It presumes, first, that an empirical test is a bilateral
contest between theory and experiment (or data), and second, it
constrains scientific interest to outcomes that are refutations.

Lakatos has proposed a more "sophisticated" variety of methodo-
logical falsificationism that is immune to criticisms of the sort
mentioned above. According to this view, falsification is understood
to be the process of giving up one theory in favor of another more
desirable theory under a stringently defined set of conditions:

The sophisticated falsificationist regards a scientific
theory T as falsified if and only if another theory T'
has been proposed with the following characteristics:
(1) T' has excess empirical content over T: that is,
it predicts novel facts, that is, facts improbable in
the light of, or even forbidden by T; (2) T' explains
the previous success of T, that is, all the unrefuted
content of T is contained (within the limits of observa-
tional error) in the content of T'; and (3) some of the
excess content of T' is corroborated (Lakatos, 1970:
116).

Thus, science advances and knowledge accumulates as theories succeed
one another in accordance with the strict regimen outlined by
Lakatos. Furthermore, it is this progression of theories, or
"research program" as Lakatos prefers to call it, that must bear the
brunt of critical appraisal.
Two quite distinct but complimentary sets of assumptions are incorporated in the body of a research program. One set of assumptions will comprise the program's "hard core"; these are simply accepted as givens at the outset, thereafter to be kept safe from critical scrutiny. They are protected by two methodological maxims: the "negative heuristic" which deflects criticism away from the hard core, and the "positive heuristic" which draws criticism toward a second, less critical set of assumptions serving as the program's "protective belt." It is the protective belt that must withstand the constant adjustment and revision brought to the program by succeeding theories. In an illuminating aside, Terrence Ball (1976: 164, f.n. 13) explains that "Lakatos' point...is that we can never get anywhere if we dwell always upon the 'fundamental assumptions' of a theory (or series of theories), instead of its 'payoff'. The 'hands-off' policy prescribed by the negative heuristic allows the scientist to get on with his work instead of having to constantly defend his core assumptions."

How does Lakatos' sophisticated falsificationism inform our study of the international system? First of all, we embrace his interpretation of the falsification process as a three-sided struggle with two contending theories pitted against empirical observation. Accordingly, our purpose is not to determine whether the multiple systems interpretation presents a true rendition of global politics, but only whether it does better than the competing single system perspective. At this point we must address a matter of usage.
Strictly speaking, this study is concerned with contending approaches or frameworks that cannot be themselves be designated theories. They can, however, provide alternative interpretations, or models, for a theory. It is in this interpretative sense that we will henceforth refer to the single system and multiple systems models (see Kaplan, 1964).

Secondly, we can identify a hard core of assumptions that remain intact when we shift from a single system to a multiple systems interpretation. Several of these assumptions define the stratification theory of system structure that is applied within the confines of both approaches. These tend to become rather specific and concern such matters as the hierarchical ordering principle that dominates the system and the patterns of behavior that are expected to occur at various points within the system. Other hard core assumptions are framed at a more general level. For example, both approaches assume that there is a discernible quality of "systemness" to world politics and both assume an interplay between the structural configuration and events occurring beneath the systemic level.

Thirdly, we contend that the conventional image of a single global political system is most appropriately viewed as a protective belt assumption. That this image is rejected or at least modified by a sizable literature on regional subsystems would seem to reinforce our estimate of its protective belt status. Finally, regarding Lakatos' criteria of falsification, we can state with certainty
that the multiple systems model predicts novel facts forbidden by the single system model. Of course, how well these predictions stand up when confronted with observational data is the fundamental question of our inquiry.

Systems in International Politics

References to "international systems" have appeared repeatedly throughout the short space of this chapter, yet conspicuously absent has been any effort to explicitly convey the meaning imputed to this much overworked concept. This task is addressed in the present section. We begin by briefly considering systems in general and then move to more specific interpretations of international systems. We will examine a variety of measures for dividing the system into more manageable components paying particular attention to efforts organized around distinct substantive areas. The section concludes with a discussion of a positional interpretation of system structure. Numerous surveys and critical reviews of the international systems literature have been completed in recent years and it should be noted that what follows is not intended to duplicate those efforts (for example McClelland, 1972; Stephens, 1972; Weltman, 1973; Waltz, 1975; Mitchell, 1978).

The System Construct

The emergence and subsequent growth of systems approaches has been one of the more noticeable developments to have occurred in the natural and social sciences since the end of the Second World
Practitioners of every nearly every branch of scientific inquiry—from astronomy and geology to sociology and political science—have adopted one or another variant of the systems perspective as a way to organize an increasingly complex array of observations and impressions. Many early enthusiasts saw the potential for unifying science under the systems idea and diligently set about formulating abstract and highly generalized definitions of the concept able to span the subject matter of ostensibly divergent disciplines.

The classic definition of systems as "sets of elements standing in interaction" is from Ludwig von Bertalanffy (1957: 3), universally considered to be the founder of general systems theory. Another frequently cited definition stipulates a system to be "...a set of objects together with relationships between the objects and between their attributes" (Hall and Fagen, 1957: 18). Similar in denotation though somewhat more engaging in its simplicity is the following definition taken from an introductory text on political analysis: "A system is a set of things related in some way, so that changing or removing any one thing in the set will make a difference to other things in the system" (Strickland, Wade and Johnston, 1972: 7). Oran Young (1968: 15) has summarized the common idea expressed in these and other definitions of abstract systems as "...a group of objects or elements standing in some characteristic structural relationship to one another and interacting on the basis of certain characteristic processes."
One of the most bedeviling aspects of systems concerns their physical existence. Some would argue that systems do exist as pockets of interdependence in an otherwise random or empty world. In this view it is the analyst's job to locate empirical systems, define their boundaries, and map their internal structure. In contrast, others hold that systems are intellectual devices imposed by the analyst to circumscribe and focus attention to relevant aspects of some empirical domain. Earlier we introduced a distinction between the system construct and the system referent. That distinction will prove instrumental in charting a course between these two positions. Recall that a system construct is wholly an analytical device, a mental image that has no existence apart from the analyst who conceives it; a system referent, on the other hand, does have an independent existence and represents that portion of reality modeled by the system construct. We do not deny that some parts of the world are more interrelated than other parts, but the analysts' chief task is not to ferret out these independencies, nor is it to impose systems at will on arbitrary conglomerations. In our view the principal intellectual task of the systems analyst is to determine the degree of congruence between the system construct and the system referent and, where the degree of congruence falls short, to make the necessary adjustments.

For the present we will confine our interest to system constructs, specifically to international system constructs. It should be no
surprise that students of international relations were quick to embrace the systems idea embodied in the definitions given above; indeed, it seemed tailor made for a discipline described by one prominent scholar as "...the study of interactions between certain kinds of social entities, including the study of relevant circumstances surrounding the interactions" (McClelland, 1966: 18). The strong intuitive correspondence between the subject matter of international relations and the abstract notion of a system was a hopeful sign that there might also be significant areas of correspondence, or isomorphisms, with other, more easily grasped system constructs residing at a lower level of abstraction. Even without the benefits of possible isomorphisms, the systems idea supplies a valuable perspective for viewing large scale aspects of international affairs.

What, then, is encompassed by an international system construct? At minimum, an answer requires specification of the objects or units that populate such a system and delineation of the interactions or relationships that bind them together. Unfortunately, but not unexpectedly, there is little evidence of widespread agreement on either of these two points. In a recent work pointing out the system's role in shaping governmental foreign policies, Maurice East (1978b: 145) declared that "an international system refers to the patterns of interactions and relationships among the major territorially based political actors existing at a particular time." Among other things, East suggests that international systems are bounded in time, as do
Kaplan (1957) and Rosecrance (1963) for example. But this position is strongly at variance with the timeless image depicted by Charles McClelland (1966: 20):

The conception of the international system is an expanded version of the notion of two-actors-in-interaction. A view of a whole phenomenon is involved. The outermost boundaries of international relations are suggested if we imagine all of the exchanges, transactions, contacts, flows of information, and actions of every kind going on at this moment of time between and among the separately constituted societies of the world. To this picture in the mind we should add the effects created within societies from all such interflowing events in earlier times, both of the immediate and the more remote past. Finally, the stream of these actions and responses should be conceived as moving on into the future of tomorrow and beyond, accompanied by expectations, plans, and proposals of all observers of the phenomena. This total picture is the reference intended in the term the international system.

McClelland has envisaged an astoundingly complex picture of the international system, but it is not a comprehensive one in at least two respects: the reference to "separately constituted societies" is too restrictive regarding the objects comprising the system and the emphasis on interactions seems to exclude more passive types of relationships. We find the system formulation proposed by J. David Singer (1969) to be instructive on both counts. Singer has urged that the nation-state be recognized as only one--albeit an exceedingly important and longlasting one--among several reasonably distinct classes of entities that regularly engage in international political interactions. Others would include intra-national entities, extra-national or transnational entities, international coalitions, and international organizations. In addition, Singer has advanced some
useful distinctions regarding the interconnections that hold systems together. He suggested that accumulations of discrete actions, or interactions, be distinguished from relationships, which do not involve behavior or physical movement; the latter are then further divided into relationships of attribute similarity and interdependence relationships, which indicate some sort of bonding or linkage between entities that is not merely a summarization or derivative of their separate attributes.

Clearly an effort to develop a comprehensive image of the international system must take account of a bewildering edifice of interactions and relationships linking together a multitude of vastly differentiated entities existing at various levels of social organization. Furthermore, assuming that such an image could be constructed, it is uncertain that it would be of any value beyond instilling in us a new appreciation for the complexity of international relations. In fact, few researchers make an attempt to deal with the international system comprehensively and those that do (for example, Kaplan, 1957; East, 1978b) refrain from a complete specification of its objects and interrelationships. The far more common approach is to reduce this overwhelming complexity by dismantling it, that is, by focusing attention to comparatively small segments of the system at any one time. We can do this and usually do it successfully because, as McClelland (1966: 97) has observed, "The international system is multidimensional; what is transpiring in one sector or at
The Search for Boundaries

The very suggestion that researchers routinely split apart the international system seems to contravene the holism and interconnectedness so fundamental to the systems idea. Nonetheless, the literature abounds with examples of studies that do just that. In this subsection we briefly examine the rationale underlying this practice and then point out some familiar applications.

Once again, it will be helpful to have in mind the distinction between a system construct and a system referent. When we speak of narrowing our attention to a selected portion of a system we are actually reducing or limiting the scope of our system construct. Obviously, this will also have an effect on the congruence obtained between construct and referent. In most cases the effect will be slight—-not so much because of the way the construct is formulated, although that is certainly relevant—-but because of a particular organizational form associated with the system referent. Put differently, the world—-that is, reality—-tends to be structured in a manner that permits and even encourages selective attention to only a few of its aspects at any one time. Designating this ubiquitous structural form a "hierarchy," Herbert Simon has brilliantly elucidated its amenability to decomposition. 3 Simon's (1969: 99) great insight was that complexity is easily managed when it is organized hierarchically
since it is possible to "...distinguish between interactions among subsystems, on the one hand, and the interactions within subsystems—that is, among the part of those subsystems—on the other. The interactions at different levels may be, and often will be, of different orders of magnitude." Successively dividing a complex system into simpler, more comprehensible parts is much like disassembling a set of Chinese boxes. We can discriminate between these separate parts, or subsystems, by considering their interaction patterns. Only occasionally will hierarchies be completely decomposable with no appreciable interactions among subsystems, but frequently they will be "nearly decomposable" and subsystems will appear as dense clusters of interactions separated by perceptable discontinuities. Now let us examine how these discontinuities have been pursued by students of world politics.

Perhaps the most common and intuitively plausible scheme for decomposing the international system is one based on the regional subsystem. Since the time Leonard Binder (1958) initially proposed the idea in response to Kaplan's (1957) single system analysis, a sizeable literature has evolved around the theme that certain geographical areas, such as the Middle East or Southeast Asia, constitute partially self-contained systems, or subsystems, that are only loosely connected to the larger global political system. Although conceptualizations vary, most authors contend that regional subsystems are indicated where geographically proximate actors are
bound together in a dense pattern of recurring interactions and relations such that the area is recognized as distinctive by both internal and external observers (Thompson, 1973). The notion of hierarchic complexity is transparent in the presumption that interactions within regional subsystems are discernible from interactions across subsystems. This is not to deny the existence of the latter, nor is it to diminish their significance. In fact, the intrusion of one subsystem into another, such as Western Europe's penetration of certain African subsystems, is often an important element in the regional configurations studied by the proponents of this approach (Cantori and Spiegel, 1970). Nevertheless, the main focus of attention remains within a subsystem whose boundaries are suggested by qualitative as well as quantitative discontinuities of members' interactions.

Although its logic is reasonably compelling, the regional subsystem approach is by no means the only formula for decomposing the international system. Three other methods come to mind, each of which represents a loose extension of the regional subsystem logic. The first employs a systemic perspective to portray the interactions and relationships between any two (or three or four) actors irrespective of their geographic location. Authors that opt for this approach usually acknowledge that their chosen dyad is only a small part of a larger system, but they justify their narrowly circumscribed view as a necessary measure to bring forth the finely textured detail characteristic of many international relationships. Bruce Russett's
A fine study of twentieth century British-American relations is perhaps the classic example of a systems-oriented inquiry reduced to the dyadic level. A second approach combines elements of the regional subsystem perspective with the dyadic formula. The subsystem—whether it be regional or some other type, such as an alliance—is conceived as a unit that interacts with other subsystems. Illustrating this inter-subsystemic approach are the studies of NATO and Warsaw Pact interactions by Galtung (1966) and Tanter (1972).

A third device for simplifying the international system focuses less on overt interactions than on more passive types of relationships. One manifestation of this approach is exemplified by Russett's (1967) attempt to delineate clusters of nations, or regions, on the basis of their similarity on certain shared attributes. Russett demonstrated that geographic position was only one of an endless number of attributes on which nations could be proximate or distant, and that these attributes could serve as the relational basis for defining diverse subsystems. Hayward Alker (1968) presents us with a different interpretation of this approach. By showing that certain causal processes need not be invariant across groups of countries, Alker has provided a firm basis for a decomposition scheme that postulates discontinuities in causal processes rather than discontinuities in interaction patterns.
We can identify yet another popular formula for untangling complexity in the international system. At this point let us move away from formulations premised, however loosely, on the decomposability of hierarchic systems to consider an approach that utilizes a more functional or substantive mode of differentiation. The following passage by John Burton (1968: 8) outlines a model of a world constructed of numerous functionally specific systems:

The map of this society would appear like millions of cobwebs super-imposed one upon another, covering the whole globe, some with stronger strands than others representing more numerous transactions, some concentrated in small areas, and some thinly stretched over extensive areas. Each separate cobweb would represent a separate system—trade flows, letters exchanged, tourist movements, aircraft flights, population movements and transactions in ideas, cultures, languages, and religions...

Although not everyone will agree with Burton that separate interaction networks constitute distinct systems, it can hardly be disputed that concentrating on a particular type of transaction is a familiar tactic for investigating systemic patterns (Brams, 1966; 1969; Bernstein and Weldon, 1968; Christopherson, 1976). The idea that each kind of interaction or transaction can be treated as a separate system is especially relevant to the present study because of its close resemblance to the multiple issue-based systems model discussed in the opening paragraphs of this chapter. Now that we have examined several of the more common procedures for dismantling the international system it is time we take a closer look at the issue-based model.
Issue-Based Systems

Despite any impression to the contrary left by our earlier discussion, it is by no means a novel idea to assert that there exist many international systems each revolving around a distinctive issue area. In the field of international politics the basic notion of the issue area concept—that international political actors vary their behavior depending on what substantive interests or values are at stake—is most closely associated with the work of James Rosenau; and indeed, it is Rosenau (1963: 115) who first posited a connection between issue areas and international systems:

I find myself intrigued by the idea of systemic analysis which gives central prominence to the concept of issue-areas. Persuasive evidence is available to show that lesser political systems—that is, local and national ones—function differently in different issue areas, that each area elicits a different set of motives on the part of different actors, that different issues, and that therefore the different interaction patterns which result from these variations produce different degrees of stability and coherence for each of the issue-areas in which systemic processes are operative.

The "persuasive evidence" to which Rosenau made reference is drawn from the community power literature, particularly Robert Dahl's (1961) famous inquiry into the politics of New Haven. Dahl's findings, as Rosenau suggests in this passage, evinced practically no overlap of political leadership in three types of local issues, urban renewal, education, and nominations. While not apparent at the outset, issue-based systems were clearly the dominant theme by the conclusion of this short essay:

The functioning of international systems, in other words, is a consequence of the particular complex of issue-areas
that are active at any one time. The complex is constantly changing because some issue-areas are continuously becoming dormant and others are continuously being activated. By probing and comparing the major types of issue-areas, therefore, we ought to be able to penetrate more incisively and expose more clearly the dynamics of international systems (Rosenau, 1963: 117).

Although the emphasis on issue areas would surface again and again, the major thrust of Rosenau's work began to veer toward the foreign policy orientation for which he is best known. Nevertheless, there is still an occasional glimmer of the issue systems idea. For example, in his widely celebrated pretheories article, Rosenau (1966: 74, f.n. 90) proposed a distinction between horizontal systems—those based on geographic units or functional institutions—and vertical systems which, not surprisingly, are "...conceived to encompass a set of interdependent procedures whereby a cluster of values within an issue-area is allocated by either a single horizontal system or a fusion of horizontal systems."

Speculating as to why the issue system concept had been so long neglected, Rosenau suggested what is potentially its most damaging counter argument. Because political actors ("horizontal systems" in his terminology) continuously engage in cross-issue bargaining and compromise, the boundaries of issue systems may be so permeable as to "...at best be obscure and at worst be nonexistent..." (Rosenau, 1966: 78). He then responded to his own challenge by maintaining that at least some issues, arms control being a notable example, seem to be reasonably well insulated from adjacent issues. Such
insulation is facilitated under two sets of conditions: the first is when actors share goals in one area but lack a basis for agreement in most other matters; the second is when actors disagree in a single area but otherwise are in agreement. Furthermore, he reasoned that even if issue areas are poorly insulated it does not follow that their boundaries will be imperceptible: "The resolution of the conflict in the vertical system, i.e., the way in which values are allocated, may be affected by the intrusion of other issues, but the distinctive behavior which forms the boundaries of the system may well continue unaltered." (Rosenau, 1966: 79, f.n. 99). Finally, Rosenau (1966: 81) concluded consideration of the boundary maintenance problem by declaring: "None of the argument against the construction of vertical political systems out of identifiable issue-areas fully offset the compelling evidence that horizontal systems function differently in different areas."4

A year later the issue systems idea reappears in a piece on foreign policy as an issue area, but it is mentioned only in passing. Rosenau (1967: 15) interpreted findings from the United Nations voting literature (particularly Hovet, 1960) to suggest that, "insofar as the functioning of international organizations are concerned, there are several international systems and that to comprehend the processes of any one of them is not necessarily to grasp how the others function."
Apart from Rosenau's work, the issue area concept has received a varying amount of attention from students of foreign policy (O'Leary, 1976; Zimmerman, 1971), regional integration (Nye, 1971; Lindberg, 1971), and UN voting behavior (Alker and Russett, 1965), but until the recent contributions of Lampert and his associates (Lampert and Mansbach, 1976; Lampert, Falkowski and Mansbach, 1978) there was virtually no thought given to reviving the issue systems idea. Nevertheless, there have been studies that, with only a modest amount of interpretive liberty, can be seen as pointing in this direction. Let us consider two.

Some years ago in an article focusing on the relations between the United Nations and the international systems, Stanley Hoffman sketched an imaginative metaphor illustrating the diminished significance of the customary distinction between high and low politics. Hoffman (1970: 401) observed that "...the competition between states takes place on several chessboards in addition to the traditional military and diplomatic ones: for instance, the chessboards of world trade, of world finance, of aid and technical assistance, of space research and exploration, of military technology, and the chessboard of what has been called 'informal penetration'." Clearly, Hoffman is invoking the issue area theme, but more importantly he extends the metaphor by declaring that each chessboard (or issue area) has "rules of its own," and thus implies the separability of issue areas. Do such "rules" supply the basis for distinctive issue-based systems?
Although an affirmative answer might reasonably be argued, Hoffman himself would resist such an interpretation for two reasons: first, he raises the old question of issue permeability, though in a new light: "...depending on the national situation a state may be able to offset its weakness on one chessboard thanks to its strength on another or else by prevented from exploiting its strength on one because of its weakness on another" (Hoffman, 1970: 401). Second, nowhere—neither here nor in other works—does Hoffman ever entertain the possibility of concurrent international systems. He only rarely uses the plural form, but when he does the reference invariably is to successive systems.

A more direct and apparently more sympathetic reference to the issue systems idea is contained in a recent volume by Keohane and Nye (1977) on international regime change. International regimes are arrangements of generally accepted procedures governing interstate and transnational relations, the "rules of the game" in Hoffman's metaphor. In the course of their analysis the authors consider several possible models or putative explanations of the regime change process. One of these, labeled "issue structuralism" by the authors, reconstructs the logic of an "overall structure model" introduced earlier, except for one exceedingly important difference: "...unlike the overall structure explanation, issue structuralism does not predict congruence of power across issues" (Keohane and Nye, 1977: 50). Insofar as regime change is concerned, both models place a high
premium on power as the instrument effecting creation of new regimes and enforcement of old ones. "A basic assumption of the issue structure model, however, is that although states may be tempted to draw linkages among issues, such linkages will be generally unsuccessful. The premise of issue structuralism is that power resources in one issue area lose some or all of their effectiveness when applied to others" (Keohane and Nye, 1977: 50).

Although Keohane and Nye do not say so explicitly, it does not seem too far from their meaning to infer that issue structuralism might operate as a set of issue specific international systems. In fact, they ventured close to this interpretation themselves when, commenting on the necessity of separate analyses of issue areas, they asserted that "within each issue area one posits that states will pursue their relatively coherent self-interests and that stronger states in the issue system will dominate weaker ones and determine the rules of the game" (Keohane and Nye, 1977: 50-51, emphasis added). We should note, however, that in all probability this usage of "system" merely provided an expedient substitute for "issue area" and was not intended to convey the more formal meaning of the concept. Ultimately, we must conclude that the issue structuralism of Keohane and Nye, along with Hoffman's chessboard metaphor, are only suggestive of a multiple issue systems model of world politics.5

As we have demonstrated in this short review, the issue systems idea is not a new one--Rosenau discussed it more than a decade ago.
More recently this idea has been expressed in a series of collaborative papers by Richard Mansbach, Donald Lampert, and Lawrence Falkowski. Earlier in this chapter we indicated that their formulation was the prime impetus for the present inquiry; it is time we examined that formulation in greater detail.

The systems model advanced by Lampert and his associates is seen by its authors to differ in two important respects from more conventional frameworks. In addition to postulating multiple systems, their model assumes a transnational perspective by rejecting the commonly held notion that international systems are populated solely by nation states. The assumption that global systems encompass a broad spectrum of actors seems to be a sensible one on its face and deserves little in the way of further comment (see Keohane and Nye, 1971; Young, 1972, Mansbach, Ferguson and Lampert, 1976). We shall focus instead on the presumption of multiple systems, the essential features of which are summarized in the following passage:

Political systems, including international ones, may be said to vary along several significant dimensions including, though by no means limited to, time, space, participant actors, and issues at stake. The last two are of particular significance because they provide us with the analytic tools to discriminate among systems, compare them, and examine the degree to which they overlap or are linked. Although the structure of a political system is characterized by several sets of properties that are complexly related, in the last analysis each system features a unique cast of purposive actors behaving around a discrete set of issues (Lampert, Falkowski and Mansbach, 1978: 151).
Three interrelated aspects of their formulation demand attention: first, how do the authors conceive of issues and issue areas? Second, what sort of linkages occur between systems (or between issues)? Third, what are the criteria to discriminate between systems? We shall consider each of these questions in turn.

If international systems are to be conceived as revolving around "a discrete set of issues" then the meaning imputed to the issue concept is of critical importance to the formulation's internal logic as well as its analytic tractability. The authors tell us that "issues provide the substantive basis for a given system...an issue is essentially a subject matter concerning which actors can desire a diversity of possible outcomes" (Lampert, Falkowski and Mansbach, 1971: 151). Others who have explored the political implications of issues and issue areas (for example Rosehau, 1966; O'Leary, 1976) have stipulated that issues are inherently divisive and arise only when certain values or potential outcomes are disputed by the parties involved. In other words, an issue ensues only when some allocation of values is, quite literally, "at issue". Lampert and his associates object to this interpretation as too restrictive because it eliminates from consideration a broad range of areas over which there may be genuine consensus regarding desired outcomes.

The fundamental aspect of an issue, then, is its subject matter; whether or not it is controversial is irrelevant for definitional
purposes though, of course, not necessarily for analytic purposes. Furthermore, these subject matters must be inductively derived, although the authors do not tell us what level of abstraction is most appropriate. The authors refrain from using the familiar term "issue area", preferring instead such designations as "issue package" and "superissue" to express the idea that a set of related issues cluster together within an issue system. Given their definition of issue, it might be expected that an issue package would encompass a group of issues particularly close in substantive content. In fact, an issue package is based not on a substantive relationship but on a behavioral one. It is assumed that issues induce certain patterns of behavior on the part of affected actors; an issue package, then, consists of those issues eliciting similar patterns of behavior regardless of subject matter.

One of the chief advantages of the multiple systems approach is the clarity it brings to the often murky distinction between an international system and its environment, since "each system forms part of the environment of every other system..." (Lampert and Mansbach, 1976: 4). In addition, an issue system is assumed to be open and, hence, may be linked to its environment—that is, to other systems. Two types of linkages are discussed by Lampert and his associates. One of these they label "actor sharing" to indicate a situation in which two or more systems involve behavior of a common set of participant actors—the greater the number of common actors, the tighter the linkage between systems. The second type is an "issue linkage"
that occurs when the pattern of behavior associated with a specific issue exhibits characteristics common to two or more issue packages and thus is considered part of two or more issue systems.

But, as we have seen, these are not the only ways issues can be interconnected. How do the authors deal with the possibility of less congenial linkages of the sort discussed by Rosenau and Hoffman? Recall that Hoffman assumed issues to be pervasive in the sense that actors' influence and capabilities—or conversely, weaknesses—in particular issues were fungible and thus could be transferred to other issues. The authors discount this possibility and take a position similar to Keohane and Nye's issue structuralism argument by asserting that "actor objectives...express desired future states in relation to specific issues; actor capabilities can generate influence only in specific situations that derive from particular issues" (Lampert, Ferguson and Mansbach, 1978: 152, authors' emphasis).

Rosenau, on the other hand, was concerned about issue systems maintaining their boundaries in the face of linkages resulting from cross-issue bargaining processes. Although Lampert and his associates cite Rosenau's discussion of the boundary maintenance problem, they fail to address it with regard to their own formulation. What makes this omission significant is the fact that the "actor sharing" linkage, which they do discuss, is all that is necessary for actors to engage in bargaining across common issues.
The question of system distinctiveness or distinguishability is obviously crucial for any approach premised on the notion of multiple systems, for unless it is possible to invoke clearly defined and easily operational criteria to discriminate between systems, the fundamental premise must remain in doubt. To ask what distinguishes one system from another is to ask how a system is set off from its surrounding environment. An acceptable answer to this question must go beyond consideration of system boundaries to address the very foundation of the system concept, namely, its internal composition or structure. Lampert and his associates are reasonably clear about what they mean by system: "We proceed from the classic definition of a system as a 'set of elements standing in interaction.' The elements in each system in world politics consists of a diverse cast of purposive actors interacting around a discrete set of issues that imply different stakes for the participants" (Lampert and Mansbach, 1976: 4). Hence, systems are defined by interaction structures, that is, by concentrations of issue-specific behavior into discernible patterns. Furthermore, the authors claim that such patterns can be delineated by using a factoring technique to reveal groups or "packages" of issues that covary according to the amount of behavior they elicit from particular actor-recipient combinations. They conclude that each factor emerging from such an analysis "...would represent a unique concatenation of actors and issues recognizable through behavior; in other words, a system" (Lampert, Falkowski and Mansbach, 1978: 152).
It would appear that Lampert and his associates have followed a logic that in some ways resembles Simon's principle of hierarchical decomposition. By arguing that interrelationships within subsystems could be distinguished from the more feeble interrelationships occurring among subsystems, Simon assumed that systems and subsystems were structured as distinctive networks of interrelated elements. Likewise, Lampert and his associates also advance a network structure interpretation of system composition, although it is not at all clear whether their networks (or systems) represent sets of similarly covarying issues—as their methodology suggests—or whether they consist of sets of actors that pattern their behavior around a common class of issues—as their conceptual discussion suggests. In other words, there is an appreciable degree of slippage between operational and conceptual definitions of system.

Other problems, ranging from the restrictive assumptions of the factor model to problematic interpretation of results, arise either directly or indirectly as a result of the authors' factor analytic approach. Let us mention just two of these difficulties. First, virtually all factor models assume linear covariation among the factored variables and an additive linear combination of variance components. Now, it may be that issue behavior in world politics does conform to these linearity assumptions, but it is equally conceivable that issues covary in response to some threshold level or according to some other nonlinear principle. Until linearity is
confirmed empirically it would seem unwise to adopt such a restrictive view of the world.

The procedure outlined by Lampert and his associates attaches a substantive label to each system on the basis of the specific issues loading highly on each of the corresponding factors. It is possible, even likely, that on occasion issues will exhibit high negative loadings on at least some factors, thus indicating the presence of inverse covariation between a factor and a particular issue. On the one hand, the analyst is obliged to acknowledge an issue with a high negative loading as a powerful contributor to the factor structure; yet, on the other hand, negative loadings reveal an absence of issue behavior. Surely it would be misleading to define a system by the type of issues it does not involve.

These problems and others that derive from the authors' factor analytic approach have serious implications for their multiple systems model, but it is not our intention to pursue this line of criticism further. Instead, we turn to the far more severe difficulties associated with their interaction network conceptualization of system structure. We do not deny either the legitimacy or the potential value of network structure interpretations of world politics. Karl Deutsch's well known studies of international integration (for example, Deutsch and Eckstein, 1961) and, more recently, Anthony Judge's (1978) innovative "networking" approach to global problem solving are but two examples of how network structure perspectives can be
used to good advantage. Our criticism, then, is of Lampert and his associates' application of the network structure approach and not of the approach in general.

Earlier we stated that the prime requisite of a multiple systems model of world politics must be its capability to unambiguously distinguish one international system from another. Systems structured as interaction networks as distinguished from one another by relative discontinuities in the intensity of interaction patterns such that interactions among different systems are distinctly weaker than interactions within systems. Because these discontinuities seldom, if ever, entail a total absence of interaction, they must be based on some threshold level set by the analyst. The implication, of course, is that system boundaries are ultimately an arbitrary matter reflecting whatever is thought to be a tolerable amount of interaction between a system and its environment. Once systems are defined as interaction structures it will be relatively easy to locate a threshold level that will yield several putatively distinct systems, but by the same token it is likely that what appear to be separate systems at one threshold level could be consolidated into a single structure at a more comprehensive level. From this perspective, the multiple systems argument is transformed into an argument concerning what threshold level is most appropriate. Furthermore, so long as at least some interactions among systems are tolerated there can be no basis for the assertion that these structures
represent distinctive systems in their own right rather than merely subsystems encompassed by some overarching, hierarchically organized system. Unless the multiple systems hypothesis is reduced to nothing more than a semantic preference for "system" rather than "subsystem," it must be possible to demonstrate not only that systems can be differentiated from one another but also that the existence of multiple systems precludes the possibility of a single comprehensive system.

There is yet another reason why the network-of-interactions representation of system structure is unsatisfactory as the basis for a multiple systems model of world politics. At the beginning of this chapter we noted that a chief concern of systemic level analysis in international relations is to delineate the effects of system structure on the behavior of units within the system. If structure is defined at the outset as merely a result built up from units' actions and interactions the researcher is effectively denied access to this line of inquiry. Moreover, we should not expect a multiple system theory ever to supersede (that is, falsify) a single system theory if the former, by virtue of its network structure approach, entails less empirical content than the latter.

It is for these reasons that we are compelled to dispense with the network-of-interactions interpretation of system structure as a viable foundation for investigating the multiple systems hypothesis. What is needed, then, is an alternative conceptualization of system
structure that avoids arbitrary distinctions between systems, discriminates between system structures and subsystem structures, and does not preclude the possibility of systemic effects on actors' behavior. In the following subsection we introduce a positional approach to system structure as just such an alternative.

**Positional Structure**

If we alter our focus from networks of interactions to the positions units occupy vis-a-vis one another, then we are adopting a positional interpretation of system structure. A positional approach to international system structure is not unfamiliar to students of world politics. Studies of system polarity, for example, normally proceed from an assumption of how actors are positioned along a system-wide distribution of power (for example Deutsch and Singer, 1964; Waltz, 1967). Similarly, Keohane and Nye (1977: 20) employ a positional interpretation when they state that "the structure of a system refers to the distribution of capabilities among similar units" (authors' emphasis). Kenneth Waltz (1975: 46) has furnished us with a useful summary of a positional perspective: "To define a [positional] structure requires ignoring how units relate with each other (how they interact) and concentrating on how they stand in relation to each other (how they are arranged or positioned). By abstracting from unit attributes and interactions, one arrives at a purely positional picture of society."
If a positional approach is to provide the necessary conceptual leverage to investigate the multiple systems hypothesis, it must be able to discriminate between different systems in a nonarbitrary manner. Once again, Kenneth Waltz (1975: 46) has proved to be singularly enlightening by devising a twofold definition of positional structure that seems well-suited to our purposes:

If one is concerned with the different expected effects of different systems, one must be able to distinguish changes of systems from changes within them. A two-part definition of structure enables one to discriminate between those types of changes. Structure is defined, first, according to the principle by which a system is ordered: in politics, that is, according to whether the realm is one of super- and subordinate units or of coordinate ones. Systems change, then, if structures vary through alteration of their ordering principle. Structure is defined, second, according to the arrangement of parts. Within a given type of order, systems change if variations occur in the arrangement of parts. Any changes in units and any changes of their behavior that leave the structure intact will then be changes within systems rather than shifts from one system to another.

If should be noted that Waltz's purpose in advancing these criteria was to facilitate detection of system change through time. Our intention, on the other hand, is to employ Waltz's criteria to discriminate between contemporaneous systems. In both cases the principle remains the same: a change of structure indicates a change of system.

Application of these criteria requires, first of all, determination of the ordering principle(s) underpinning the structure of world politics. Does the ordering principle remain constant for all international systems, or should we be alert to changes occurring at even this fundamental level? Waltz suggests that systems contain either
"super- and subordinate units" or "coordinate ones". He illustrates these two organizational forms by contrasting domestic and international systems: "Domestic systems are centralized or hierarchic. International systems are decentralized or anarchic. The ordering principles of the two systems are different in type" (Waltz, 1975: 46). It is evident that Waltz' ordering principle is derived from the degree of autonomy exhibited by a system's units. Accordingly, the anarchy attributed to the international system indicates a situation in which autonomous units maintain an equal footing regarding one another in an environment with no fixed pattern of authority.

Yet there is something bothersome about an ordering principle that stresses equality in the face of the glaring inequalities prevailing throughout the world. Speaking of the distribution of national power, Waltz (1967: 229) himself has elsewhere remarked that "because inequality is inherent in the state system...it cannot be removed." From the vocabulary of international relations—wherein such terms as "superpower" and "middlepower" are firmly entrenched—to the constitutions of international organizations—where institutions such as the World Bank, the Common Market, and even the United Nations formally entitle some of their members to privileges not available to others—one can find evidence of a global hierarchy (Wallace, 1971). Even studies of international interaction networks have uncovered structures consistent with hierarchical ordering principles (for example Bernstein and Weldon, 1968; Brams, 1969; Russett and Lamb, 1969). In short, if we relax the assumption that
hierarchy describes a vertical channeling of authority and instead use the term more inclusively to indicate a continuum of deference and reward much like a barnyard pecking order, then in many respects international systems would have to be treated as hierarchically ordered system.\textsuperscript{10}

If we concede that international systems are characteristically hierarchical in the sense described above, then we must shift our attention to the second of Waltz' two criteria in order to detect a change of system. In a multiple system world we should find substantial differences in the arrangements of parts as we move from system to system so that taken altogether the world would approximate the "multi-hierarchical" configuration envisaged by Stanley Hoffman (1968: 356-360). Of course, Hoffman would reject the multiple systems premise and imbed these discrete hierarchies in a single, comprehensive, multi-hierarchical international system. This raises a most critical question for the present analysis since surely there must be more to an international system than mere structural hierarchy.

These last few pages have dwelt exclusively upon the importance of systemic structure, but this is only part of the picture; we must not overlook the potential importance of systemic effects. J. David Singer (1969: 39) has put the matter incisively: "...any system which serves as a social and/or physical environment imposes constraints and provides incentives which must exercise an appreciable
impact on the behavior of the component sub-system entities, and therefore on the interaction patterns which these entities display."

Thus, having previously discarded patterned interactions as the structural basis of an international system in favor of positional approach, we now find it necessary to reassess the role of interactions in the system model. Note, however, that this in no way implies a retreat to an interaction network representation of system structure; quite to the contrary, interaction patterns are presumed to be the effects of a system's positional structure. Consequently we would anticipate that a change of system evidenced by a change of system structure would ordinarily be accompanied by some sort of change in the interaction patterns associated with that structure.

We have now arrived at a position where it is possible to sketch in the rough features of a research strategy for investigating the multiple systems hypothesis. In accordance with the logic of inquiry outlined in an earlier section, such an investigation should be conducted as a three-cornered affair with two competing theories or models set against observable facts. Our task, in other words, is to determine which of two competing system constructs--the single system or multiple systems model--more closely approximates its intended empirical referent.

Let us assume for the moment that our real world referent is organized in a manner that is consistant with a model positing multiple, concurrent international systems. Each separate system would
exhibit a distinctive structural arrangement of units and a characteristic pattern of unit interactions. Moreover, the structure and interaction components of these systems are not merely coincident, they are functionally related such that interactions are patterned, at least in part, by the prevailing system structure. Now, if these various structures and their corresponding interaction patterns were aggregated to fit a single system representation, as they might be by a naive analyst unaware of the underlying multiple systems configuration, it would be very nearly impossible to detect anything resembling the authentic functional relationship linking structure and interaction. Likewise, if the world were actually more accommodating to the single system construct but our analyst dissected it into putatively distinct systems, the result would be much the same: a badly distorted view of the genuine functional relationship. What emerges, then, is a research strategy that places a high premium on a prior understanding of the actual relationship that obtains between a system's positional structure and patterned interactions. It would then be a relatively straightforward matter to evaluate the congruence between this theoretically derived relationship and the relationships determined empirically under the single system and multiple systems interpretations of world politics.

Let us briefly summarize our development thus far. We have seen that any attempt to devise a truly comprehensive formulation of
a single overarching international system rapidly becomes emmeshed in enormous complexity. Most researchers cope with the complexity by dismantling the system into more manageable components. Lampert and his associates have gone a step beyond this approach by claiming that a single system construct is fundamentally untenable and that it must be replaced by a model postulating the existence of several distinct international systems each revolving around a discrete set of issues. We examined their formulation in some detail and concluded that their network-of-interactions approach to system structure was inadequate for investigating the empirical veracity of the multiple systems hypothesis. We then advanced an alternative conceptualization based on Waltz' notion of positional structure that could provide the foundation for a research strategy focusing on the presumed functional relationship between structure and patterned interactions. Put simply, this strategy says that if we know a priori what sort of interaction patterns should be produced by what sort of structural arrangements, then we can use this information as a set of expectations against which to assess the relationships derived from empirical analyses of the single system and multiple systems models. The key to this strategy rests upon our prior expectations regarding how structure affects interactions. This matter is taken up in the following section.
Many who have studied the hierarchical ordering of international systems have done so from a stratification perspective. Gustavo Lagos (1963) is widely recognized as the first to explicitly utilize the stratification framework at the international level. Other notable contributions to this literature would include Johan Galtung's (1964; 1966a; 1966b; 1966c) many works on rank theory and Rudolph Rummel's (1977a) more recent formulation of status-field theory. Substantial differences separate these various approaches, but common to all of them is the fundamental assumption that international systems, like all social systems, are composed of entities, usually but not exclusively conceived as nationa states, that are layered through an unequal distribution of socially desirable characteristics such as power, wealth or prestige. This ordering defines each member's status position or rank in the social structure which, in turn, leads to an intriguing set of hypotheses linking patterns of behavior to certain structural positions. Only a brief overview of the stratification framework will be presented in this section; a more detailed treatment is advanced in the next chapter. We will begin by considering stratification as a characteristic of a single, composite system and then outline the modifications necessary for its application to a multiple system world.
International Stratification

Generally speaking, stratification is a result of two constituent social processes termed differentiation and evaluation (Duvall, 1975; Lagos, 1963). The former is a process whereby the members of a social system recognize that they possess different attributes and behave in different ways; the latter then interprets these differences according to a commonly accepted scale of value. "Consequently," writes Lagos (1963: 9-10) "the system of international stratification would be the result of differences between the various nations appraised in terms of the prevailing values." Stratification thus entails some measure of inequality among the members of a social grouping. In large societies an unequal distribution of income, education, occupational prestige, or other social rewards gives rise to variously defined spectra of social classes that routinely occupy a central role in contemporary social analysis. Often references are made to specific classes or social strata as a way of verbally summarizing a variety of attributes characteristic to particular elements of a large population. To use the familiar language of stratification as a form of scholarly shorthand is to assume, usually with ample justification, the social rewards tend to accumulate in uneven amounts to identifiable segments of society. Yet stratification and related concepts are by no means limited in their application to large populations, even small groups of five or ten individuals may be hierarchically ordered through an unequal
distribution of leadership capabilities, formal authority, expertise, and so forth. In sum, a social system is stratified through a combined effect of differentiation and evaluation, and not as a function of its size.

How is stratification related to behavior? Numerous studies of stratification conducted at various levels of social organization have drawn attention to two fundamental behavioral propositions. Proceeding from the assumption that social systems are ordered on multiple status dimensions, the first argues that status dis-equilibrated actors—those ranking high on some dimensions and low on others—will experience certain frustrations and tensions that, under certain conditions, may lead to various forms of deviant behavior. At the international level this proposition has been extensively investigated as a possible explanation for the occurrence of international conflict and war. The second proposition is concerned with how interactions are channeled throughout the social hierarchy and claims that the amount of behavior, whether transmitted or received, tends to increase as a positive functions of the actor's overall rank position. Put more simply, interaction is said to be "rank-dependent."

In order to facilitate comparative evaluation of the single system and multiple systems perspectives we will posit a stratification theory of international systems that entails a priori expectations of systemic effects derived from extensions of the
rank-dependent interaction hypothesis. Although our inquiry follows an hypothesis testing mode and represents a test of the two models, it is still very much an exploratory venture. Therefore, to maximize both tractibility and comparability with previous studies our stratification theory is designed to encompass only a narrow range of international actors and interactions. For this effort we confine our attention to the foreign policy behaviors of national governments. Foreign policy behaviors are defined as "... those discrete official actions of the authoritative decision-makers of a nation's government, or their agents, which are intended by the decision-makers to influence the behavior of international actors external to their own policy" (Hermann, 1972: 72; see also Hermann, 1978).

Two features of this conceptualization deserve special notice. First, the myriad of international activities conducted by other than governmental actors are eliminated from consideration. This represents a significant departure from the approach taken by Lampert, Mansbach, and Falkowski since, as we noted earlier, they assumed an explicit transnational perspective by incorporating nongovernmental actors into their multiple systems formulation. Second, our restricted view of foreign policy behavior excludes phenomena such as aggregate trade flows and international organization memberships because they are not discrete actions. We should point our, however,
that specific trade agreements and decisions to join or withdraw from
an international organization would qualify as discrete actions
and, hence, are encompassed by our definition of governmental foreign
policy behavior.

As a fundamental premise of our stratification theory we will
assume that international systems populated by governments are struc-
tured according to a hierarchical arrangement that, for the single
system model, is defined by the distribution of national power. Power,
certainly one of the thorniest of concepts, often means an ability
to exert control over the actions of others or over the outcome of
events (Hart, 1976). Frequently, however, power is thought of as
the potential for control associated with the possession of power
resources. Writers in the realist tradition, for example, typically
conceive of power as a composite of various ingredients—some of the
more prominent being military capability, economic strength, indus-
trial capacity, and the availability of human and natural resources—all complementing and reinforcing one another in a gestalt-like
process that in the end reveals a government's overall power rating
(for example Organski, 1968; Morgenthau, 1973).

This composite view of national power supplies the structural
element for the single system model of intergovernmental politics.
We can expect foreign policy interactions to be sensitive to this
power structure and as a consequence to display certain patterns
that reflect its configuration. But for our purposes the signal
feature of this model is the absence of issue area differentiation as a complicating factor. "On the contrary," wrote Keohane and Nye (1975: 43) of overall structure approaches, "it predicts a strong tendency toward congruence of outcomes among issue areas. Since power, like money, is considered fungible, power resources will be shifted by major states to secure equal marginal returns in all areas. When outcomes on one issue area are markedly different from those on others, we should expect shifts to make outcomes in the divergent area more consistent with the world structure of military and economic power." Their words supply a fitting description of our single system model of world politics.

In sum, the distribution of overall national power plays a dual role in our model of a single, comprehensive intergovernmental political system. First, it defines the specific arrangement of units within the system's hierarchical structure and, second, it provides the basis for stratification of the system. This power structure will have a pervasive effect on the foreign policy behaviors taking place within the system without regard to issue area. As a result, the single system model as seen through the lens of the stratification approach predicts the emergence of behavioral patterns that should be discernible both across and within specific issue areas. We will be more explicit about just what these patterns look like when we develop the stratification theory more fully in the following chapter, but before doing that let us consider the multiple systems model.
Issue-Based Stratification

Having already outlined the gross features of a single system model of international stratification, we can now move expeditiously toward a similar characterization of the multiple systems version. As the title of this subsection indicates, we will propose a marriage of the issue-based systems idea and stratification theory. Our starting point is a simple and largely self-evident proposition that customarily precedes discussion of the status disequilibrium hypothesis: "...the status position of nations is not the same for all issue areas" (Wallace, 1973: 5). Here, though, we intend to explore its implications for rank-dependent interaction.

If, indeed, the status position of nations as determined by their power resources differs appreciably from one issue to another then there is some basis on which to argue that the arrangement of units has changed, and thus that the system itself has changed. Recall that such a change does not require an alteration of the system's basic ordering principle—in fact, in this analysis we assume that all international systems are ordered hierarchically—but merely a shift in the positions units occupy vis a vis one another. If an issue-specific structural hierarchy is to constitute the foundation of an issue-specific international system, then we would expect to find evidence of structural effects of the sort predicted by stratification theory. However, the distinguishing feature of the multiple systems approach is that for any given hierarchy, structural effects will be confined to interactions within that
hierarchy's associated issue area. Thus, from a multiple systems perspective the distribution of overall national power would appear to be an overly aggregated and artificial representation of system structure. Moreover, an overall power structure should prove to be a very poor predictor of issue-specific foreign policy interaction patterns and, contrariwise, issue-specific structures should do poorly when set against interactions aggregated across issues. Of course, if an issue-specific structure and an overall power structure were found to replicate one another's positioning of units—an eventuality that must not be discounted—then we would expect both to perform equally well irrespective of how interactions are represented.

A necessary assumption in all of this is that issue areas are relatively encapsulated behavioral domains resistant to certain kinds of linkage. For example, the occurrence of cross-issue bargaining processes, if persistent, could severely, dilute interaction patterns in specific issues and thereby mask or even erase altogether any expected effects of the system's structure. Even more troublesome would be a linkage that permitted influence or capabilities to be transferred across issue areas, since that would directly contradict our earlier premise that structural effects are circumscribed to particular issue areas. To be sure, both kinds of linkage can be found in the real world of international politics and to claim otherwise is, at best, a "simplifying assumption" and, at worst, simply erroneous. But all models, all intellectual constructs, are built
upon simplifications of reality and the dividing line between what
is erroneous and what is merely a simplification is sometimes
obscure. Ernest Nagel (1973) has done a great deal to clarify
matters by maintaining that theoretical assumptions can be unrealistic
in three different senses: they may be incomplete descriptions; they
may refer to "ideal types" or "theoretical terms" which are never
actually found in the world; or they may be utterly false or im­
probable on the weight of available evidence. It is our contention
that the "no linkage assumption" be considered unrealistic in the
second sense rather than the third, at least until there is empirical
evidence to the contrary.

One final point deserves mention before we conclude with a
brief overview of the remaining chapters. In our earlier discussion
of Lakatosian falsificationism we noted that certain "hard core"
assumptions remained untouched in the shift from a single system to
a multiple systems model of world politics. Several of these
assumptions have been advanced at various points throughout this
chapter. We have assumed, for example, that a positional approach
to system structure is appropriate to both models, that all inter­
national systems are structured hierarchically, that all international
systems are subject to stratification and its constituent processes
of differentiation and evaluation, and that stratified systems
reveal similar patterns of interaction. The "protective belt"
assumptions, on the other hand, have concerned such matters as
the criteria for determining the arrangement of units within a system and the linkage between issue areas. In short, the vital center of our system construct remains intact across both models.

Overview

The main purpose of this first chapter has been to unfold a viable strategy for a rational and systematic appraisal of the proposition that international political relations are organized into several distinct issue-based systems. What has emerged at the end is a research program contingent upon a theory of intergovernmental stratification that entails specific and empirically operational predictions of structural determined patterns of foreign policy behavior. These predictions will be used as a standard for comparatively evaluating sets of empirical realtionships observed, first, according to the conventional image of a single, overarching international system and, second, under the competing assumption that world politics more closely approximates a model of multiple, issue-based international systems.

The next chapter develops the theoretical rationale motivating the stratification hypotheses. To provide some background for this development, the first part of the chapter surveys the extant literature on international stratification. This review suggests that most previous efforts have taken concepts and principles originally applied to individuals and facilely transferred them to
collective entities at the international level without further explication or reformulation of their intended meaning. To avoid this difficulty, an explanatory logic is introduced that characterizes governmental foreign policy behavior as a problem-solving enterprise and international structure as a function of the system-wide distribution of problem-solving capabilities. The chapter concludes by drawing upon this logic to delineate ten bivariate hypotheses. Each hypothesis specifies a linkage between a single government's structural position or the positional configuration of a pair of governments, and certain dimensions of intergovernmental behavior. The hypotheses refer to three general categories of behavioral properties: level of involvement in international affairs, degree of behavioral autonomy, and sensitivity or attentiveness to various types of foreign policy problems.

Chapter III deals with issues pertaining to the operationalization of international positional structure. This chapter begins with a brief discussion of seven criteria for sound measurement. The main part of the chapter then presents twelve candidate indicators along with a preliminary evaluation of each according to the seven criteria specified earlier. Each measurement procedure focuses on some aspect of national capabilities as a means for delineating each government's relative structural position in the putative systems under consideration. The first four measures assume a single system approach by emphasizing overall capabilities in foreign
affairs undifferentiated by issue; in contrast, the remaining eight measures reflect the multiple systems approach by being more narrowly focused on capabilities relevant to a limited range of problems in the areas of international trade or the settlement of military security disputes. In the last section the empirical convergence and divergence of the twelve indicators are examined prior to the final selection of those best suited to represent the structural arrangements underlying a single, comprehensive system on the one hand, and the two issue-based systems on the other.

The fourth chapter continues the task begun in Chapter III by outlining the procedures used for the observation and measurement of governmental foreign policy behavior. These measures are acquired from an established data base containing foreign policy events for thirty-eight governments during the decade of the 1960s (Hermann, et al., 1973). The operational procedures underlying these data are examined in separate stages dealing with the identification of events from public source material, the classification or coding of identified events according to relevant behavioral properties, and the aggregation of events into monadic and dyadic patterns of behavior.

The second section of Chapter IV presents the methods used to empirically determine the fit of the stratification hypotheses under the competing models of systemic organization. After consideration of some alternative techniques for delineating bivariate association, simple regression analyses is deemed most appropriate to the
requirements of the present study. A set of procedures is then introduced for compiling two types of evidence from comparisons of separate regression analyses of each hypothesis. These comparisons are designed to yield outcomes that can be evaluated as supportive of either the single system model or the multiple systems model, or they can be judged "ambiguous" and thus supportive of neither approach. The chapter closes with a brief discussion of some technical considerations involved in applying the regression model to the stratification hypotheses.

Chapter V reports the empirical findings of the regression analyses and provides an initial assessment of the two approaches based on the comparisons described in the previous chapter. The first section presents results from the hypotheses concerning monadic patterns of foreign behavior. The second and third section deal with dyadic behavior patterns and the effects of structural proximity and structural direction, respectively. The final section summarizes the results and considers some overall patterns or trends that may be indicative of systemic organization.

Chapter VI concludes the research by exploring what, if any, implications these findings have for the single system and multiple systems approaches. In the first part of the chapter, consideration is given to the possibility that the findings are merely an artifact of the measurement or analysis procedures. The second part then presents three alternative interpretations of the results along with some initial suggestions for the further investigation of each.
NOTES

CHAPTER I

This paragraph and the three that follow draw heavily from Lakatos (1970). For an extremely lucid summary of Lakatos' position, particularly as it contrasts to Kuhn's (1970), see Ball (1976). For evidence that Lakatos' views are becoming increasingly influential among specialists in international politics see Alker (1976), Ashley (1976), and various contributions to Hoole and Zinnes (1976).

^The stratification model is briefly examined further on in this chapter. A detailed treatment is contained in Chapter II.

^As used here, hierarchy does not refer to the vertical authority schemes that are common to most formal organizations. Simon (1969: 87) employs the term more generally to mean "...a system that is composed of interrelated subsystems, each of the latter being, in turn, hierarchic in structure until we reach some lowest level of elementary subsystem.

^It would appear that the more recent proponents of the multiple systems approach badly misread Rosenau's purpose throughout this argument. It seems abundantly clear that Rosenau's intention was (1) to set forth the issue (or vertical) system idea, (2) to raise certain criticisms against this conceptualization, (3) to counter these criticisms in defense of issue systems, and (4) to conclude that the issue system concept is both workable and potentially valuable. After quoting a passage from Rosenau outlining the distinction between vertical and horizontal systems, Lampert, Falkowski, and Mansbach (1978: 147) state that "Rosenau himself recognized such a distinction as highly arbitrary and unsatisfactory in terms of boundary-maintenance." Immediately following this statement, and presumably in an attempt to justify it, they quote the following passage:

The argument that vertical systems are vulnerable because of the stability requirements of horizontal systems may not be as valid as it appears at first glance. Precisely opposite considerations may prevail in some instances. To maintain stability, horizontal systems may have to insultate certain
issue-areas and prevent bargaining across their boundaries (Rosenau, 1966: 80).

What the authors do not seem to realize is that this passage is intended to bolster Rosenau's argument, not undermine it. Rosenau is simply saying boundaries may be firmly insulated from cross-issue bargaining processes.

5 Not to be outdone by Hoffman, Keohane and Nye (1977: 60) develop their own metaphor to graphically illustrate the key aspect of issue structuralism:
If the overall power structure in world politics determined patterns of regime change, we would not need to have introduced such a complex set of models. World politics would be like a single great lake: often turbulent, but with a uniform level. Changes in the amount of water flowing into one part of the lake quickly have effects on the whole body of water. We assume, however, that world politics is highly differentiated rather than homogeneous. The appropriate image for our analysis is therefore not a single lake, but a river divided by sets of dikes, dams, and locks, which separate and connect various levels and "lakes".

6 However, they do provide an instructive example: ". . . the escorting of Soviet guards to the Soviet war memorial in West Berlin by American troops becomes part of a stack [issue] labeled 'Postwar Treaties' whereas the dispatch of U.S. armored vehicles along the Berline autobahn is included in an issue termed 'Access to Berlin'" (Lampert and Mansbach, 1976: 12). Still, it is unclear why either more specific (e.g. "Surface Access to Berlin") or more general (e.g., "Berlin") designations were deemed inappropriate.

7 Virtually the same behavioral criterion underlies Rosenau's (1966: 81) conception of issue areas: "If a cluster of values induces support building on the part of the affected actors but if their behavior is not distinctive from that induced by another cluster of values, then the issue-area is considered to encompass both clusters, and both are also regarded as being processed by the same vertical systems."
The following paragraph may aid the reader in comprehending the factor analytic procedure advocated by the authors:
By taking all possible dyads as the rows and the distribution of numbers of events [i.e., behaviors] across all of the issue classifications as the columns, we can construct a matrix that can be factor analyzed. The factors which emerge from an R factor analysis (in which the columns, in this instance, the issues, are factored) describe the interrelationships which exist among the issues in the data. They do so on the basis of the patterns characteristic of behavior along each dyad. (Lampert and Mansbach, 1976: 12).

For a more closely argued presentation of this objection see Waltz (1975: 43-45).

Note that this use of "hierarchy" differs considerably from Simon's usage. See Note 3.
CHAPTER II

INTERNATIONAL STRATIFICATION

One reason for the early enthusiasm toward systems-oriented inquiry was a widely held belief that carefully formulated systems constructs could be profitably applied beyond their intended empirical referents. Some general systems theorists have elevated this belief to the status of a basic ontological axiom by insisting that all natural (empirical) systems are organized in much the same way and behave according to a single set of general laws. A less extreme variant would claim only that natural systems may share certain similarities to one another that are made apparent by application of a common system construct as a metaphor or heuristic device. Whatever the rationale, the operational consequences involve modeling an empirical domain using concepts borrowed or adapted from a system construct originating elsewhere.

Students of world politics were quick to exploit this logic and in the early 1960's began to investigate international systems guided by concepts and methods previously introduced by analysts of less inclusive social and political systems. For example, the structural-functional framework pioneered by Talcott Parsons was employed by Modelski (1961), Riggs (1961), and subsequently even by Parsons (1961) himself to examine the extent to which functional requisites are maintained at the international level. Others have pointed out the
resemblence between international systems and various lower order political systems, ranging from the inchoate forms characteristic of primitive societies (Alger, 1966; Masters, 1964) to the competitive systems found in modern democracies (Russett, 1963b). Still others who have followed this pattern have assumed that all social systems, including international ones, may be conceived as stratification systems.

What is the basis of international stratification? Are systems composed of collectivities stratified in the same manner as systems of individuals? How does stratification influence behavior? What is the explanatory logic that accounts for this influence? These are questions to be addressed in the present chapter. Our primary aim is to advance a theory concerning the operation of hierarchically ordered intergovernmental political systems. In order to locate this effort in its broader intellectual context we will begin by surveying the extent literature on international stratification, particularly that portion of the literature dealing with the behavioral effects of stratification. We then discuss the underlying logic of the stratification approach. As a theory deriving its inspiration from studies at the individual level yet applied here to national governments, our stratification theory is in particular need of a well articulated explanatory logic since the transference of assumptions from one level to another is not automatic. We conclude the chapter by drawing upon this explanatory logic to delineate hypotheses linking hierarchical positional structures to a diverse set of foreign policy behavior properties.
The Study of International Stratification

Stratification is normally thought of as a condition present whenever there is appreciable inequality of desired rewards received by the members of a group or society. But stratification can also be conceived as a process describing the interrelationship between a system's prevailing inequality structures and its patterns of behavior. We will adopt the latter usage because it more accurately reflects the concerns of those who have applied the stratification framework to international relations and because it helps illuminate an important division in the international stratification literature. By process we mean a causal relationship such that "...a change in one or more input variables produces a regular and predictable change in an output variable" (Brunner and Liepelt, 1972; 539, f.n. 3).¹ The literature encompasses two distinct but complementary interpretations of the stratification process that differ on the basis of whether inequality structures or interaction patterns are assigned causal priority. The view expressed briefly in the preceding chapter presumed that inequality was causally antecedent to patterns of behavior; the other interpretation reverses this ordering by assuming that hierarchical social structures are the product of certain types of interactions. These two perspectives—which we shall label the "interactionist" approach and the "structuralist" approach, respectively—are illustrated in Figure 1.²
FIGURE 1

The Stratification Process
Despite the impression conveyed by the diagram in Figure 1, we should resist the temptation to infer that the interactionist and structuralist approaches focus on opposite sides of the same nonrecursive social process. Apart from the directionality of causal influence, the two approaches incorporate differing interpretations of inequality and emphasize different aspects of social interaction.

For most structuralists inequality is based on an international division of labor that engenders functional differentiation of social roles and interaction is conceived as an exchange relationship that delivers certain benefits to the interacting parties. Because the benefits of interaction generally accrue in uneven amounts the existing inequality structure is reinforced or even exacerbated by the prevailing patterns of interaction. On the other hand, the interactionist's view of inequality emphasizes the established distribution of socially esteemed characteristics and does not depend on concomitant differentiation by social tasks. Moreover they tend to regard interaction more as an influence attempt than an exchange relationship.

Admittedly, these distinctions are roughly drawn and are apt to obscure areas where the two approaches overlap and interrelate; nonetheless, they are sufficiently distinct to warrant separate treatment. We will briefly consider some of the central ideas expressed in the structuralist literature and then move quickly to a more extended review of interactionist approaches.
The Structuralist Approach

Let us begin our examination of structuralist literature by advancing two preliminary observations. In the first place, structuralists tend to focus their attention of the forms and consequences of global economic relations, particularly those associated with capitalist modes of production and distribution. The second observation is a corollary of the first: in general, structuralists are only marginally concerned with international relations in the traditional sense of relations among separately constituted nation-states. To be sure, nation-states are still very much a part of the global scene, but the advent of capitalism on virtually a worldwide scale has generated a need for alternative units of analysis to accommodate functional differentiation of economic and political tasks that transcend conventional political-legal-territorial boundaries (Little, 1978; Targ, 1976).

Argentine economist Raul Prebisch (1950) was among the first to call attention to the now familiar global division of center and periphery. Despite the unfortunate imagery of a planar surface, the terms are intended to connote a vertical dimension of control and accumulation. Galtung (1971) has expanded this dichotomy into a fourfold categorization by claiming that national societies, whether they are located in the worldwide Center or Periphery, are themselves internally partitioned into central and peripheral elements. The center-periphery distinction has been elaborated in a different way by Helge Hveem (1973). The "primary level", which is at the bottom
of his four-tiered scheme, engages in production of agricultural commodities and extraction of raw resources; the "secondary level" processes and distributes raw materials and includes some low level management functions; the "tertiary level" coordinates and manages activities at the primary and secondary levels; and, finally, the "quaternary level" is responsible for the most fundamental decisions concerning operation of the overall system. But regardless of how one chooses to differentiate role functions in the global economy, the critical point for the structuralists is that this differentiation also defines levels in the structural hierarchy. Positions in this hierarchy are arranged in order of increasing control over factors of production and distribution and, correspondingly, by increasing concentration of accumulated wealth.

From this basic framework the structuralists have argued that interaction is essentially an exchange of value. When exchange occurs across levels benefits tend to be distributed upward in disproportionate amounts with the result being further accumulation at the top and consolidation of the entire structure. This principle of unequal exchange has been a recurrent theme in structuralist literature used to explain the persistent gap between developed countries and the Third World (Galtung, 1971; Emmanuel, 1972; Amin, 1973). The dependency (dependencia) literature has developed this line of argument with particular reference to United States and Latin America (Bath and James, 1976) and Galtung (1973) has applied it to the European
Community's relations with the less developed countries of Africa and Asia. In addition, Galtung (1971) has extended the argument beyond the economic sphere by asserting that differentiation and accumulation of value are to be found in political, military, communication, and cultural areas, as well.

Structuralist authors since the time of Hobson and Lenin have employed this general frame of reference to account for great power imperialism and penetration into less developed areas. There have been many variations of this argument but fundamental to all of them is the premise that capitalist centers require some degree of control over peripheral areas to insure continuance of the structural order. Hveem (1973) has proposed a taxonomy of control mechanisms ranging from "enforcement" through the use or threat of direct physical force to the more subtle and indirect "demonstration effect" which relies on the "trickle down process" for dissemination of attitudes, motivations, and ideas congenial to the existing hierarchy. Between these two extremes is "assimilation", a form of control which operates like Galtung's (1971) bridgehead theory by instilling in elites of the controlled unit interests and values parallel to those in the controlling unit. In addition to being instruments of imperialism, these various types of control and the dominance relationships that underlie them have been thought to constitute an important structural source of the two major contemporary international conflicts between North and South and between East and West (Gantzel, 1973). In short, there is a strain of structuralist literature that partially overlaps
the interactionist approach by postulating a structural explanation for certain forms of international behavior such as imperialism and conflict. Still, the structuralist classification of these studies seems appropriate because even though they presume that inequality structures are able to exert some influence on behavior, they generally do so within a framework of reciprocal causality in which the effect of such behavior is thought to be maintenance of the structure itself and the further integration of its strata.

The structuralist view of inequality encompasses two complementary and self-reinforcing processes. These are the transference of value from the bottom of the hierarchy to the top and the exertion of authority or control from the top down. Together these two processes entail a twofold conception of inequality involving (1) relational differences in the existing distribution of valued rewards and (2) differential access to such rewards. This notion of inequality encourages structuralists to embrace a normative position advocating a more equitable and just social order. Indeed, a strong normative orientation is expressed by most structuralist writers who would concur with Galtung (1971: 83) that "... the goal should not be an 'objective' social science freed from all such value premises, but a more honest social science where the value premises are made explicit".

In sum, the world envisaged in structuralist literature is marked by clear functional divisions usually but not exclusively defined in economic terms. Tasks are not merely differentiated, they
are also arranged in a structural hierarchy of accumulation and control benefitting the levels nearest the top through asymmetrical exchange relationships often referred to as exploitation. As interaction proceeds across levels its effect is to preserve the structural configuration by diverting value upward and by reinforcing and consolidating the differential characteristics of individual strata. Structuralists are virtually unanimous on this point, although some maintain that the effect is largely unintended whereas others consider it fully purposeful if not conspiratorial. Another area of wide agreement is found in the structuralists' firm disapprobation of the vast inequality and structural disadvantage so prevalent in the current global order.

We should reiterate that the stratification theory advanced further on in this chapter is somewhat removed from the structuralist approach capsulated in this short and highly selective overview. The more immediate research antecedents of our formulation are drawn from the interactionist perspective. We will consider some of these antecedents in the following subsection.

The Interactionist Approach

Earlier we characterized the interactionist approach as being primarily concerned with effects of the inequality structure on the system's interaction patterns. Recall from the previous chapter that such effects may be subsumed under two behavioral propositions involving rank disequilibrium\(^3\) and rank-dependence. Although some
writers have developed theoretical formulations incorporating both propositions (for example Galtung, 1966a; 1966b; Rummel, 1977a) our plan in this subsection is to treat each proposition independently. We begin our discussion of interactionist literature by considering studies of international rank disequilibrium.

Social psychologists have long theorized that aggressive behavior in individuals is an outward manifestation of an internal psychological state of frustration. Some have speculated that one source of frustration may be disequilibrated rankings in a social structure comprised of multiple status dimensions. A status dimension is a criterion such as an inherent trait or an acquired ability on which the members of a social system evaluate one another. "In practice," writes Zetterberg (1966: 130), "ranks become convenient bundles of evaluations of their occupant". Disequilibrium, then, is the unsettling result of enjoying a positive evaluation (high rank) on some dimensions while at the same time experiencing a decidedly lower evaluation (low rank) on others. Johan Galtung (1964) was the first to recognize that this same reasoning could provide an explanation of aggression and conflict at the international level. It should be noted, however, that Galtung's pathbreaking study was intended as more than an explanation of international conflict; its purpose was to elucidate a generalized structural theory to account for aggressive behavior occurring at different levels of social organization.

Beginning with an unorthodox definition of aggression as "drives toward change, even against the will of others", Galtung (1964: 95)
argued that both the motivation and the capability to engage in aggressive behavior would be concentrated in disequitibrated social positions. He reasoned that individuals and nations in disequilibrium would experience unmet expectations and feelings of deprivation and injustice regarding their low status positions, while at the same time their high status positions would entail the necessary resources to engage in aggressive acts. Galtung cautions that disequilibrium is neither a necessary nor a sufficient condition for aggression. He states only that aggression is a probable consequence of high levels of disequilibrium, particularly in poorly integrated social systems exhibiting few means for legitimately rectifying discrepent status dimensions.

Galtung enriched the theory by drawing upon the sociologists' traditional distinction between ascribed and achieved status dimensions. In brief, an actor has a large measure of control over its ranking on an achieved dimension whereas an ascribed status position is fixed and cannot be improved (or diminished) through any action taken by the actor. This distinction enabled Galtung (1964; 1966a; 1966b) to hypothesize that where ascribed ranks are higher than achieved ranks aggression is more likely to be directed inward and, conversely, where the disequilibrium is reversed aggression is more likely to be directed outward, that is, toward other members of the social system. Unfortunately, the predicted consequences of these different disequilibria have not been subjected to empirical research at the international level. In fact, the only studies to have come close to
investigating the first sort of disequilibrium have examined the incidence of internal conflict as a function of the discrepancy between achieved levels of development and internationally induced aspirations for development (Schwartzman, 1972; Mora y Arujo, 1972). Most researchers concerned with the effects of rank disequilibrium have focused on international conflict as a potential consequence of either type of disequilibrium.

Empirical studies of Galtung's structural theory have investigated national war involvement as a function of discrepancies of military or economic capabilities and international prestige. Prestige is usually conceived as a form of reputational status indicated by the number of diplomatic missions accredited to a nation's capital by other states in the international system (Singer and Small, 1966). Maurice East (1972) discovered that between 1948 and 1964 the level of disequilibrium between military power potential and prestige was consistently related to the amount of interstate conflict experienced throughout the system as a whole. A similar study by Michael Wallace (1973) over a much longer time span produced comparable results. Both authors found that the predicted relationship tended to increase in strength when time lags were introduced, although East explored only one and two year lags whereas Wallace found the most supportive results using a time lag of fifteen years. We must caution, however, that because both authors chose to frame their investigations at the level of the international system their findings suggest only that when some nations experience disequilibrium some nations will engage
in conflictful behavior. To explain this linkage East and Wallace introduced a number of intervening factors such as role expectations, uncertainty and misperception, alliance configurations, and so forth, as possible elements in the causal sequence from disequilibrium to war involvement. A more recent theoretical effort by John Romanga (1978) proposed a similar expansion of the disequilibrium hypothesis by provisionally mapping the complex relationships involving policymakers' perceptions and predispositions likely to mediate the behavioral effects of status discrepancies.

For rank disequilibrium theorists the system's structure is most interesting at those points where discrepancy between ranks is highest. One reason that rank discrepancy is accorded so much attention is that it represents a significant departure from the normal situation of rank concordance, that is, a situation in which rankings generally coincide with one another. East (1970: 115), for example, concluded that "less than one out of every four states in the international system from 1946-64 can be classified as being status discrepant". If disequilibrium, itself an aberration, is thought only to account for a form of deviant behavior, then we must look beyond disequilibrium to uncover a structural source of more conventional types of behavior. In the interactionist literature this means looking to the rank-dependent interaction hypothesis and, once again, to the work of Johan Galtung.
Studies of small group dynamics have repeatedly demonstrated that high status members—those ranking highest in power or leadership skills—are more often a party to group interactions than are low status members and, furthermore, that behaviors directed to high status members frequently tend to be more positive in tone and content than behaviors directed elsewhere in the hierarchy (Shaw, 1971). It was Galtung’s (1966b; 1968) contention that essentially the same patterns of interaction could be found in all hierarchically structured social systems—including international ones. He postulated that the higher the level of concordance among the system’s rank dimensions, the higher will be the probability that generalized rank-roles will emerge and become internalized by the role occupants (Galtung, 1966a). When this occurs interactions become dependent on rank (or rank-roles) and the system is designated a feudal system. What patterns of interaction are associated with disparate rank-roles? Galtung (1966a; 1968) hypothesized that the higher a unit’s position in the status hierarchy, the more behavior it would initiate and receive. For the dyad this means that interaction increases as the combined rank of the pair increases. Other hypotheses cast at the dyadic level state that more interaction is expected from high to low ranks than vice versa and that the closer two units are ranked the greater the amount of expected interaction between them.

The underlying logic that explains these patterns will be examined in some depth further on in this chapter. At this juncture it will suffice merely to note that the primary motivating principle in Galtung’s theory is the desire to attain higher ranking. "Thus,
one should only speak about rank-dimensions to the extent that there
is consensus in the system for which it is defined that high rank is
to be pursued and low rank to be avoided" (Galtung, 1966a: 122).
Not surprisingly, Galtung followed the structuralist assumption that
interaction provides a way to improve one's ranking and, for those
already in the topmost positions, to reinforce the existing order.
Nils Petter Gleditsch (1970) has devised a slightly different theory
of rank-dependent interaction. Gleditsch was heavily influenced by
Galtung's work and his behavioral hypotheses closely resemble Galtung's;
nonetheless, the two formulations differ in one important respect:
Gleditsch postulated that interaction was itself a rank dimension.
This axiom, coupled with the conventional proposition that units seek
balanced (concordant) rank profiles, allows Gleditsch to argue that
the volume of interaction increases with rank because units attempt
to balance their interaction rank with their other rank positions.7
In other words, Gleditsch arrived at precisely the same behavioral
prediction as Galtung even though he began with different initial
premises.

The empirical evidence accumulated from the research literature
on rank-dependent interaction has generally been supportive of the
theory. Most researchers have analysed rank and interaction exclu-
sively in terms of nation-states, either within a regional context
or across the entire international arena. Two partial exceptions
to this pattern are Vayrynen's (1970) study of stratification among
international governmental organizations and a piece by Alger and
Hoovler (1975) examining both international governmental and nongovernmental organizations for evidence of feudality. In a study of twenty Latin American nations first published in 1966, Galtung and his associates (Galtung, Mora y Arujo, and Schwartzman, 1972) found that surface transportation links, commercial airline connections, and the volume of trade all covaried with rank position as expected even when controlling for the powerful influence of geographical proximity. As a measure of rank the authors summed each nation's rank positions on ten indicators of size and development. Rank concordance obtained within the size and development dimensions but not across them; nevertheless, the authors retained the overall rank index because it correlated highly (.93) with a subjective index compiled from questionnaire responses by Latin American college students (see also Schwartzman, 1966; Schwartzman and Mora y Arujo, 1966). In a similar study of inter-Latin American trade, Per Olav Reinton (1967) discovered that the larger the rank difference between trading partners, the greater the probability that the lower status nation would be dependent upon trade with the higher status nation but not vice versa.

Other studies have reported comparable findings for a wide variety of interactions outside of the Latin American regional setting. For example, national representation in international nongovernmental organizations and international commercial airline flights seem to follow the rank pattern, whether across East-West cold war boundaries
or over the entire world (Arosalo, 1970; Galtung, 1966c; Gleditsch, 1967; 1970; Smoker, 1966). The volume of tourist flows are the one major exception to the rank-dependent pattern, and this despite the findings concerning airline connections (Galtung, 1966c).  

Thus far we have considered only examples of nongovernmental interactions. Now let us examine how governments behave with regard to rank. East (1970) discovered that shared memberships in international governmental organizations were most frequent between pairs of low status nations and next most frequent when one of the pair was low ranking. A similar analysis by Gleditsch (1970) suggests that this contradictory finding is explained in part by the confounding influence of geographical proximity. In addition East has pointed out that low ranking nations may have a greater incentive to participate in IGOs since membership offers a relatively low cost form of interaction capable of enhancing the legitimacy of low status nations in the system as a whole as well as improving their access to higher status nations. More direct types of diplomatic activity have also been scrutinized for evidence of rank-dependence. Again controlling for geographical distance, Gleditsch (1970) found that for 8000 nation-dyads combined rank scores were moderately correlated (.46) with the number of diplomats exchanged. Similarly, Galtung (1966c) discovered that for NATO and Warsaw Pact nations a broad range of official governmental interactions—including formal diplomatic relations, state visits, trade and cultural agreements, and exchange of diplomatic notes—were generally more frequent between
higher ranking nations, both within and across the two alliance
systems, than between lower ranking nations. He also detected rank-
dependent patterns in two types of conflictual interactions—restrictions imposed on diplomats' activities and outright expulsions of
diplomats.¹⁰ A study by Peter Wallensteen (1973) has shown that war
involvement, the most extreme form of conflictual interaction, has
on average been more than ten times more frequent for a handful of
top ranking states than for all others in the period from 1920 to
1968.¹¹ In addition, according to Wallensteen, most war relations—
that is, dyads of opposing nations—occur neither at the top nor at
the bottom of the hierarchy but rather between structurally unequal
pairs.

A common thread running through the research literature mentioned
thus far is the central role given to the theoretical ideas articu-
lated by Johan Galtung (1966a; 1966b). Other authors, however, have
treated Galtung's innovative contributions more as a point of depar-
ture for their own extensions and refinements of rank theory. Cer-
tainly the most conspicuous example is Rümmel's (1977a) status-field
theory, but before we discuss Rümmel's formulation and the empirical
research it has precipitated let us briefly consider another line of
development pursued by Heide Birgit Dechmann (1972) as part of a
cooperative venture between European and Latin American social
scientists under the intellectual leadership of Peter Heintz (1972a).

Dechmann's study was motivated by Heintz' (1972b; 1972c) theo-
retical work on stratification of large scale social systems. The
main thrust of Heintz' theory is aimed at the implications of
disequilibrated national statuses for economic development policies and the incidence of domestic conflict. Building upon ideas expressed by both Heintz and Galtung, it was Dechmann's intention to examine governmental foreign policy interactions with regard to national status position. Unlike most other authors, however, Dechmann was interested in the presence of certain behavioral properties rather than merely the volume of interactions. Assuming a three-tiered rank system, he hypothesized that governments would engage in goal convergent interaction within their own stratum and goal divergent interaction across strata. Furthermore, he anticipated that the higher a government's stratum the greater the probability of goal divergent interaction across strata and the lower the stratum the greater would be the share of a government's goal divergent interaction within its own stratum. To account for these patterns Dechmann reasoned that high rank position would facilitate adoption of similar goals and norms of behavior and, hence, goal convergent interaction. The hypotheses were supported reasonably well when tested empirically with foreign policy event data gathered for over one hundred and fifty national governments in the mid-1960's. To explore these patterns further Dechmann separately factor analyzed the interaction data for each stratum. He found clear divisions between conflictual and cooperative behavior dimensions for the upper and middle strata but not for the lower stratum suggesting that lower ranking governments tended to be isolated from the central norms and values of the system.
Possibly the most rigorous elaboration of the interactionist approach to international stratification has developed out of Rudolph Rummel's (1977a) work on social field theory. The fundamental idea underlying social field theory is relatively simple: the foreign behavior from one nation to another is, in part, a function of social forces emanating from distances separating the nations in a multi-dimensional attribute space. Considering status to be a type of attribute, Rummel argued that questions of rank distance and disequilibrium could be effectively investigated within the larger context of field theory. In Rummel's (1977a: 201) words, "status theory will be a special case of field theory". The hybrid status-field theory was seen as a vehicle for eliciting more specific and substantively richer a priori predictions from field theory's mathematical structure. According to the theory, economic development (later designated wealth) and power correspond to achieved and ascribed statuses, respectively, and status-dependent foreign behavior is depicted as a linear function of the combined positional configuration of these two status dimensions for both the acting nation and its recipient. Note the qualification: status-dependent foreign behavior. Nowhere does Rummel claim that all foreign behavior is influenced by the dynamics associated with status and related concepts. The status-dependent portion of a nation's external behavior divides into cooperative and conflictful dimensions but beyond this rather vague characterization there is no a priori specification of the kind of conflict or cooperation susceptible to the effects of status.
Rummel and his associates (1977a) have conducted numerous empirical analyses of status-field theory following the conceptual and methodological pattern established for earlier investigations of field theory. Multifarious national attributes, including but not limited to those associated with the two status dimensions, are first factored to produce a moderate number of attribute space dimensions along which nations are scored and inter-nation distances are computed. Behavior space dimensions are derived from a similar procedure that begins with measures of dyadic foreign behavior. In the final step separate canonical analyses for each nation, as required by field theory's Model II, are performed to determine the linearly dependent linkages between inter-nation distances in attribute space and patterns of dyadic behavior in behavior space. The "status" portion of status-field theory concerns any of the resulting canonical variates encompassing the status dimensions of power and wealth. Results are considered supportive of the theory if significant canonical equations are found which at least approximate one or more of the main status-field theorems. On balance, the many empirical tests performed by Rummel and his associates are clearly supportive of only the "Economically Developed Status Behavior Theorem" which asserts that "the status-dependent cooperation and conflict behavior of economically developed nations to others is a function of their power incongruence ..." (Rummel, 1977a: 243).

The empirical evidence on status-field theory is not limited to the investigations carried out by Rummel and his associates. A group of researchers under the direction of Jack Vincent (1973)
conducted somewhat different and in some ways more severe tests of the three status-field theorems that treat cooperation and conflict independently. They employed two composite behavior indices—one for cooperation and one for conflict—as dependent variables in a series of multiple regression analyses to determine directly how well the theorems performed when confronted with actual data. Their most encouraging findings indicated only modest support for the theorem stating that the higher the joint rank of two nations the more the cooperative behavior between them (Cooperation Theorem). The status-field analyses conducted by the Vincent group as well as those by Rummel and his associates all share in the unstated premise that the fundamental behaving unit in international relations is the nation, not just its central government. Thus, the behavioral indicators examined in these studies included activities such as foreign trade and participation in international nongovernmental organizations alongside exclusively governmental behaviors like diplomatic representation and United Nations voting. How does status-field theory fare for governmental behavior alone?

Tomlin and Bhulman (1977) recently investigated status-field hypotheses with respect to the official foreign policy actions of Black African national governments. According to Rummel's original formulation, the generally low level of economic development in Black Africa suggests that analysis should concentrate on the theory's two "economically underdeveloped" theorems (see note 13). However, Tomlin and Bhulman introduced certain modifications that removed this
restriction by focusing on relative status position—that is, the actor's position relative to its recipient—rather than absolute position on economic development status. In addition, they performed separate analyses on groups of nation-dyads having similar relative status configurations. For these reasons the results obtained by Tomlin and Bhulman are not directly comparable to those from earlier studies. They found that distance on the power dimension tended to be more important determinant of foreign policy behavior although they were able to correctly predict the direction of this relationship only sixty percent of the time, hardly an improvement over chance alone. In another study Tomlin and several associates (1978) examined Canada's governmental behavior toward six highly industrialized Western nations from a status-field perspective. Once again power distance (or its converse, power parity) proved to be the more salient status dimension in accounting for behavior. Moreover, despite some minor differences of both a conceptual and methodological nature between their approach and the original theory, their findings appear entirely consistent with the relationship posited by the Economically Developed Status Behavior Theorem. This particular study is interesting for another reason as well. The authors discovered that Canadian behavior to the United States seemed markedly different from other Canadian behavior: "... Canadian foreign behavior is clearly dependent on status distance; but, it is the relatively greater asymmetry between Canada and the United States that distinguishes Canadian behavior toward its North American neighbor from
that which is directed to the other industrialized countries" (Tomlin, et. al., 1978: 7). As a result they went beneath the superficial structure of status-field theory to devise a special theory of foreign policy behavior in asymmetrical dyads. In a sense their effort to expand status-field theory parallels Romanga's (1978) elaboration of the status disequilibrium hypothesis.

Taken as a whole, the interactionist literature reviewed in the preceding pages appears as the confluence of three more or less distinct streams of theory and research all flowing in a common direction in the sense that each posits a causal path from structural inequality to patterned behavior. To be sure, the various authors we have examined here have suggested a number of different causal mechanisms and mediating factors to account for the causally potent linkage between inequality and behavior. But whether the argument has been framed in terms of the tensions and frustrations associated with status disequilibrium, the emergence of generalized rank-roles, the propagation of "social forces", or some other device, the primary focus of these studies has remained on structurally induced interaction. Another point at which interactionist theories converge is indicated by their emphasis on the nation-state as the central analytic unit. The structural hierarchy is conceived as a hierarchy of nation-states and the ensuing behavior is in one way or another attributed to national actors. Interactionist writers view the international hierarchy as a positional structure defined by the established distribution of valued rewards in the possession of these
national actors. This hierarchy is not dependent on the functional differentiation of social tasks—indeed, at least one author (Wallace, 1973) has explicitly rejected the idea that there is an international division of labor. The normative concerns expressed in the interactionist literature tend to be focused on the conflict potential of behavioral outcomes rather than the inequality of the structure. This normative posture most clearly motivates research on rank disequilibrium but it is also behind many of the studies mapping rank-dependent interactions (for example, Galtung, 1966c; Smoker, 1966).

It is time we consider this literature in the immediate context of our present research objectives. To reiterate, our plan is to derive a series of theoretical expectations regarding the linkage between system structure and interaction patterns and then to utilize these expectations to comparatively evaluate the single system and multiple systems models of world politics. Does the interactionist approach to international stratification furnish us with a sufficiently precise picture of international systems to permit implementation of this research strategy? The empirical findings gleaned from the research literature are mixed and inconclusive on this question. It may be that the stratification theories examined here are simply off the mark and incongruous with the real world, in which case we would do well to look elsewhere for a theoretical representation of international systems. However, there may be other reasons that would account for the varying results reported in the empirical literature. Let us consider two in particular.
We have characterized the interactionist approach as that area of stratification scholarship centrally concerned with structural influences on system-wide patterns of behavior. But, what constitutes behavior? Who or what is designated the behaving unit? Answers to these questions would seem to be indispensable for even the most superficial inquiry into the stratification process, yet the literature is ambiguous and at times almost capricious on these matters. Is it, for example, reasonable to interpret the presence of regularly scheduled airline flights as indicative of interaction? What about formal diplomatic representation or shared membership in an international organization? These are only channels through which interactions can pass, they do not in themselves tell us anything about discrete actions like issuance of a diplomatic note or casting a vote in an international assembly. Compounding this ambiguity is the fact that interaction is frequently attributed to the nation-state as though it were a unitary actor. Even Rummel, whose work is perhaps the most rigorous of all interactionist theorists, is guilty of this error. These difficulties, however, are only symptomatic of a larger problem: the lack of a well articulated explanatory logic. Too often stratification at the international level is discussed as though it were simply a larger version of the stratification process found in small groups or societies of individuals. Clearly this practice is of questionable validity since, as several authors have pointedly observed, the sociological and psychological principles that purport to explain behavior of individuals are not directly applicable to collectives much less to
such abstract entities as the state (see East, 1972; Galtung, 1968; Wallace, 1973). What is needed, then, is an explanatory logic encompassing specification of appropriate actors and behaviors involved in the stratification process operating at the international level.

A second and necessarily more speculative factor that might account for the mixed showing of interactionist research is suggested by the very question that motivates our present research effort. In general, stratification analysts have subscribed to the conventional image of a single, composite international system. Although there have been studies of stratification circumscribed to geographically or politically defined regions, invariably these presume an overall structural hierarchy that remains intact across separate issues. Authors that posit multiple rank dimensions normally consolidate these ostensibly distinct rankings through some summative procedure such as the simple additive index devised by Galtung (1966a) or the cannonical technique employed by Rummel (1977a) to test his status-field theory. The latter, although not a summative procedure per se, provides an inductively derived weighted combination of dimensions based upon their joint ability to statistically account for variance in a set of dependent behavior variables. If this single system view is incorrect—that is, if the multiple systems model more closely approximates the configuration of international activities—then on balance we would expect to find a rather unsupportive record of empirical research. Actually, the overall record appears quite
supportive in some areas, less so in others; however, we must keep in mind that much of this research is not comparable because of the wide-ranging assumptions regarding the nature of international behavior and the behaving unit (see above).

In sum, the interactionist approach to international stratification would seem to provide a valuable starting point toward a reasonably precise specification of the linkage between system structure and interaction patterns. However, it is only a starting point. In further developing this perspective it is critical that we avoid the facile and clearly erroneous assumption that status dependent behavior at the international level may be understood solely in terms of the sociological and psychological principles that explain behavior of individuals. This will require articulation of an explanatory logic that entails explicit designation of the behaving unit and, hence, the behavior being theorized about. Moreover, it can be argued that the mixed results reported in the research literature actually favor adoption of the interactionist approach for our present purposes. According to the research strategy outlined in the preceding chapter, the case for investigating the multiple systems hypothesis from an interactionist perspective would be less compelling had the literature exhibited consistently strong empirical support for interactionist propositions. Because interactionist research has generally presumed a single system, such positive results would cast doubts on the thesis that a multiple systems representation more closely approximates international reality than does its single system counterpart.
In this section we have touched upon a broad range of stratification literature, beginning with a brief synopsis of the structuralist approach and ending with a somewhat more extensive survey of interactionist writings. From this review we concluded that the behavioral propositions of stratification theory are adequate for our present needs only to the extent that they are supported by an underlying logic appropriate to international level phenomena. The next section addresses this theme more directly by presenting a closer examination of some putative explanations advanced in the interactionist literature.

The Explanatory Logic of Stratification

In general, we use the term "explanatory logic" to denote a chain of reasoning that stands behind and gives force to an hypothesized relationship linking two (or more) variables. More simply, an explanatory logic is what explains such a relationship. What should we expect of an explanation? "All explanations", according to Robert Brown (1963: 41), "are attempts to explain away impediments of some kind. They are efforts to deprive puzzles, mysteries, and blockages of their force, and hence, existence". Likewise, Abraham Kaplan (1964: 330) has observed that "the etymology of the word 'explain' is to take out the folds, to make something level or even. An explanation makes something intelligible or comprehensible". At present our particular concern is with an explanatory logic that illuminates the connection between hierarchical positional structures and patterns of international behavior.
As in the previous section, our discussion again conveniently divides into two parts. The first extends our review of interactionist literature to the explanatory mechanism posited by various authors. The second then elucidates an explanatory logic that will be used to generate specific behavioral hypotheses to guide our investigation in subsequent chapters.

A Critique

When we speak of a causal connection between positional structure and interaction patterns it should be understood that we are referring to a very passive form of causality. System structure does not serve as a stimulus that, once activated, somehow precipitates a given behavioral act in response. The process is more subtle and moves more slowly than suggested by the typical image of a cause and effect relationship. It is better to envisage the structure's influence on behavior as a kind of background condition or contextual effect that operates by constraining some types of actions and facilitating others.

This highly generalized description of the stratification process leads to two further observations about how that process might be explained. First, even though we are primarily interested in stratification as a systemic phenomenon our explanatory logic must take account of events transpiring beneath the system level. If we are to understand why behavior tends to coalesce into patterns dictated by the structural hierarchy we must understand something about why behavior occurs at all, and that requires reasoning on the level of the
individual actor. We shall see shortly that most authors have indeed followed this tack by framing their explanatory premises in terms of actor's purposes and dispositions. Incidentally, there is no logical prohibition against utilizing concepts at one level of social organization to enhance our understanding of phenomena on a different level, despite occasional misplaced protestations to the contrary (for example, Singer, 1961; Hanrieder, 1971).20

The second observation is actually a qualification of the first. Social behavior is most satisfactorily conceived as a purposeful enterprise that reflects the intentions and goals of the actor, but conceding this does not necessarily commit one to the proposition that the consequences of behavior are intended, or even that they are fully anticipated. Donald Moon (1975: 183) has suggested that this is particularly true of systemic level consequences: "Rarely do the 'gross' characteristics of social or political systems correspond to the intentions or conscious designs of anyone..." Galtung (1966c: 149) seems to imply otherwise by his observation that "...topdogs draw more attention than underdogs, and...they control their systems and, hence, prefer interaction patterns consonant with this." But this statement is exceptional; most interactionist writers, including Galtung, eschew the assumption that rank-dependent interaction patterns figure in the plans of the system's actors.

These preliminary remarks aside, let us take a closer look at the explanatory arguments contained in the literature. We shall begin with a fundamental premise embraced by virtually all interactionist
theorists: social units or actors are purposeful, goal-seeking entities. Earlier we alluded to behavior as purposeful, here we are simply referring to the other side of the coin. Frequently this goal-seeking presumption is built into the definition of rank or status. This is the implication of Galtung's (1966a: 122) contention that "...high rank is to be pursued and low rank to be avoided." Similarly, Rummel (1977a: 211) defines a status dimension as "...a continuum involving virtually universal international consensus as to which end is better or more desirable," and from this definition infers that actors will normally seek to improve their rank position. This desire for higher ranking, or upward mobility, is expressed as a central postulate in Gleditsch's (1970) axiomatic formulation. We might note here that in addition to acquiring higher rank position, this goal-oriented activity is also presumed to be directed toward achieving balance among a set of disparate rank positions (see note 7).

Loosely speaking, a goal can be defined as some preferred state of affairs and behavior is said to be goal-seeking or purposeful if it is consciously designed to bring about such a state. There is a neat, intuitive connection between goals and behavior at the individual level; however, the very concept of goal-oriented activity becomes problematic when imputed to collectivities. Must there be some centralized means of control before a collectivity can possess a goal? Is a group goal merely a distillation of individually held goals? Can goals be imposed from outside? What degree of concurrence is
necessary amongst the members to speak meaningfully of a collective
goal? Unfortunately, these perplexing issues have received very
little attention from stratification theorists in general or from
interactionist writers in particular.\textsuperscript{21} Interestingly enough, the
reasons why these issues have not been addressed are less puzzling
than the issues themselves.

A central theme expressed in much of Galtung's (1964; 1966a;
1966b; 1968) work, and to a lesser extent in Gleditsch's (1970), is
that concepts of rank and interaction constitute the core of a
general "structural" theory of social systems that is isomorphic across
levels of social organization. Isomorphism refers to a relation be­
tween two sets (or systems) such that the elements and relations of one
correspond to the elements and relations of the other (Galtung, 1968;
Kaplan, 1964). To initially explore the degree of isomorphism that
obtained between systems of individuals and international systems,
Galtung simply assumed that individuals and "nations" were comparable
units, although it is clear that he did so merely as a heuristic
expedient. Thus, for example, after developing his rank theory per­
spective exclusively at the individual level, he proceeded to "...de­
liberately commit what is known as the fallacy of treating the inter­
national system as if it were a system of individual actors—to see
how far this carries us" (Galtung, 1966b: 172). Elsewhere he
observed that "...to the extent the isomorphism holds [between indi­
vidual and international systems] we are actually developing a general
(social) systems theory, just as analogies between domestic and
international law point to a general legal theory. However, one cannot conclude without some words of caution, by specifying in concrete terms some of the implications of the truism that 'nations are, after all, not human beings'" (Galtung, 1968: 293). He went on to note that one point at which the isomorphism breaks down is in a decision-making body where multiple individual preferences compete against one another to determine the collective preference.

Galtung, then, was chiefly concerned with ascertaining similarities between individual and international systems as a basis for developing a general theory of rank and interaction. Although he recognized the conceptual difficulties associated with this approach, one gets the impression that he was willing to postpone treatment of these issues in order to advance the general outlines of the theory. In other words, much of Galtung's work seems to be pitched on the level of an analogy. Rummel, on the other hand, appears to be advocating the much stronger proposition that international stratification is actually an extension of the same process that operates at the individual level. We shall return to Galtung in a moment, but first let us consider Rummel's rationale.

To begin with, Rummel (1977a: 218) repeatedly refers to the actors in his system as "nations", although he does acknowledge that "each nation is itself a complex social system filtering outside events, transforming them into foreign policy decisions and actions". In order to accommodate various concepts and propositions of the stratification perspective normally applied only to individuals,
Rummel (1977a: 219) postulates what he terms the Elite Corollary: "A nation's elite identify with their [nation's] rank and status configuration. This corollary is the linchpin of Rummel's (1977a: 219) theory since, in his words, it "...enables us to apply individual level sociological propositions at the national level to understand how national statuses influence international behavior; it enables tying together status theory and foreign policy decision-making research." Thus, the assertion that nations aspire to positions of higher rank is to be understood in the context of this formulation to mean that a nation's elite are motivated to improve their nation's standing in the international hierarchy. At first glance, Rummel's assumptions appear to offer a reasonably clear path to the substantive propositions of individual level stratification theory, yet a closer inspection reveals a path strewn with just the sort of difficulties we have alluded to these past few paragraphs. It is time we spelled out these difficulties in greater detail.

The basic thrust of Rummel's explanatory logic is that national elites orient much of their international behavior toward enhancing and balancing their national statuses. Why should these goals dominate foreign policy? The desire for upward status mobility, according to Rummel (1976a: 138) "...is a commonplace of everyday observation, introspection, and sociology. The why of this universal urge is explained by its nature as a fundamental psychological need." Here is where Rummel's logic breaks down. Even if we are willing to concede that there is a need for status that parallels
other basic drives such as hunger, thirst, sex, or safety, we can find no good reason to suppose that such a need would be routinely fulfilled by national status rather than by a more personal representation of status, the Elite Corollary notwithstanding. But let us go a step further and assume, for the sake of argument, that national status does contribute to one's psychological status-needs. Once we admit national status what is to prevent us from considering other group affiliations—such as primary work group, political organization, or ethnic heritage—as additional sources of status gratification? From this vantage point, the effects of a psychological need for status are apt to become rather muddied as a principal source of international behavior, although we would not recommend discounting the possibility of such effects altogether. In essence, the ambiguity created by treating individual status and the status of one's group as functional equivalents is a special case of the more general problem of drawing inferences about traits of individuals from a knowledge of the groups to which they belong, and vice versa. Another manifestation of this problem arises whenever there is reference to goals and goal-oriented behavior of collectivities. When we characterize behavior of an aggregate as purposeful, we, like Rummel, usually mean that purposes and goals reside in the minds of individuals comprising the aggregate. This line of reasoning falters, however, if we admit the not unlikely possibility that individual behavior, even in a collective context, is actually motivated by such privately held goals as personal power, income, prestige, security, convenience, and so forth.
Rummel's explanatory rationale lacks credibility for other reasons in addition to those already mentioned. Implicit in his formulation is the dubious assumption that upward status mobility and balanced ranks are universally regarded as the predominate goals of national elites. Furthermore, he seems to imply that all individuals share a common and presumably accurate interpretation of their nation's position in the international structure. In short, Rummel makes no allowances for individual differences in perception or preference, a procrustean approach that stems from his exclusive reliance on the supposed universality and dominance of status-related psychological needs as the ultimate explanatory concept. The conclusion that emerges from all of this is that we must look beyond the sociological and psychological principles responsible for stratification at the individual level if we are to arrive at a satisfactory account of the linkage between hierarchical structures and patterns of behavior in international systems.

Thus far in our discussion we have considered only those explanatory assumptions that deal with actor motivations. We have seen that the typical strategy of attributing goal-oriented behavior to collectivities is marked by certain conceptual difficulties that have yet to be resolved. How does positing the purposefulness of behavior contribute to our understanding of the stratification process? One possibility is that rank provides the capacity to achieve objectives through interaction. By this rationale, interaction will increase as rank increases because the higher an actor's
rank the greater is its capacity to interact successfully to accomplish its goals. Dechmann (1972: 197) puts it this way: "Since rank level is defined as the possession of central goods and since the possession of central goods implies the ability to manipulate others a high rank position is associated with a high degree of ability to achieve one's own goals...through interaction while a low position is associated with a low degree." In contrast to Rummel's psychological approach, this might be termed an instrumental interpretation of stratification since rank is treated as a kind of resource conducive to realizaion of goals independent of the actor's status position. In other words, rank is interpreted more as a means than an end in itself.

If an actor's own rank is associated with its ability to interact successfully with others then it would seem reasonable to infer that success might further depend on the ranks of those to whom interaction is directed. This argument has been developed by Galtung to extend the logic of this instrumental view to the dyadic level. His reasoning is best expressed in the following passage:

Rank is a kind of resource, and rank concordance means a heavy concentration of resources among the people who are high on all dimensions and a similar deprivation of resources from the units that are low on all dimensions. Interaction will often presuppose resources just as much as it will beget resources; for that reason there will be more interaction, the more resources are present. But, in addition to that, the top dog unit will prefer to interact with another top dog unit for the simple reason that he can get more rewards from a top dog than an underdog. The top dog unit will at times want to interact with an underdog unit to get the kind of services the underdog can give him, and the underdog will certainly want to interact with the top dog unit. But to the extent
that we assume that any unit will try to interact with
the top because that is most rewarding, two top dog
units will be at an advantage because their wishes cor­
respond to each other, whereas the wishes of two under­
dog units will never correspond to each other and the
wishes of one top dog and one underdog unit only some­
times. (Galtung, 1966a: 150-151).

In summary, Galtung's argument is that actors, whether they be individ­
ual or the collectivities that normally operate at the international
level, will, on the basis of an evaluation of potential rewards, elect
to interact at the highest points in the system under most circum­
stances. In other words, Galtung here appears to be adopting Rummel's
approach by articulating more than merely an analogical relationship
between individual and international level stratification: he implies
that the very same explanatory logic is applicable at either level.

How is this rationale extended beyond the isolated act to account
for system-wide behavioral patterns? Galtung (1966a: 149) tells us
that "For every single unit there is a sort of intra-unit harmony,
based on equilibrium and generalization of role expectations, and for
every pair of units there is a similar harmony based on congruence
and generalization of role patterns". Two fundamental ideas, neither
originating with Galtung or others in the interactionist school,
underlie this contention. The first deals with the concept of roles
and affiliated role expectations as a set of more or less firmly
established patterns of behavior. Roles serve as a repository of
integrated and mutually supportive behaviors that come to be ex­
pected of the role occupant on the basis of widely recognized
characteristics, including social position or rank. Interactionist
Theorists commonly rely on the notion of role and related concepts when discussing conformity to the patterned behavior that is characteristic of stratification systems. Reinton (1968: 50), for example, refers to status as "...a denomination of role sets...."

The second idea, that social units display a strong tendency toward equilibrium or balance, is borrowed from students of social psychology. Interactionist scholars have posited a dual focus to this equilibrating mechanism. The first involves the unit's status set. It is this putative desire for balanced ranks that has given force to the status inconsistency hypothesis that has for so long occupied the attention of stratification theorists within the field of international politics and elsewhere. The second of these two foci concerns the desire for equilibrium between rank and behavior. Gleditsch, it will be recalled, addressed this point directly by defining behavior as a rank variable to be kept in balance with other variables in the unit's status set. Galtung is less clear on this but the passage just cited suggests that role patterns and expectations are generalized at least partially on the basis of this need for balance or "harmony".

To summarize, the instrumental logic advanced by Dechmann and Galtung interprets rank as a resource that contributes to the probability that an actor's goals can be achieved through interaction. Furthermore, this calculus takes into account the rank position of the recipient as well as the position of the acting party. Because units located at similar points in the structural hierarchy tend to
arrive at similar evaluations of potential rewards, their behavior begins to assume certain shared characteristics that become identified as roles associated with the positions they occupy. The emergence of structurally induced interaction patterns is thus a consequence of the behavioral manifestations arising from the maintenance of these international roles and role expectations.

Does this view fare better than Rummel's rendition of the stratification process? On balance our answer is yes, though with some reservations. In the first place there is still the problem of imputing goals and purposes to collectivities. The use of designations such as "unit" or "actor" to encompass entities at different levels of social organization only serves to perpetuate the erroneous image that such concepts can be transferred from one level to another without further explication or reformulation of their intended meaning. Precisely the same kind of difficulty arises with respect to the notion of roles and role expectations. Social theorists have long recognized the role concept to be an extraordinarily rich theoretical construct for analyzing the behavior of individuals, but without additional specification it is entirely inappropriate at the collective level. Nevertheless, the instrumental perspective does present an intuitively plausible logic that does much to illuminate the causal connection between positional structure and patterns of international interaction. In fact, we will find it useful to draw upon elements of this rationale when devising our own explanatory logic further on in this chapter.
Thus far this subsection has focused on explanatory assumptions either expressed or implied by interactionist writers. We have, to use a popular idiom, considered only the "sins of commission" contained in this literature. Are there significant omissions as well? For the most part, interactionist researchers have taken a fairly parochial view of the behavioral implications stemming from their explanatory arguments. The result has been an inordinate amount of concern for patterns based solely upon the frequency of rather specific classes of interactions or interaction channels with little regard for the variety of theoretically interesting properties that cut across discrete classes. For example, the numerous theoretical and empirical studies conducted by Galtung and his colleagues have examined behavioral patterns comprised of the incidence of diplomatic communication or representation, the number of shared organizational memberships, and the volume of international trade flows. Such patterns reflect the amount of participation in the system and indicate over what portions of the hierarchy participation is most frequent, however, beyond the nature of the specific categories used for recording purposes they can tell us little about the overall quality of participation. Several scholars, including Rummel, Deichmann and Galtung, have taken a preliminary step toward qualitative assessment by investigating patterns of cooperation and conflict resulting from stratification processes, but aside from this rather modest and predictable development the qualitative aspects of rank-dependent interaction have remained virtually unexplored.
The sources of this parochialism are reasonably clear. Earlier in this chapter we noted that the study of international stratification has been heavily influenced by findings gleaned from the research literature on small group dynamics. Many of these studies focused on the quantity of group interactions (see Berelson and Steiner, 1964) and this interest was subsequently carried over to the international level. Conflict and cooperation have long been a predominant concern in the field of international relations so it is not surprising that stratification researchers would consider these properties to be descriptive of the behavior patterns associated with the stratification process. But beyond efforts to characterize behavior as cooperative or conflictual there has been a singular lack of creative conceptualization and innovative instrumentation in stratification research as well as in international relations more generally. Of course, we must not underestimate the formidable difficulties that confront any attempt to expand our view of the world. In the first place, there is what Kaplan (1964: 53) has aptly termed the paradox of conceptualization: "The proper concepts are needed to formulate a good theory, but we need a good theory to arrive at the proper concepts". However, as Kenneth Boulding (1966) has observed, conceptualization alone is often not enough to insure the desired outcome—conceptualization must develop hand in hand with instrumentation, that is, the methods by which we detect and process relevant information about the outside world.
In summarizing this rather critical look at the explanatory arguments underlying interactionist approaches, we conclude that previous efforts have suffered in two general areas. First, they have failed to take account of the fact that most international actors are not single individuals but collectivities; consequently, they have mistakenly assumed that concepts meaningful at the individual level are directly transferable to collective entities. Even where there is explicit recognition that actors are aggregates of individuals, either the implications of this are ignored for heuristic purposes or the separate individuals are treated as though they shared identical preferences, motives, personality traits, and so forth. Second, there has been a serious lack of development concerning the range of behavioral properties susceptible to influences from the stratification process. It seems unlikely that the structural hierarchy could noticeably affect levels of participation in the system without also influencing other dimensions of behavior. Such implications should ensue directly from one's explanatory logic, yet they are seldom found in the extant literature. The following subsection presents an explanatory logic designed to overcome these particular difficulties.

A Problem-Solving Logic

Before beginning the task of developing our explanatory logic let us be clear about the limited purposes of this exercise. In the first place, we shall be concerned with the stratification process only as it applies to one species of international actor: the national government. Admittedly, this constitutes a rather severe
limitation since governments are responsible for no more than a fraction of all the activities occurring in the international system at any one time. Moreover, in light of the considerable influence exerted by some non-national actors, such as the giant multinational corporations or certain guerrilla organizations, one would be hard-pressed to argue that governments are necessarily more important or powerful than other types of actors. Nevertheless, confining our attention to national governments allows us to retain some comparability with previous stratification research. In addition, there is the pragmatic benefit of having greater accessability to data for national governments than for the panoply of non-governmental actors. Finally, it must be remembered that the central aim of this research is not formulation of a comprehensive theory of international stratification, rather, it is to evaluate the multiple systems hypothesis outlined in the previous chapter. The limitation imposed here does not appear to be incompatible with that purpose, although this is not to say that such a circumscribed view would be appropriate for other purposes.

Secondly, we should reiterate that our intention here is to advance an explanatory logic that will elucidate the hypothesized connection between positional structure and patterns of foreign policy behavior. We employ the special term "explanatory logic" to underscore the bounded nature of this enterprise. What follows is not intended, and should not be construed, as an attempt to explain the external behavior of national governments. Although we would hope that this effort might eventually be read as a modest contribution
to that more ambitious objective, our exclusive focus on the stratification process guarantees that it can be no more than that—a modest contribution. This is because foreign policy, at least insofar as it is currently understood, is believed to be the product of numerous factors, most of which are entirely independent of the stratification process outlined below (See Rosenau, 1966; Breecher, Steinberg and Stein, 1969; Andriole, Wilkenfeld and Hopple, 1975; Cohen and Harris, 1975; East, Salmore and Hermann, 1978).

Our explanatory logic is presented in what some (for example Waltz, 1975) would term a reductionist style. The major premise motivating this logic can be stated quite simply: Foreign policy, like all matters of public policy, may be depicted as a continuing effort by a government's duly constituted officials to recognize, confront and ultimately eliminate problems. The concept of problem serves as the fulcrum from which this logic acquires its theoretical leverage. A problem arises when there is noticeable divergence between some consciously preferred state of affairs and the state of affairs as presently apperceived or anticipated. Although we do not regard the traditional domestic-foreign policy distinction as a particularly useful taxonomic device, 24 we can identify certain conditions under which problems assume a "foreign" component "...including when the perceived source of the problem is beyond the political jurisdiction of the government, when all or part of the remedy requires actions by foreign entities, or when the effects—intended or otherwise—slip over and alter the relevant environments for external entities" (Hermann, Hermann and Dixon, 1979: 3).
Governments, then, are assumed to be problem-solvers or problem managers in the sense that their efforts are directed toward the avoidance or resolution of problems as defined above. To be sure, governmental officials are not always successful in their problem-solving endeavors but our immediate concern is with motives, not outcomes. There are those who would contend that this problem-solving assumption only glosses over the real motivations that underlie foreign policy behavior. For example, it might be argued that problem-solving is actually a means toward more fundamental objectives such as retaining political office or acquiring the support of society's politically active segments (for example see Salmore and Salmore, 1978). Without denying the considerable merit of this approach, let us simply assert that all explanatory arguments must come to rest somewhere. For our purposes the problem-solving assumption provides a suitable resting point.

From what has been said thus far it is clear that problems presume the existence of goals. Furthermore, because governmental behavior typically is the product of group level decision-making processes imbedded in a larger organizational context, we are immediately confronted with the necessity of attributing goals to collectivities. Students of organizational behavior traditionally have dealt with this issue in one of two ways: either by endowing the corporate entity with human characteristics in a transparent example of the anthropomorphic fallacy or by summing preferences across the individual members of the organization. For obvious reasons neither of these
solutions is satisfactory. As a way out of this impasse James Thompson (1967) has recommended treating the goals of those in a dominant coalition as the organization's goals. "In this view, organizational goals are established by individuals—but interdependent individuals who collectively have sufficient control of organizational resources to commit them in certain directions and to withhold them from others" (Thompson, 1967: 128). We find Thompson's strategy engaging for several reasons. It does not prohibit an organization from possessing multiple or even conflicting goals at any one point in time. Moreover, it allows for alternate permutations in coalition membership through time and it implies that often there will be less than unanimous agreement concerning the organization's goals. Finally, it makes no assumption regarding the size of a dominant coalition—it may be one individual or many. All of these are characteristics frequently ascribed to governments, particularly in the area of foreign affairs (see Allison, 1971; Halperin, 1974).

Lawrence Mohr (1973) has criticized Thompson's formulation on two counts. First, he noted that it is not always possible to identify those who are dominant or have power in an organization. Mohr's point is well taken in any event, but with the cloak of secrecy that customarily obscures high level governmental deliberations it is even more true of governments than of other types of organizations. Nevertheless, it does not damage the conceptual argument elucidating the meaning of organizational goals and the procedures underlying their selection. Mohr's second objection is more troublesome. Thompson's
strategy, he observed, "...discreetly attempts to substitute for the burden of defining the goals of an organization the equally difficult task of defining the goals of a group within it" (Mohr, 1973: 473). He suggested that a more reasonable conceptualization would incorporate group consensus as the pivotal concept: "Consensus means explicit or tacit agreement among those concerned that a certain behavior will under the circumstances be followed, notwithstanding the possibility that some might prefer other available alternatives. A group goal, then, connotes collective intent, e.g., a consensus of intent" (Mohr, 1973: 473-474, emphasis in original). In sum, we conclude that governments, as a type of complex organization, possess goals reflecting "a consensus of intent" reached by a dominant coalition of its authoritative decision-makers.

Foreign policy behaviors occur in response to perceived problems. Note that the existence of a problem does not necessarily precipitate a foreign policy behavior, although all such behaviors can be traced back to some problem. In the vocabulary of the decision-making literature, the problem supplies an "occasion for decision" (Snyder, Bruck and Sapin, 1962). We are now in a position to amend Hermann's (1972) definition, cited in Chapter I, by stipulating that foreign policy behaviors are discrete official actions intended to influence the behavior of other international actors for the purpose of alleviating, resolving, evading or otherwise coping with a problem. Interpreting foreign policy as a type of problem-solving activity establishes an immediate stimulus prompting the behavioral act as well as the motivating force underlying the behavior.
How does this problem-solving logic lead to expectations regarding systemically induced patterns of foreign policy behavior? We can bring these connections into clearer focus with the help of three additional auxiliary propositions. The first is offered as a definitional statement stipulating what we believe to be the key aspect differentiating the various positions arrayed along the system's structural hierarchy. Drawing upon the instrumental view expressed by Dechmann and Galtung, we envision a government's structural position as indicative of the available resources that can be brought to bear throughout the problem-solving process. We use the term "resources" in its broadest sense to embrace both human and nonhuman factors capable of facilitating the identification or resolution of perceived or anticipated foreign policy problems. Hence, our formulation posits international positional structure as the product of a system-wide distribution of problem-solving capabilities. To the extent that problem-solving requires freedom of action and an ability to influence the behavior of relevant others, we can conceive of international structure in more familiar terms as a positional representation of the distribution of national power.

Our second proposition is a simple extension of the first and provides an element of continuity linking our approach to previous interactionist literature. This proposition expresses an explicit reaffirmation of the familiar relationship between a government's hierarchical position and its level of involvement in global politics. Informed by our problem-solving perspective, the explanatory logic
underpinning this relationship is compelling and straightforward. From our amended definition of foreign policy behavior it follows that international involvement presupposes some concern on the part of governmental decisionmakers with what are perceived as extra-national problems. The number of such problems encountered by a government will depend for the most part on two factors: the diversity and specification of its goals and its capacity to detect and evaluate pertinent information. Recall that a problem arises only if there is an established preference regarding some state of affairs and if there is noticeable divergence from that preferred state. But actual involvement requires more than the emergence and recognition of a problem; it also requires the capability to formulate a response and to act accordingly. These interrelationships are displayed schematically in Figure 2.

As the figure indicates, each of the three requisite factors of international involvement is directly related to the government's position in the structural hierarchy. The capacity to monitor the environment and assimilate information and the ability to respond to identified problems are subsumed under our more general notion of problem-solving capabilities. In addition, compared to governments situated farther down in the hierarchy, those in the upper eschelons are likely to have larger administrative and bureaucratic support structures facilitating the establishment of preferences in a broader array of geographical and substantive arenas. In short, the relationship between structural position and involvement in international affairs is mediated by the government's ability to establish goals, to monitor its environment, and to respond to identified problems.
FIGURE 2

Structural Position and International Involvement
Our second proposition concerned the relevance of structural position to international behavior. The third extends this concern to the implications of positional proximity by asserting that structurally proximate governments—that is, those located near one another in the hierarchical structure—will tend to exhibit similar patterns of international behavior. For reasons explicated above, structural proximity will entail similar goal setting, monitoring, and response capabilities. Beyond these rather obvious sources of similarity we would further expect proximately positioned governments to settle on comparable sets of preferences and to encounter similar types of problems. In effect, we are arguing that a policymaker's point of view and interpretation of events are influenced in part by the positional context of his or her government in the world structure. Hence, it is no accident that governmental leaders from Third World countries tend to emphasize problems stemming from global inequality and acute poverty whereas individuals from industrialized nations are more likely to underscore problems endangering world peace and stability. Obviously, countless other factors are also involved in shaping an individual's world view; the point we wish to emphasize, however, is that a government's structural position may be a particularly potent factor for individuals engaged in the formulation and implementation of governmental foreign policy.

Let us summarize our development thus far. We began by characterizing the foreign policy process as a problem-solving enterprise. A problem was defined as a recognized discrepancy between what is
preferred and what actually exists or is anticipated. Because problems presuppose goals we found it necessary to explicate the meaning of that concept when attributed to collectivities. Organizational goals were said to represent a consensus within a dominant coalition of individuals able to commit the resources of their organization. Building upon this problem-solving orientation, we argued that the international system's hierarchical structure is defined by inequality in the distribution of problem-solving capabilities. Finally, two very general propositions were advanced to illustrate some important conceptual linkages between problem-solving behavior and our positional interpretation of system structure.

One further ramification of this problem-solving logic deserves special notice before we turn to the formulation of specific stratification hypotheses. We refer to the utility of the problem construct for delineating separate systems in accordance with the multiple systems model. A fundamental premise of this model is that each discrete system encompasses behavior from a relatively encapsulated substantive domain. This is the "no linkage" assumption discussed in the first chapter. In our estimation, reliance on the problem-solving framework significantly enhances the plausibility of this assumption since it focuses attention on the substantive component of the problem precipitating the behavioral act and it implies that individual systems are separately structured by substantively circumscribed problem-solving capabilities. In other words, it is assumed
that successful management of substantively distinct problems will require utilization of different types of resources. Thus, the problem construct serves double duty for the multiple systems model by facilitating differentiation of system specific behavior on the one hand and differentiation of corresponding system structures on the other.

If the problem construct is to be employed to its fullest advantage, great care must be given to the task of problem specification. This is especially true with respect to the level of abstraction used to characterize a problem's substantive content. For example, a situation in which the valuation of a country's currency experiences steep declines in international money markets could be construed very broadly as an economic problem or more narrowly as a problem of foreign exchange. The level at which problems are interpreted will affect the specification of relevant problem-solving capabilities as well as the inclusiveness and homogeneity of the problem systems. If problems are framed too narrowly then we face the needless proliferation of problem systems to the detriment of theoretical parsimony and generality; on the other hand, if problems are too broadly specified then systems will be overly aggregated and vulnerable to the same difficulties allegedly plaguing the single system perspective. For purposes of this research, problems are framed at a level suggested by the middle gauge typology of substantive problem areas developed by Hermann and Coate (1979). As indicated in the previous chapter, our analysis of the multiple systems hypothesis
will focus on two types of problems selected from this scheme to represent the substantive core of two prototypic problem systems. The various features of this problem area classification will be examined more thoroughly in Chapter IV; for the present let us limit our consideration to a brief description of the substantive content demarcating these two putative systems.

We have already mentioned that our two systems divide into the general areas of military security and international trade. The former is based on problems arising from conflict negotiations or military settlements of intergovernmental disputes. More specifically, this system embraces problems relating to such matters as the cessation of military hostilities through cease-fire arrangements, peace agreements, or formal capitulation; negotiations concerning exchanges of prisoners or agents; the determination of national borders; the creation, maintenance, or use of peacekeeping forces; and so forth. Expressly omitted are problems having to do with negotiations or agreements regarding nuclear or conventional arms control or disarmament; the formation or maintenance of military alliances; and the threat or actual use of military force (except peacekeeping forces) either in wartime or in isolated incidents. The trade system is concerned exclusively with governmental actions addressed to problems involving transactions of specific goods, commodities, or services; import protection through tariffs or quotas; negotiations or agreements regarding preferential trading status; and the establishment or maintenance of intergovernmental associations including common
markets, customs unions, and cartels. Again, it may be helpful to consider some types of problems not within the confines of this system. Specifically excluded are problems dealing with private transactions (except those subject to governmental agreement); foreign investment and regulation of multinational corporations; fiscal or monetary conditions, including problems of currency valuation and balance of payments; and any sort of economic assistance, whether through sales or purchases or through loans, credits or grants.

The Behavioral Effects of Stratification

This part of the chapter advances a set of empirically testable hypotheses derived from the problem-solving logic outlined above. Each hypothesis establishes a linkage between a government's structural position or the positional configuration of a pair of governments, and certain properties of intergovernmental foreign policy behavior. Before we proceed to the main task of this section let us summarize some key aspects of the behavioral dimensions represented in these hypotheses.

Behavioral Properties

Earlier in this chapter it was argued that stratification researchers generally have assumed a rather parochial view of the behavioral effects potentially attributable to system structure. We believe that a strong case can be advanced for the examination of multiple, qualitatively different properties of governmental behavior.
In the first place, such an approach constitutes explicit recognition of the richness and complexity that characterizes foreign policy behavior (Hermann, 1972; Brady, 1979a). Secondly, investigation of multiple behavioral dimensions is really the only way to determine how pervasive and multifaceted are structurally induced stratification effects. This is particularly important for the explanatory logic developed here since, as we will soon show, it generates expectations of observable effects across several dimensions of behavior. Finally, such a multipronged investigation can provide some assurance that the conclusions drawn from our research will be indicative of more than the peculiarities of a single behavioral dimension or its operationalization.

Having accepted the need for a multidimensional representation of behavior, our choice of particular properties must take account of three important criteria dictated by our strategy for investigating the multiple systems hypothesis. The most critical of these is that there be a clear and unequivocal conceptual linkage from our explanatory logic to the behavioral properties under consideration. This requirement follows from our decision to utilize a set of a priori expectations as a benchmark for comparing the empirical performance of the single system and multiple systems models. The second criterion is of a more pragmatic nature in that it constrains our choice of behavioral concepts to those that are reasonably susceptible to empirical inquiry. Although further elaboration must await our discussion of operationalization in Chapter IV, at this juncture we
should point out that this requirement involves more than "in principle" measureability. Our final criterion is necessitated by the problem or issue orientation motivating this study. Stated in the most general terms, our research question is aimed at uncovering differences between comparable analyses conducted at alternative levels of problem specification. To keep the comparison as tight as possible we must avoid introducing behavioral properties that are likely to be affected by, or to interact with, a problem system's characteristic substantive content.

Let us examine the rationale for this last criterion in a little more detail. If our research design were stated in the language of experimentation we would designate the single, comprehensive system as the control group and the two problem-specific systems as treatment groups, the treatment being a reduction in the variety of substantive problems represented within the groups. The purpose of the experiment would be to determine whether the treatment or the control provides the better interpretation of international political behavior. In order to conclude that any differences resulted from the treatment alone we would have to insure that the treatment and control groups were identical in every respect, save for the treatment itself. But this is just what we cannot do since application of the treatment—that is, narrowing our consideration to a fairly homogeneous set of problems—necessarily introduces a contaminating factor into the experiment in the form of the problem area's substantive core. Hence, it would be impossible to determine whether treatment group effects
resulted from the intended treatment or from some unknown dynamic associated with the substantive content of a particular problem area.  

Clearly it is impossible to eliminate the contaminating factor of problem substance from a design such as ours; nevertheless, there are precautions that can be taken to isolate or diminish its consequences. The surest method would involve constructing an exhaustive list of problems at a very low level of specification, analyzing each problem individually, and then comparing results from the separate analyses to locate specific instances where the problem substance and behavioral property evidence some interactive effect. An alternative measure that proved more feasible for the present study required only that we take care to select behavioral properties unlikely to be significantly affected by our problem systems' substantive content. In the absence of any empirical evidence in the extant literature, we found it a useful guideline to focus on properties tapping the style of diplomacy rather than those more indicative of the substance of policy (Brady, 1979a). Obviously, this procedure is far from foolproof since it rests almost entirely on the prudent judgement of the analyst. We will leave it to the reader to decide how well our behavioral properties meet this criterion.

The behavioral properties to be examined in this study divide into three general categories. The first, which we have already mentioned, refers to a government's level of involvement in international political affairs. There are two interwoven dimensions of involvement that seem worthy of investigation within our stratification framework.
One of these is simply the amount of foreign activity exhibited by a government. Clearly there is wide variation in the frequency with which governmental decisionmakers engage in activities beyond their national borders. The second dimension of involvement concerns what East and Hermann (1979: 180) have designated "scope of action", which they broadly define as "...the distribution of a government's foreign policy activity across potential targets or recipients in the international arena". Scope of action is used here to refer to the extent of geographical dispersion (or concentration) characterizing a government's external activities.

The second general category has to do with the degree of autonomy expressed in foreign policy behavior. It is important to distinguish between autonomy of foreign policy and autonomy of foreign policy behavior. Following Hermann (1979), we will focus on the latter concept since it entails implicit recognition that some types of actions may be more or less autonomous than others. Autonomy, in other words, is taken to be a property of a discrete behavioral act; thus, it is neither an attribute of a government nor a characteristic of the policy process. As with involvement, autonomy can be decomposed into two separate dimensions. The first interprets behavior as a response to some prior action by an external entity or, conversely, as an initiative with no prior external stimulus. "The initiative-reactive dimension", writes Hermann (1979: 416) "is intended to tap the extent to which the content of a government's present behavior is structured for it by entities outside of its
control". The second dimension concerns whether foreign behavior is conducted unilaterally by an individual government acting alone or multilaterally in collaboration with other national actors. Hermann (1979) has combined these two aspects of behavior into a single construct termed "independence/interdependence of action"; however, for our purposes it seemed advisable to preserve the distinction between the two dimensions by examining them independently.

For the lack of a better label we will refer to the third type of behavioral property as problem sensitivity. This property, which is derived directly from our problem-solving logic, locates the entity (or entities) that experiences the problem in relation to the governmental actor. Recall that a problem is defined as a perceived discrepancy between some goal and the condition or state of affairs that actually exists or is anticipated. Existence of a problem thus implies that some deprivation or foregone indulgence is directly experienced by one or more entities. The reason this is at all interesting is that the jeopardized entity need not be identical to the acting government or its domestic constituency. In general, we can identify three analytically distinct classes of entities on whose behalf an actor may address a problem.26 One encompasses the acting government itself or its entire national society, specific subnational groups, or individual citizens. In contrast, a second class includes only foreign governments, societies, groups, or citizens. The third class refers to the actor in combination with one or more external entities or to international organizations. From the actor's perspective these
three classes demarcate three types of problems which we can designate internal, external, and communal, respectively. Problem sensitivity, then, relates to the actor's sensitivity or attentiveness to these various types of problems and to the entities experiencing them. It will be useful to conceive of problem sensitivity as a property manifesting the inward or outward orientation of a government's foreign policy behavior.

Earlier in this chapter it was pointed out that stratification effects are more readily discernable in behavioral patterns than in terms of discrete behaviors. Having outlined the main features of the behavioral properties selected for empirical analysis, it is time we considered the patterning of these properties. Individual behaviors are discrete in that they are located in both space and time; patterned behavior, on the other hand, represents extension through one or both of these dimensions. Patterns extended through time are easily conceived as aggregations of temporally adjacent behaviors. The meaning of spatial extension in this context may be less self-evident but if we think of the spatial dimension in geographical terms, then spatial extension can be interpreted as aggregations of behaviors across multiple geographical units. According to our definition, discrete foreign policy behaviors are geographically fixed with reference to the acting government and to one or more recipient entities. Hence, spatially derived patterns can be generated for groups of actors, groups of recipients, or various combinations of both actors and recipients. In what follows
we will be concerned with two of the more frequently studied types of behavioral patterns. We refer to patterns of monadic behavior, which are aggregated by actor across time and recipients, and dyadic behavior, which are aggregated across time for particular actor-recipient combinations. Let us now consider the emergence and form of those behavioral patterns in relation to their surrounding structural context and our problem-solving logic.

Hypotheses

Each of our hypotheses is stated in a simple, declarative style and followed by a brief elaboration of its supporting rationale. We should point out that our repeated use of probabilistic language is quite deliberate and is intended as an implicit acknowledgement that system structure is only one of several factors shaping patterns of foreign policy behavior. The first four hypotheses deal exclusively with monadic behavior and the potential effects of absolute structural position.

1. The higher a government's position in the structural hierarchy, the more likely it is to engage in a high level of foreign activity.

Level of foreign activity is one aspect of a government's involvement in international affairs and has already been discussed in some detail. To recapitulate, government's enjoying a high structural position have available a relative abundance of problem-solving capabilities compared to lower positioned governments and this
implies that more goals can be initially specified, then monitored and potentially identified as problems, and finally acted upon if deemed necessary. The rationale for this hypothesis rests solely on the government's capability to assume an active and prominent role in world politics—it says nothing about the decisionmakers' disposition to assume such a role. Dispositions do enter into our explanatory logic but only with respect to the problem-solving motivation for external behavior. This is one reason why it is imperative that this hypothesis, and the others as well, be couched in probabilistic language. Hence, it would not be a contradiction of our explanatory logic, although it would be contrary to expectations, to find that a highly positioned government had withdrawn from international affairs through a conscious policy of isolationism.

2. The higher a government's position in the structural hierarchy, the more expansive its scope of action is likely to be.

Scope of action, which refers to the dispersion of external activities across potential foreign recipients, represents another aspect of international involvement. The explanatory rationale for this hypothesis closely parallels that given above. Here, however, the sheer number of specified goals is less important than their geographical distribution. Likewise, this hypothesis emphasizes the breadth of monitoring and response capabilities that accrue from a government's location in the structural hierarchy. We might also point out that
this hypothesis, along with the first, is a direct implication of the rank-dependent interaction proposition that figures so prominently in the interactionist writings of Galtung, Gleditsch, and others.

3. The higher a government's position in the structural hierarchy, the more likely it is to initiate its foreign activity.

A behavior is considered an initiative if it is not explicitly elicited through the prior actions of another actor or through some routinized set of interactions such as an ongoing negotiation or a formal vote in an intergovernmental assembly (Hermann, 1979). In general, it probably requires more organizational resources to formulate and implement an initiative type behavior than to react to some highly structured external stimulus. Of course, both types of behavior involve some expenditure of resources and therefore both are likely to occur with greatest frequency among the highest ranking governments. This hypothesis, however, focuses on the overall mixture of initiative and reactive behaviors, not on their raw frequencies. What, then, would account for the pattern indicated by our hypothesis? Owing to their more diverse interests and widespread involvement, the higher ranking governments will probably encounter more opportunities for initiative type behaviors than actors positioned further down in the structure. Moreover, with ample resources available to them higher ranking governments are better prepared to act on those opportunities. At the other end of the structural continuum are actors
with fewer resources and fewer opportunities to use them. We would expect that these actors will be better able to respond to previously structured situations than to formulate and enact initiatives of their own. The result of all this is that higher ranking governments will tend to initiate a relatively large proportion of their foreign activity whereas lower ranking governments will display a countervailing tendency toward more reactive patterns of behavior.

4. The higher a government's position in the structural hierarchy, the more likely it is to act unilaterally in foreign affairs.

Like the previous hypothesis, this one also concerns the autonomy of foreign policy behavior. Here again, we must overlook the frequency of behaviors in order to focus on the overall pattern revealed in the mixture of unilateral and multilateral activities. Some authors have suggested that governments with limited resources will tend to emphasize multilateral forms of diplomacy—whether through global or regional organizations, international conferences, temporary coalitions, or other collaborative activities—as the more economical way to conduct their foreign relations. This argument is quite reasonable up to a point, but we should also note that compared to unilateral behaviors, concerted actions frequently impose additional costs necessary for coordinating policies and maintaining intergovernmental cohesion. Yet even these additional costs would seem to be worthwhile if we further assume that collaboration serves to enhance
the visibility of international behavior and to dramatize the problem(s) underlying that behavior. Such measures will be particularly important for lower ranking governments only peripherally involved in international affairs. Higher ranking actors, on the other hand, are generally better equipped to address their problems unilaterally and may be less willing to compromise or coordinate their positions for the sake of what to them are only marginal benefits accruing from multilateral behavior.

The remaining hypotheses deal with patterns of intergovernmental behavior occurring between different points in the structural hierarchy. The monadic patterns of behavior considered thus far accommodate only those systemic influences emanating from the actor's own position in the structure. In effect, we have based our first four hypotheses on the assumption that a government's foreign behavior is pretty much the same regardless of the recipient(s) to whom it is directed. Though clearly an oversimplification, this assumption has served a useful purpose in focusing our attention on the gross effects of the actor's structural position. It is now time to relax this assumption, to introduce a refinement in our thinking to take account of the actor's position relative to that of the recipient. We will implement this modification by shifting our attention to dyadic patterns of behavior. Note, however, that the dyad entails a simplifying assumption of its own, namely, that each discrete foreign policy action involves no more than one recipient. We cannot accept
this assumption as descriptive of governments' actual behavior in the international arena; nevertheless, as a first approximation it can provide a useful vehicle for investigating the behavioral implications of relative structural position.

In the following six hypotheses we consider two analytically distinct sources of relative structural effects. The first three hypotheses concern the potential influence of the absolute (rank) distance between the positions occupied by the actor and its recipient, or what we shall designate as structural proximity. In any given dyad the actor may be located either above or below the recipient's position. The implications of dyadic directionality—that is, whether behavior is directed upward or downward in the hierarchy—are examined in the final three hypotheses.

5. The closer the structural proximity of two governments, the more likely they are to act on communal problems in their mutually relevant foreign behavior.

Mutually relevant behavior encompasses any activity occurring along either of the two possible actor-recipient dyads uniquely defined for a pair of governments. This hypothesis derives from our earlier assumption that the preferences and perceptions of individual policymakers are to some degree responsive to their government's structural position. If closely positioned governments possess similar problem-solving capabilities and express similar goals then it seems reasonable to suppose that they experience somewhat similar problems as well.
Hence, we conclude that foreign policy actions dealing with communal types of problems are likely to involve structurally proximate recipients. Similarly, we would anticipate that structurally distant governments would encounter few of the same problems and therefore have few opportunities to engage one another on communal problems.

6. The closer the structural proximity of two governments, the more likely they are to engage in high levels of mutually relevant behavior.

This hypothesis has to do with the level of activity between governments rather than the nature of the problems precipitating that activity; nevertheless, it follows closely from the explanatory rationale presented above. It was assumed that structurally proximate governments will experience many of the same types of problems because they share similar concerns and possess similar capabilities. In many cases we would expect these similarities to nurture a sort of interdependence or mutual relevance eventually leading to comparatively high levels of activity between closely positioned governments. We must qualify this statement with two caveats. First, structural proximity is not necessarily a positive bond encouraging cooperation and friendliness between governments; in some instances it may underlie a competitive or even a hostile relationship. For this reason the hypothesis concerns only the amount of activity and not its affective content. Second, we do not consider structural proximity to be a sufficient condition for high levels of intergovernmental
activity. Other factors, such as geographical proximity or ideological similarity, are likely to be of more importance in building mutual relevance between governments. Structural proximity should be viewed as a more subtle influence contributing to the relevance engendered by these other factors. This qualification is needed to account for the large number of structurally proximate though behaviorally empty dyads typically located in the middle and lower levels of the hierarchy.

One other matter deserving comment is the system-wide pattern of behavior implied by the conjunction of hypotheses one and six. If these hypotheses are correct we would expect to find the highest levels of activity at the top of the hierarchy among closely positioned governments, a pattern consistent with the stratification theories of most interactionist writers. However, our approach departs from previous efforts by suggesting that structural proximity will tend to encourage interaction even among lower ranking governments.

7. The closer the structural proximity of two governments, the more likely they are to engage in collaborative foreign activity.

This is the final hypothesis concerning the effect of structural proximity. Once again we refer back to our discussion of hypothesis five and the assumption that closely positioned governments will tend to encounter similar types of problems. Since collaboration entails joint action on a commonly perceived problem we would expect
collaborative activity to be heaviest between structurally proximate governments. Furthermore, when this hypothesis is considered alongside hypothesis four there emerges another overall pattern, this time having to do with collaborative behavior. According to these hypotheses, collaboration should be proportionately more frequent between proximate, lower ranking governments than between either structurally distant or higher ranking governments.

Although the following three hypotheses remain fixed on the dyadic unit, they do involve a shift in emphasis from the proximity of structural positions to the distance between them. However, the more distinctive feature of these hypotheses is their special concern for the structural directionality of dyadic behavior. Direction is defined relative to the acting government rather than with reference to some international average or threshold level. In what follows, we use the term "positively distant" to indicate that behavior is directed upward in the hierarchy by an actor positioned beneath its recipient; likewise, "negatively distant" refers to dyads in which the actor ranks above the recipient.

8. Relative to the acting government's position, behavior directed toward positively distant recipients is likely to concern more internal or communal problems than otherwise and, conversely, behavior directed toward negatively distant recipients is likely to concern more external problems than otherwise.

Perhaps the first thing to notice about this hypothesis is that it is actually composed of two separate assertions delineating the
countervailing effects of structural directionality. At first glance, the reference to communal problems in the first part of the hypothesis may appear to contradict hypothesis five, which maintains that communal problems are more pertinent to structurally proximate governments. An examination of the supporting rationale indicates that this contradiction is more apparent than real. We can assume that in most cases behavior is directed to positively distant recipients in order to take advantage of their superior problem-solving capabilities. Often such behavior will be motivated by the actor's own internal problems. On the other hand, low ranking governments may find it a useful strategy to act on more broadly based problems as a way to gain the attention of higher ranking recipients. The compatibility of hypotheses five and eight rests on this important point: whenever a foreign policy action is motivated by a communal problem, the recipient(s) may or may not directly experience that problem. More specifically, the logic underlying hypothesis five tells us that if actor and recipient are fairly proximate then communal problems are likely to affect both governments; in contrast, if the recipient is ranked far above the actor then communal problems are more likely to involve only the acting government and its more proximate associates. In sum, we conclude that positively distant foreign policy behavior generally will be precipitated by the actor's concern for its own internal problems or for more broadly experienced communal problems.

The rationale supporting the second part of the hypothesis mirrors that given above. We can conceive of two very general
conditions under which a government is likely to be addressed as a recipient of a foreign policy action pertaining to the acting government's internal problems: (1) if the recipient government were causing (or perceived as causing) the actor's problem, or (2) if the recipient were in a position to alleviate the problem through some subsequent action of its own, such as assisting the actor directly or influencing the behavior of some third party. Owing to their diminished problem-solving resources, negatively distant recipients are not very likely to fulfill either of these conditions. Of course, the same reasoning would apply to communal problems which the actor shares with its more proximate structural neighbors. By a process of elimination we are left with only the external class of problems as the most likely subject of negatively distant behavior. Taking a more direct approach, we would argue that under most circumstances actors will be more willing to take on others' problems if they can do so by addressing a negatively distant recipient. Whatever the actor's strategy for coping with an external problem—whether it is to help the deprived entity or to influence a third party causing the problem—it will probably be less costly, more successful, or both when directed to a lower ranking recipient.

9. Relative to the acting government's position, behavior directed toward positively distant recipients is likely to be more reactive than otherwise and, conversely, behavior directed toward negatively distant recipients is likely to be more initiatory than otherwise.
This hypothesis builds upon the rationale previously developed in conjunction with hypothesis three. We have already seen that higher ranking governments will have a tendency to initiate much of their foreign policy behavior and that lower ranking governments will tend to be more reactive. Having said this, we might conclude that the hypothesized pattern of dyadic behavior is a natural result of a closed system in which reactive behaviors emanating from the lower part of the structure are responsive to initiatives formulated at higher levels. This "initiative-begets-reaction" sequence is congruent with the hypothesized behavioral pattern and no doubt occurs with some regularity; however, we must not assume that initiative and reactive behaviors necessarily exist in one-to-one correspondence. Some initiatives, for example, may never generate any sort of responsive actions or may indirectly precipitate actions that appear to be more initiatory than reactive; on the other hand, some reactive behaviors may ensue from prior reactions rather than prior initiatives. This caveat notwithstanding, we can still assume that governments engaged in positively distant behavior typically defer to the superior problem-solving capabilities of their recipient by responding to actions previously tendered by the higher ranking government. Similarly, we would expect most governments to take the lead or to counter with their own proposals or initiatives when dealing with negatively distant recipients.
10. Relative to the acting government's position, behavior directed toward positively distant recipients is likely to be more multilateral than otherwise and, conversely, behavior directed to negatively distant recipients is likely to be more unilateral than otherwise.

The logic underlying our final hypothesis rests on the important distinction between collaboration, the subject of hypothesis seven, and multilateral behavior. Both terms refer to foreign policy actions taken in concert with other governmental actors. We use the term "collaboration" to identify actions taken jointly with the dyadic recipient whereas "multilateral behavior," as used here and in our discussion of hypothesis four, is a broader term that includes all collective activities regardless of the particular recipient's active participation. Keeping this distinction in mind, we note that the first part of the hypothesis asserts that positively distant behavior tends to be multilateral but not necessarily collaborative. In fact, hypothesis seven tells us that collaboration between structurally distant governments should occur rather infrequently if at all. The implication is that positively distant behavior will involve the acting government in concert with its more structurally proximate neighbors. In addition, this pattern reflects our earlier contention that multilateral activity is a useful strategy for lower ranking governments because it dramatizes their problems and enhances the visibility of their foreign behavior. Furthermore, we would expect higher ranking governments to engage in collaborative activity among themselves but not in their dealings with negatively distant recipients.
Summary

Since this part of the chapter completes the conceptual foundation of our inquiry, it will be useful to take stock in what has been accomplished thus far and to preview the tasks that lie ahead. In addition, it is important that this chapter's near exclusive focus on international stratification not obscure the main objective of the investigation, which is to compare and evaluate the single system and multiple systems models of world politics. These concerns suggest the need for a summarization of our stratification theory within the context of our overall research strategy.

The primary aim of this chapter was to articulate a set of theoretically motivated expectations or hypotheses regarding empirically observable linkages between system structure and patterns of foreign policy behavior. These expectations form the central core of our research strategy since they are to be used as a kind of benchmark against which we can assess the linkages actually observed under alternative system representations. It is obvious that whatever confidence we have in the research strategy as a whole will depend a great deal on our confidence in these ten hypothesized relationships. In this regard, special notice should be given to three features deliberately built into the set of hypotheses. First, we have departed from previous stratification research by formulating hypotheses to deal with the autonomy and problem sensitivity of foreign behavior, in addition to more conventional hypotheses concerning international involvement. To recapitulate, analysis of several
behavioral properties will insure that our overall results and the conclusions based on them are broadly reflective of stratification effects and are not merely peculiar to a single type of behavior. Second, our hypotheses are framed in such a way as to accommodate structural effects on both monadic and dyadic patterns of behavior. Here again, we assume that examination of both types of patterns will provide a broader and more conclusive basis for conducting an evaluation of the two system models.

The third and possibly most important feature is that the hypotheses are not a series of mere ad hoc speculations; on the contrary, they represent a set of interrelated propositions held in place by a common explanatory logic. In fact, we would contend that these hypotheses together with the explanatory logic constitute a theory of international stratification, albeit one lacking rigorous development in axiomatic-deductive form. It is designated a stratification theory because it delineates a process that accounts for interaction patterns in terms of the social system's hierarchical structure. Unlike most other theories of international stratification, our hypothesized connections between hierarchical structure and international behavior do not rely on loose or misplaced analogies to individuals but on definitions and assumptions about national governments and their foreign policy behavior.

The remaining tasks called for by our research strategy are reasonably well defined. The most pressing matter involves operationalizing the concepts in the ten hypotheses enumerated above.
System structure is the pivotal concept in each hypothesis and the key to system differentiation; its operationalization is the subject of the next chapter. The operational procedures for the behavioral properties are quite different and are discussed in Chapter IV. Once these operational tasks are completed we will be in a position to test the hypotheses and interpret the results.
1Note that this definition of process is virtually identical to what some writers would designate a system (see Blalock, 1964; Nurmi, 1974).

2The terms are borrowed from Little (1978) although our usage of them differs slightly from his. If only to point out the obvious, we should mention that the label of each approach refers to what is posited as the output variable of the stratification process.

3In the literature, the terms "discrepancy" and inconsistency" are frequently used as synonyms of "disequilibrium".

4For a concise review of the social psychological research literature investigating the frustration-aggression hypothesis, see Larsen (1976: chapter 4). Galtung (1964; 1966b) presents a short but useful bibliography and summary of the literature on status disequilibrium.

5Operationally, the studies by Schwartzman and Mora y Arujo closely resemble the more familiar approach linking internal conflict to relative deprivation (see Gurr, 1970). Unlike relative deprivation theorists, however, Schwartzman and Mora y Arujo postulate that the stratification of the international system generates expectations of upward mobility and internal tensions that may be released through anti-status quo internal conflicts.

6Prestige was first employed as a relevant dimension of international status by Lagos (1963). Subsequent writers, however, have been divided regarding its significance. East (1972) used a reputational measure of prestige as an indicator of ascribed influence. Wallace (1973), on the other hand, rejected the sociological distinction between achieved and ascribed status as inappropriate to analyses conducted at the international level. He maintained that the considerable mobility in reputational status precludes designating it "ascribed" in the traditional sociological sense. Rummel (1977a) refrained from using prestige as a status dimension altogether because he claimed it was linearly dependent on two other dimensions, wealth and power. Incidentally, Rummel does retain the achieved-ascribed distinction for wealth and power, respectively.
The desire for balanced ranks is a core assumption of most interactionist theories of international stratification. Galtung (1966b) refers to this balancing principle as the "axiom of rank equilibriation" and Rummel (1977a) terms it the "equilibriation theorem". In addition, Rummel uses the concept of balance as a bridge to other bodies of theory dealing with cognitive dissonance and relative deprivation. For sociological applications at the individual level see Zelditch and Anderson (1966) and Kimberly (1966).

Galtung (1966c) explains these findings by arguing that tourism is less a matter of governmental politics than individual choice and thus is likely to be more influenced by a country's scenic value than its rank position. To the extent that these findings are representative of the global network of tourist flows—and it is not clear that they are since Galtung examined only tourism to Warsaw Pact countries from NATO and Warsaw Pact countries—tourism may be an effective countervailing influence on international feudalism.

Commenting on this rationale East (1970: 120) writes: "It is the unequal distribution of resources and influence in the system that is the key to this explanation and this is also the key element in the concept of stratification". We should note too that East's reasoning is not inconsistent with Gleditsch's results since he found only a modest .3 correlation between combined rank scores and shared IGO memberships when controlling for geographical distance.

Galtung (1966c) attempted to gather data on other types of conflictual interactions as well—including trade, tourism, and visa restrictions and the closing of cultural institutions—however, those events occurred too infrequently to allow anything but very tentative and impressionistic conclusions.

Unlike other authors discussed thus far, Wallensteen defined high ranking nations as those most centrally located in the global trading network.

Originally social field theory was cast in a rigid determinism but Rummel has since altered his views significantly. This shift in Rummel's thinking is outlined in a recent autobiographical sketch (1976b) and in his annotations to a volume of collected research reports (1977a). One result has been that field theory has taken on certain aspects of a multifaceted world view anchored in a unique combination of social science and philosophy (see Rummel, 1975; 1976a; 1977b). For a somewhat dated but still useful summary and critique of field theory, see Hilton (1973).
These procedures are discussed in some detail in Rummel (1977a; 1979), but a few brief comments are pertinent here. For those unfamiliar with the evolution of field theory, Model I presumed that the linear combination of attribute and behavior dimensions were weighted identically for all nations, whereas Model II allowed these weights to vary from nation to nation. For further elaboration of the differences between Model I and Model II, see Hilton (1973).

Status-field theory as formally developed in Rummel (1977a: 199-225) contains thirteen theorems, however, only five of these represent hypotheses suitable for empirical investigation. The five concern status-dependent cooperative behavior (Cooperation Theorem), status-dependent conflict behavior of developed nations (Economically Developed Conflict Theorem), status-dependent conflict behavior of poor nations (Economically Underdeveloped Conflict Theorem), overall status-dependent behavior—that is, cooperation plus conflict—of wealthy nations (Economically Developed Status Behavior Theorem), and overall status-dependent behavior of less developed nations (Economically Underdeveloped Status Behavior Theorem). Not one of these five specifies the type of conflictual or cooperative behavior predicted—that is left to be determined empirically by inspection of the variates on the behavior side of the canonical equation.

Finally, we should note two subtle but meaningful distinctions between the methodology initially used to test field theory and the methodology applied to status-field theory. The first involves the "principle of diminishing salience" which asserts that actors are better able to accurately perceive and discriminate distances of small magnitudes than distances of large magnitudes. In other words, as the magnitude of distances increases it becomes less salient for a given actor. Rummel incorporates this principle into his analyses by log transforming the attribute data. The second consideration evolved from a recognition that part of the variance in attributes and behavior is unique to each actor and therefore does not contribute to the common dimensions in attribute space and behavior space. This notion is reflected in the shift from the principal components factor model to image analysis, a common factoring method. For explication of the differences between these two models, see Rummel (1970).

This is in contrast to the Rummel and Vincent analyses which were conducted separately for each acting nation as required by Model II (see note 13). In effect, Tomlin and Bhulman have developed a field theory model that stands midway between Models I and II.

This characterization is not meant to imply that the interactionist approach adheres to the state-centric paradigm any more closely than most other areas of international relations scholarship. For a valuable if somewhat impassioned discussion of the state-centric model see chapters one and two of Mansbach, Ferguson and Lampert (1976).
Rummel has addressed this point directly. In his original formulation of status-field theory he used the terms topdog and underdog, however, in a more recent piece he noted that "...these terms imply a moral-political view I do not accept: that low-status people are the victims of social injustice and the high-status people are exploiters" (Rummel, 1976a: 147, f.n. 16).

Emphasis here is on "directly applicable". Typically international actors are aggregates of individuals who are subject to the same sociological and psychological forces that govern the behavior of other individuals. It may very well be that these principles account for the behavior of the collectivity as well as for the behavior of individuals within it; nevertheless, in the case of collective behavior this linkage is mediated by an additional step (or series of steps) in which individual preferences and behaviors are somehow transformed into group level action.

In the extant literature Galtung's (1971) theory of imperialism probably comes closest to positing distinct issue-based structural hierarchies, although this is done more from a structuralist than an interactionist perspective. We should note, however, that Galtung's extensive writings defy easy classification into the structuralist or interactionist categories since he has dealt so effectively with the concerns of both. Nevertheless, we have tried to identify the major thrust of his approach on an article by article basis.

The nature of social explanation continues to be a vigorously contested issue among both practitioners and philosophers of social science. Although there are numerous subsidiary issues contributing to this controversy, the main battle lines are drawn between those who would emulate the natural sciences by adopting the nomological or covering law model of explanation and those who insist that human behavior is governed by purposes, intentions and ideas that cannot be fully accounted for by the nomological model. The literature on social explanation is immense. Moon (1975) provides a lucid overview of the major issues along with an intriguing proposal that incorporates elements of both positions. Arguments on both sides particularly relevant to international relations are found in Harf, Moon and Thompson (1975); Raymond (1975); Reynolds (1973); Tanter (1972b); and various contributions to Knorr and Rosenau (1969).

The term "explanatory logic" is borrowed from Hermann and East (1978). We prefer this expression to the more commonly used "explanation" for two reasons. First, it avoids any preconceptions regarding the appropriate structure of an explanatory argument. Second, "explanation" is the term usually applied to an argument that accounts for a particular event or behavioral act rather than a general linkage between two variables.
The reticence concerning mixed level explanations stems, in part, from a misunderstanding of the ecological fallacy, which, of course, refers to fallacious inference across levels of aggregation. For a detailed treatment of permissible ecological analyses see the volume edited by Dogan and Rokkan (1969) and the monograph by Langbein and Lichtman (1978).

Bath and James (1976) deal with many of these issues in their perceptive and compelling critique of the structuralist practice of treating actors as rational, monolithic entities. See Hermann (1978) for a more general discussion of goals as a concept in foreign policy analysis.

A more sophisticated and compelling argument for considering the influence of personality characteristics on foreign policy behavior is proffered by Margaret Hermann (1978). Unlike Rummel, Hermann maintains that individuals differ in the extent to which they are driven by different psychological needs, that personality characteristics are likely to be more potent in some situations than in others, and that any psychological effects that finally do influence behavior must find their way past the structure and dynamics of the decision-making body. For an empirical analysis of part of Hermann's formulation see Hermann, Hermann and Dixon (1979).

The use of roles and related concepts in international political analysis is by no means limited to the stratification literature. For example, Holsti (1970) and Wish (1977) have studied role conceptions held by national leaders and Volgy and Quistgaard (1975) have examined delegate roles within the context of the United Nations. Each of these studies considers roles of national governments but only insofar as they are expressed in the statements or actions of individuals representing those governments.

One of the fundamental requirements of a good taxonomy is that its categories be mutually exclusive. It is clear that in the present era of energy politics and economic interdependence the division between domestic and foreign policy is unable to meet this criterion. See Meehan (1971) for an illuminating discussion of the difficulties involved in defining foreign policy.

For illustrative purposes, suppose we elected to examine a stratification hypothesis dealing with structural influences on the affective content of intergovernmental behavior. In this case we might expect to find some interaction between affective behavior and the substantive nature of our problem systems. More specifically, it seems reasonable to expect that the trade system would involve behavior generally more positive in affect than either the military system or the overall system. (In fact, this example is based on more than mere supposition. The behavioral data described in Chapter IV...
indicate that trade related behaviors are considerably more positive in their affective content than behaviors in either of the other two systems). Just how this situation might interfere with our analysis (if it interferes at all) will depend on the precise statement of the hypothesis. The point, however, is that affect is one behavioral property that seems to be closely connected to the substance of behavior (also see O'Leary, 1976).

26 The language we use to discuss these entities will depend in part on whether we assume a problem oriented or a goal oriented perspective. For example, Brady's (1979b) emphasis on goals leads her to characterize these entities as beneficiaries (that is, entities who benefit if the goal is achieved) whereas Hermann and Coate's (1979) problem oriented approach encourages reference to deprived or jeopardized entities. The meaning is the same in either case.

27 In essence, we are distinguishing between two subclasses of communal problems on the basis of whether or not the recipient is itself directly affected by the problem in question. We have argued that this distinction is conceptually important for maintaining consistency among our hypotheses. In principle, the implications of this distinction are empirically testable; unfortunately, the behavioral data used in the present research (see Chapter IV) gloss over this distinction and thereby preclude its empirical investigation.
CHAPTER III
DATA AND METHODS I:
ASCERTAINING INTERNATIONAL STRUCTURE

Every program of scientific research eventually must incorporate a set of procedures for linking its concepts to appropriate aspects of its referent world. Such a procedure, which is conventionally termed an operationalization, typically involves classification or measurement of phenomena according to explicit rules. In general, measurement is a process of mapping properties onto numbers in such a way as to preserve some of the characteristics of the set of numbers. Just what characteristics are preserved will be a function of the rules underlying a given operational procedure. These characteristics, in turn, play a central role in determining how such measurements are used in subsequent stages of the research enterprise.

The present study utilizes two rather different sets of operational procedures for measurement of the structural and behavioral concepts relevant to the multiple systems hypothesis. This chapter addresses only those issues pertaining to the operationalization of international positional structure. A corresponding discussion of the observation and measurement of foreign policy behavior will be present in Chapter IV. It should be noted that this separation is for expositional purposes only and -- as will be made clear throughout this chapter and
and the next -- we fully recognize that certain percepts of scientific operationalization apply irrespective of the definition or substantive domain of individual concepts.

The first section of this chapter offers a general measurement strategy that includes a set of seven criteria for sound operationalization. Two of the seven -- reliability and validity -- have been widely discussed in the research methodology literature and consequently need not be considered in great detail. The remaining five criteria are relevance, sensitivity, comparability, availability and simplicity. Section two comprises the main part of this chapter. Here we present a variety of plausible indicators of international structure and engage in a preliminary evaluation of each according to the seven criteria specified above. A final evaluation of the indicators based on their convergent and discriminant validity is deferred to section three.

Measurement Strategy

As we noted in the opening paragraph of this chapter, measurement involves a functional mapping of objects onto numbers according to explicit rules. If the numbers retain some property of the set of numbers, such as its monotonic ordering principle, then we have made a measurement of the object. If, on the other hand, the numbers fail to retain even their order then the numbers are actually numerals and we no longer speak of measurement but of
classification. The present chapter concerns only measurement; we will take up the additional matter of classification in the following chapter on foreign policy behavior.

In Chapter I we argued for a hierarchical interpretation of international structure. Keeping this interpretation in mind, our goal in the present chapter is to arrive at acceptable measurements of structure, first, under the assumption that there exists but a single comprehensive international system and, second, under the assumption that separate and distinct systems surround the military and trade issues described in Chapter II. In either case, structural position is considered an attribute of a national government. At minimum, the measurement of hierarchical structure requires that each government be assigned a number that is indicative of its position vis-à-vis other governments. We will consider the matter of what constitutes an acceptable measurement in a moment, but first let us briefly review a few general issues pertaining to the theory of measurement.

Przeworski and Teune (1970) point out that measurement procedures may be either direct or inferred. In the former case properties are defined by the operations used to measure them. For example, a country's population is normally defined as the number of people counted as citizens during the most recent census. Inferred measurement is more problematical because it is based on the investigator's inference that direct measurement of one or more properties can serve reasonably
well as a surrogate measurement of another property. The physical quality of life index recently adopted by the Overseas Development Council (Sewell, 1977) is an example of inferred measurement based on a combination of literacy, life expectancy, and infant mortality. Both procedures are susceptible to various sorts of measurement error but only when inferences are made is there the additional problem of validity. The indicators of international structure examined in this chapter are all based on inferred measurement procedures.

It has long been known that no measurement technique, however straightforward or direct, is entirely error free. This simple truth has spawned a specialized literature, sometimes termed reliability theory, concerned with the estimation and reduction of measurement error. The fundamental premise of reliability theory states that any score assigned to an object as the result of some measurement procedure will be a function of the object's "true" score and an increment or decrement arising from one or more sources of error (Kerlinger, 1963: 433). What sources of error are most likely to impinge upon our measurements of international positional structure? The data used throughout this chapter are of the aggregate type usually self-reported by national governments to an international organ, such as the United Nations Statistical Office, or published in individual governmental reports. Rummel (1972: 143) has observed
that data of this type are particularly vulnerable to three main sources of error termed clerical, definitional and methodological:

Clerical error results from mistakes in recording and punching data, or in any one of the processing steps from their first enumeration to the table in which they finally reside. Definitional error, or the problem of non-comparability, results from nations using different or partially overlapping definitions in collecting and publishing what are purportedly data on the same characteristic. One nation, for example, may include fishermen and miners in its data on agricultural workers; another might exclude them. Methodological errors occurs when a nation reports statistics based on poor or inefficient data collection techniques. Such error may result from using only urban areas in acquiring census data, for example, or of basing national death rate statistics on that for the largest city.

It is common practice to distinguish between random and systematic measurement error. In nontechnical terms, random error occurs as a result of transient or chance factors that may differ from one measurement situation to another. Over a series of observations random error can be expected to increase some scores and decrease others in what should appear as a random pattern. Clerical error almost always manifests itself as this random type. Definitional and methodological error, on the other hand, may be either random or systematic. In the latter case error is distributed in a nonrandom pattern and arises from some factor that consistently or systematically interferes with the measurement process. If the error across a set of observations tends to covary with the "true" scores or with some other series of measurements then the error is presumed to be systematic. In general, systematic error is considered the more severe
type because of its unpredictable effects on the results of certain statistical procedures. We will return to the problems associated with measurement error in the next section when we consider specific indices of international structure.

We have already stated that measurement of international structure must proceed on the basis of certain inferences built into the measurement procedure. This is because hierarchical structure is a theoretical construct that has no directly observable counterpart in the world of intergovernmental politics. Reliance on inferred rather than direct measurement adds to the burden already incumbent upon the researcher to explicate and justify his or her operational procedures. The remainder of this section will be devoted to the introduction of seven criteria that together serve as a benchmark of sound measurement. These criteria will be of primary importance when we consider the relative merits of different measurement procedures. The seven criteria are relevance, sensitivity, reliability, validity, comparability, and simplicity.

1. Relevance. A measurement has relevance for a particular observation only if that observation can properly be described in terms of the property being measured. A response to a survey question about a totally unfamiliar issue may result in an irrelevant measurement for that particular respondent. An example more apropos to the present study would be an
entirely autarkic society in the international trade structure. The important point about the relevance criterion is that it directs our attention to the appropriateness of a specific measurement with regard to a specific observation.

2. **Sensitivity.** A measurement's sensitivity refers to its ability to discriminate between observations. If we think of the measurement process as a function mapping a set of observations onto a set of numbers, the sensitivity of the measurement has to do with the number of elements in the range. Sensitivity, however, should not be conceived as an absolute criterion — it must be considered in light of the purposes for which the measurement procedure was devised. For example, as a measurement of hierarchical structure Galtung's crude distinction between center and periphery may be adequately sensitive for some purposes but not for others. For the purposes of this study it would be useful to obtain a measurement sensitive enough to differentiate the structural position of each national government. Even more sensitive would be a measurement that preserves the exact distances between governments.
3. **Reliability.** The reliability of a measurement concerns the extent to which it is affected by the presence of random error. Reliability has to do with the internal consistency or dependability of a measurement procedure. As we noted earlier, there is a considerable body of literature dealing exclusively with measurement reliability; unfortunately, the bulk of this literature pertains to educational or psychological testing and, increasingly, to survey research. The differences between these types of data and the aggregate data used here prevent reliability estimation through any of the standard reliability coefficients. Nevertheless, it should be possible to arrive at subjective estimates of reliability based on the probability of interference by clerical, definitional, and methodological random error. All of our data, for example, are taken from secondary sources. Moreover, many of these secondary sources acquired their data from secondary sources as well. As each new link is added to this chain of sources the probability of clerical error tends to increase.

The most common, and potentially most damaging sort of reliability problem likely to be encountered in the use of aggregate data results from cross-national incomparability. Because comparability is
not merely a problem of what is normally termed reliability, we have elected to treat it as a separate criterion (see below).

4. Validity. To assess the validity of a measurement procedure we must ask if it measures what it purports to measure. Put differently, we want to know if the measured differences between observations reflect "true" differences in the property being measured rather than systematic or random error. Several approaches to validity have been discussed in the methodology literature. In this chapter we shall be concerned with the approaches commonly designated face validity, content validity, and construct validity.4

Clearly, the simplest and most straightforward approach to validity is to ask if a measurement procedure appears reasonable "on its face."

Assessing the face validity of specific scores can provide a useful check for random error. Even more important, however, is the role of face validity in evaluating the soundness of the inferences built into the measurement procedure.

Content validity concerns whether a measurement procedure is sufficiently broad to encompass all
facets of the property being measured. For instance, a measurement of military capability based solely on the number of nuclear warheads in a country's arsenal would display a marked lack of content validity. This sort of validity is particularly important when measurement involves construction of composite indices.

Frequently theoretical constructs are defined, in part, by their relationships to other constructs or directly measurable concepts. The construct validity of a measurement procedure is revealed by empirical examination of these hypothesized relationships. One sort of construct validity, termed nomological validity, assesses the performance of a measured construct within a network of propositions supplied by a multivariate theory. If we consider overall structure and issue-specific structure as simply alternative measurements of the same general construct, then this entire exercise can be conceived as an attempt at nomological validation. Other types of construct validity concern the extent to which different measurements of the same construct produce similar results (convergent validity) and the extent to which a measured construct can be differentiated from
supposedly different constructs (discriminant validity) (Campbell and Fiske, 1959).

5. **Comparability.** In the context of this study we will use the term "comparability" to refer to the cross-national equivalence of our measurement procedures. Problems of cross-national comparisons concern both reliability and validity. Reliability is at issue when different governments employ different definitions or different methodologies in the collection and reporting of aggregate statistics. In recognition of the sometimes extensive cross-national differences in data collection, many United Nations publications and other secondary sources are careful to specify the nature of such differences. Even in measurements containing relatively little definitional and methodological error cross-national comparability may be problematical.

To illustrate this point, let us suppose that we want to measure societal stress from inflationary pressures and that we have available to us reasonably error free measurements of increases in each country's consumer price index. Would inflation rate constitute a valid measurement of inflation induced societal stress? One problem with such a measurement procedure is the questionable presumption that equal levels of
inflation produce equal amounts of stress in all societies. For example, a twenty percent annual inflation rate would seem unbearably high in the United States although it would be quite tolerable in Israel.

6. **Availability.** Whenever possible we would prefer to base our measurement procedures on information that is readily available for the largest number of national governments. Availability, however, must always be balanced against the other criteria. For example, we would never choose an invalid measurement over a valid one just because it is available for more countries.

7. **Simplicity.** Other things being equal, simple measurements are to be preferred over more complex procedures. This is particularly important when one's measurement procedures involve construction of composite indices. Simpler procedures are usually easier to communicate to others, have a smaller likelihood of measurement error, and are more accessible to efforts at face validation.

These seven criteria reflect various properties of an ideal measurement procedure. It will be useful to conceive of these properties as a set of continua rather than in terms of their
presence or absence. In this way the several measurement procedures to be examined below can be evaluated against one another as well as against our fictitious ideal. One final point that deserves brief comment concerns the existence of priority relations among the criteria. Although no strict priority applies, the least critical of the seven would seem to be simplicity and availability. On the other hand, if we were to select the most important criterion it would undoubtedly be validity. This is because a measurement shown to be valid must, by definition, also be reliable and comparable.

Indicators of International Structure

In this section we will present twelve candidate measures of international positional structure. It is worth repeating that positional structure is a construct that is not susceptible to direct measurement. Consequently, each of the twelve entails an inference based on some general proposition or law. Presupposition of such laws is an integral part of inferred measurement procedures because laws, along with the rules of logic, allow us to draw an inference about a construct on the basis of one or more direct measurements (Przeworski and Teune, 1970; Kaplan, 1963). Occasionally a law may be purely logical, as when density is calculated from direct measurements of mass and volume. In the present case the situation is not quite so simple. The putative laws necessary for measurement of positional structure must specify the empirical connection(s) between
the directly measured object(s) and the structure construct. Each of the candidate measurements discussed below presumes a one-to-one correspondence between the monotonic ordering of directly measured objects and the monotonic ordering on the construct.

There are two important features common to all twelve measurements. First, all are based on (or around) the year 1965. Two reasons governed this choice: (1) 1965 lies close to the middle of the time span of our behavioral data, and (2) it was a year for which data were most readily available. Occasional deviations from this base year occur for specific countries when 1965 data were unavailable. In such cases available data from the year closest to 1965 were substituted. A few of the indicators presented below utilize measurements that normally display considerable fluctuation from year to year (for example, exports in specific commodities or the number of politically motivated strikes). In order to increase the probability that such measurements were representative of a longer time span, we used a three year average centered on 1965. The second common feature that deserves comment concerns the units for which measurements were made. Ideally, we should measure the structural position of all existing national governments. In order to approximate this ideal within the limits of available data we chose to follow the rules of inclusion devised by Taylor and Hudson (1972). We did introduce two slight modifications, however: our analysis excluded all dependent territories, and 1965 (rather than 1968) was used as the cutoff year.
for national independence. With these modifications we obtained a list consisting of 125 national governments (a reduction of eleven from Taylor and Hudson's total of 136).

Finally, it should be noted that several of the measures are derived from composite indices previously developed by others. In no case did the original author either propose or sanction the use of his index for measurement of international positional structure as that concept is defined in this study.

Overall Structure

We shall begin with the task of measuring overall positional structure -- that is, positional structure under the assumption of a single, comprehensive intergovernmental system. Earlier we stated that the hierarchical arrangement for the single system model is defined by the distribution of national power. This relationship between structural position and power suggests that measurement of the latter would be interchangeable with measurement of the former. At first glance this fact would seem to offer little consolation since power has tended to be such an elusive and evanescent concept. Nevertheless, it is the key to a sizable storehouse of potential measurement procedures. We will consider four in some detail.

One of the simplest and yet most commonly used indicators of national power is gross national product (GNP). Russett (1965: 2), for example, has written that "GNP is the best summary measure of national power that we have." GNP is usually considered an index
of the size of a nation's economy and as such would seem to suffer
the disadvantage of being sensitive to only one of the factors normally
thought to comprise overall power. However, after an extensive
analysis of the determinants of power, Organski (1968) concluded that
of the three most important — population, political development, and
economic development — two are adequately represented in GNP. He
explains his reasoning:

The more efficiently the national economy is organized (that
is, the more "developed" or industrial the nation is), the
greater the amount of goods and services produced, that is,
the higher the GNP. The size of the GNP will also be determined
in part by the size of the nation's population, quite
apart from the level of efficiency. Even an economically
backward nation can produce a sizable GNP if it has an
immense population. On the other hand, a relatively
small nation can produce a high GNP if its economy is highly
developed. These two factors operate independently, and
both are reflected in the final figure of GNP (Organski,
1968: 209).

Furthermore, Organski (1968: 209) maintains that GNP is merely an
indicant of power since he is "...interested in the GNP, not because
the goods and services it represents contribute to power directly,
but because the GNP is determined by so many of the same factors
that determine national power."

How does the GNP index fare against the seven criteria listed in
the preceding section? In fact, it appears to be quite acceptable
in terms of relevance, sensitivity, availability, and simplicity.
It might be argued that GNP fails to be wholly relevant for many of
the poorest countries because a significant portion of their total
goods and services are produced outside of a monetary economy. Yet
in this respect GNP is probably more relevant as an indicator of power than of economic capacity since it is doubtful that a country's non-monetary product reflects power with the same fidelity as its monetary product. We should also point out that GNP is not particularly simple in any absolute sense; however, in comparison to other indices it is the simplest. The chief limitations of GNP as an index of national power concern its reliability, validity, and comparability. Let us first discuss a number of technical issues that affect the comparability and reliability of the GNP index.

Gross national product, like other national accounts statistics, is notorious for its susceptibility to definitional and methodological error. These sources of error can be expected to diminish in importance as more and more countries adopt international standards for keeping national accounts. Moreover, Russett and his associates (1964) have noted that errors in different sectors of the economy will normally cancel one another such that the resultant quantity will be more accurate than some of its components. Nonetheless, these types of errors are troublesome because they are usually related to a country's level of development, thus making them systematic rather than random. Another source of error is introduced when GNP is converted from individual national currencies to common units (in most cases to U.S. dollars). Conversion to a common currency is less desirable than conversion to some sort of "standardized" unit that takes account of cross-national differences in price structures
(see Russett, 1965: 3). Unfortunately, such standardization has yet to be worked out for most countries so it is necessary to utilize common currencies for purposes of comparison. The problem with this method is that the exchange rates used for international trade and currency exchange are often not appropriate for conversion of domestic transactions. Finally, we should mention that comparison of centrally planned and market economics is made particularly difficult by differences in pricing practices and accounting methods. For example, most socialist countries keep account of gross material product, an index that must be adjusted to reflect service production before it is comparable to GNP. Despite these many difficulties several analysts would concur with Organski (1968: 215) that GNP still provides "...a good indication of relative rank in power" (also see Russett et. al., 1964: 151).

Before we can agree with Organski's assessment we must consider the GNP index in terms of its validity. Perhaps the most obvious criticism is directed against the content validity of GNP as an index of power. One might argue, for example, that GNP fails to adequately tap the military component of national power. Yet if we conceive of GNP as indicative of total productive capacity it would seem reasonable to suppose that much of that capacity could be shifted to the military sector if necessary (see Hitch and McKean, 1961). A more serious omission, and one that has been mentioned by both Russett and Organski, is its failure to take into account a nation's political sector. Russett (1965: 3) notes that GNP
"...cannot indicate differences in the internal distribution of resources, or in the tightness of domestic political control." The implication, of course, is that certain totalitarian systems may provide their governments' with more power than their GNP's would indicate. Organski (1968) on the other hand, is more concerned with political development than tightness of control. He has argued that efficient and modern political organization stands second only to population in its importance as a determinant of national power. Although he concludes that GNP alone is not the best measure of power that can be conceived, it nevertheless will do reasonably well until it can be used in conjunction with a measurement of political development.

We have seen that the GNP index of power is a measurement procedure laden with difficulties. Yet we must remember that power and structure are constructs that are elusive of tidy measurement. Although it is tempting to discard the GNP index as clearly unreliable and of questionable validity, at this stage in our discipline it would seem the wiser course to retain it as one provisional approximation of national power. Accordingly, we have acquired estimates of GNP in millions of United States dollars for 124 countries. In the next section we will investigate the convergent and discriminant validity of the GNP index.

Throughout the vast literature on national power the one injunction heard above all others is to eschew what Morgenthau (1973)
called "the fallacy of the single factor." Recall that in his advocacy of the GNP index Organski hastened to point out that population and economic development were dual components of gross product. Other authors have gone to far greater lengths to avoid the pitfalls of a single factor evaluation of national power. An early and frequently cited example is German's (1960) attempted synthesis of twenty selected ingredients of power. German's study is remarkable on several counts, one of which is certainly his broad coverage of factors representing the four basic dimensions of land, population, military capability, and economic capacity. Moreover, because he sought to be as exhaustive as possible he combined what we sometimes (rather wishfully) call "hard data" with his own admittedly subjective judgements regarding such factors as popular morale and technical education. Perhaps the most innovative, yet at the same time most frustrating aspect of this study has to do with the method by which the disparate factors are combined into an overall power score. German (1960: 138) considered it "...necessary to combine and compare different qualities by applying selective and subjective criteria in order to reach the desired over-all impression." To accomplish this he introduced a series of adjustments and weights that, unfortunately, appear totally arbitrary and, hence, defy replication.

A more recent measurement of national power that shares many of the features of German's work appears in a book-length study by Steven Spiegel (1972). We find Spiegel's work particularly attractive
because his objective, mapping the various levels of international hierarchy, closely parallels the central aim of this chapter. "When analyzing international hierarchy," he notes, "our emphasis will be on the inequality of nation-states and on the structure, or the distribution of power, of the current international system..." (Spiegel, 1972: 90). The product of Spiegel's efforts is a seven-tiered scheme indicating the relative position of the world's states and major territories as of 1970. Let us consider Spiegel's measurement procedures in closer detail.

Spiegel's (1972: 40) analysis begins with a conventional definition of power as "...the present and potential ability and willingness on the part of one government to actively affect internal decision-making processes in other countries." The factors of power divide into five broad categories. Three of these -- material, military, and motivational -- are considered essential elements in any evaluation of national power; the remaining two -- achievement and potential -- are used to further refine and broaden the assessment. Two aspects of Spiegel's conception of power are critical to an understanding of the rationale underpinning his measurement procedures. The first concerns the measurability of the various components of power. He maintains that "...capacity, force, and aspects of potential are partially measurable; will and achievement largely are not" (Spiegel, 1972: 41). The second pertains to the factors in combination: "The five components of power are useful tools for
assessing the position of states, but they do not have an existence of their own and, as we shall make clear, they cannot simply be added to produce some combined sum" (Spiegel, 1972: 42). In accordance with these two aspects of power, Spiegel devised a measurement strategy that combined judgements of governments' motivational power and past achievements with numerous quantitative indicators of such factors as population, GNP, energy consumption, defense spending, and so forth. Moreover, he utilized an undoubtedly complex -- though unreported -- subjective weighting scheme that differentiated the various elements according to their importance in determining the power of specific governments.

When examined against our seven criteria, the main advantages of Spiegel's index appear to be its relevance, comparability, and validity. Presumably, the differential weighting scheme insures that the measure is cross-nationally comparable and relevant for each government. In addition, the extraordinarily broad coverage of power components and the careful consideration given to their combination contribute to the content and face validity of the index. The primary weaknesses of this measurement procedure are its insensitivity and doubtful reliability. As we mentioned earlier, Spiegel's index distinguishes only seven levels in the international hierarchy. Obviously, we would prefer a more finely calibrated measure for our purposes; note, however, that Spiegel's index is suitably sensitive for the purposes of his analysis. We might also point out that
Spiegel's choice of seven levels is not based on any "in principle" argument against finer distinctions: "The seven categories should be viewed as delineating the most obvious differentiations on a continuum, rather than as a self-contained grouping of nations" (Spiegel, 1972: 97).

Clearly, the most severe problem with Spiegel's measurement procedure concerns its reliability. A hallmark of reliable measurement is a property sometimes termed "equivalence," that is, the extent to which different investigators using the same procedures can achieve the same results (Selltiz, Wrightsman, and Cook, 1976). There is a definite lack of equivalence in Spiegel's procedure stemming from his failure to apply systematic criteria to his subjective judgements and weighting procedures. Be assured that we are not objecting to the use of subjective judgements per se, only to the lack of systematicness. Subjectivity influences all measurement procedures at some stage of their development, whether it be in the choice of indicators or in the assignment of actual scores. Kaplan (1963: 212) is most instructive on this point:

All measurement yields, not a property intrinsic to the object being measured taken in isolation, but a relation between that object and the others serving as standards of measurement. When the relation is to other human beings, or even to the observer himself, it is not therefore a subjective one. As always, everything hinges on the controls which can be instituted, and on the sensitivity and reliability with which the discriminating judgements are being made.
Despite its problems, Spiegel's index constitutes a legitimate attempt to measure the international distribution of power. Before it can be used in our study, however, we must reconcile the five year difference between his time period and our own. Actually, it does not seem unreasonable to simply appropriate his index as it stands under the assumption that any changes occurring between 1965 and 1970 were not of sufficient magnitude to involve changes in the levels assigned to states. We recorded the Spiegel index by assigning the states at each consecutive level a ranking from one to six beginning with the lowest level (that is, Microstates = 1, Regional states = 2, etc.). Table 1 displays the distribution of our population of 125 governments across the six hierarchical levels. The levels are labeled according to Spiegel's terminology.

The third index we shall consider is contained in an appendix to Cox and Jacobson's (1974) volume on decision making in international organization. They, too, conceive of power as the product of several component factors, most notably material and psychic resources and the skill and willingness to use them for wielding external influence. Their measurement procedure incorporates five indicators for concepts of these components: GNP, GNP per capita, population, nuclear capability, and prestige. We have already discussed GNP at some length. GNP per capita, which is simply total GNP divided by population, is employed as a rough indicator of a state's ability to mobilize its economic capacity and of the general level of skill (such as
### Table 1

**Spiegel's Index of International Hierarchy**

<table>
<thead>
<tr>
<th>Hierarchical Level</th>
<th>Number of States</th>
<th>Percentage of Total</th>
<th>Cumulative Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary powers</td>
<td>2</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td>Secondary powers</td>
<td>5</td>
<td>4.0</td>
<td>5.6</td>
</tr>
<tr>
<td>Middle powers</td>
<td>17</td>
<td>13.6</td>
<td>19.2</td>
</tr>
<tr>
<td>Minor powers</td>
<td>30</td>
<td>24.0</td>
<td>43.2</td>
</tr>
<tr>
<td>Regional states</td>
<td>24</td>
<td>19.2</td>
<td>62.4</td>
</tr>
<tr>
<td>Microstates</td>
<td>47</td>
<td>37.6</td>
<td>100</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>125</strong></td>
<td><strong>100.0%</strong></td>
<td></td>
</tr>
</tbody>
</table>

Source: Adapted from Spiegel (1972: 93-96).
literacy and education) in the population. The remaining two indicators require more detailed explication because they are derived from four-point judgemental scales. Nuclear capability can range from "no foreseeable nuclear capability" at the lower end to "developed 'second strike' capability." The authors argue that "even if a state does not possess nuclear weapons, its known or reputed ability to have them within a few years if it so decides is also to be regarded as a resource for influence" (Cox and Jacobson, 1974: 438). Following this logic the midpoints of the scale distinguish between those states in "possession of nuclear weapons" and those with the "ability to acquire nuclear weapons by 1980-85." Prestige is based on "...the degree of autonomy or independence of a state's foreign policy, considered as an indicator both of a psychic resource in dealing with other states and of willingness to exert influence in external relations" (Cox and Jacobson, 1974: 438). The low point on the prestige scale is marked by a "nonindependent foreign policy" whereas maximum prestige is accorded to states that are recognized leaders of alliances or groups of states or that maintain "active independence in a hostile environment." Both of these scales, along with those for the other indicators, are listed in Table 2.

Cox and Jacobson have devised a rather simple procedure for amalgamating these five indicators into a single, overall power rating. The initial step called for compressing each of the three quantitative indicators (GNP, GNP per capita, and population) into a scale such that each point represented a specified range on the
### TABLE 2

Scales Used in Cox and Jacobson's Power Index

<table>
<thead>
<tr>
<th>GNP (in billions of dollars)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1  under .9</td>
<td>7  30-39</td>
</tr>
<tr>
<td>2  1-3.9</td>
<td>8  40-59</td>
</tr>
<tr>
<td>3  4-6.9</td>
<td>9  60-99</td>
</tr>
<tr>
<td>4  7-9.9</td>
<td>10 100-199</td>
</tr>
<tr>
<td>5  10-19.9</td>
<td>11 200-499</td>
</tr>
<tr>
<td>6  20-29</td>
<td>12 500 and over</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GNP per capita (in dollars)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1  under 200</td>
<td></td>
</tr>
<tr>
<td>2  200-599</td>
<td></td>
</tr>
<tr>
<td>3  600-999</td>
<td></td>
</tr>
<tr>
<td>4  1000 and over</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Population (in millions)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0  Under 2</td>
<td></td>
</tr>
<tr>
<td>1  2-19</td>
<td></td>
</tr>
<tr>
<td>2  20-59</td>
<td></td>
</tr>
<tr>
<td>3  60-99</td>
<td></td>
</tr>
<tr>
<td>4  100-249</td>
<td></td>
</tr>
<tr>
<td>5  250 and over</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nuclear capability</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0  no foreseeable nuclear capability</td>
<td></td>
</tr>
<tr>
<td>1  ability to acquire nuclear weapons by 1980-85</td>
<td></td>
</tr>
<tr>
<td>2  possession of nuclear weapons</td>
<td></td>
</tr>
<tr>
<td>3  developed &quot;second strike&quot; capability</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prestige</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0  nonindependent foreign policy</td>
<td></td>
</tr>
<tr>
<td>1  alliance - aligned</td>
<td></td>
</tr>
<tr>
<td>2  neutral or nonaligned, independent foreign policy</td>
<td></td>
</tr>
<tr>
<td>3  leader of alliance system or recognized leadership of a group of states or active independence in a hostile environment</td>
<td></td>
</tr>
</tbody>
</table>

original indicator. Each country was then scored on all five scales and the resulting number of points were summed to produce an overall score. Because the individual scale scores play such a critical role in this measurement procedure, we have reproduced each of the five scales in Table 2. It should be emphasized that Cox and Jacobson view the overall ratings as indicative of only a rank order and not proportional differences.

Provided one is able to acquire ratings on the two judgemental scales, the Cox and Jacobson index is readily available for most countries and it relies on a reasonably simple method for combining scores. There is a potential problem of relevance associated with using nuclear capability as a contributing factor, but the authors' creative scaling of this property appears to have rendered it relevant for all countries. Although Cox and Jacobson's procedure is clearly ahead of Spiegel's in terms of sensitivity, it is not able to deliver a strong (untied) ranking of governments. In principle, the index ranges from 2 to 27; however, when applied to our population of 125 governments we find an empirical range of 4 to 26 with most of the ties occurring at the lower end of the scale (see Table 3). The highest rating, given to the United States, was one point under the maximum on the population factor. The lower limit of 4 is easily explained by the fact that the smallest and poorest countries (those scoring lowest on GNP, GNP per capita, and population) tend to have nonaligned foreign policies. Another
TABLE 3

Cox and Jacobson's Power Index

<table>
<thead>
<tr>
<th>Composite Score</th>
<th>Number of States</th>
<th>Percentage of Total</th>
<th>Cumulative Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>1</td>
<td>.8</td>
<td>.8</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>.8</td>
<td>1.6</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>.8</td>
<td>2.4</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>2.4</td>
<td>4.8</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>.8</td>
<td>5.6</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>.8</td>
<td>6.4</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>1.6</td>
<td>8.1</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>.8</td>
<td>8.9</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>5.6</td>
<td>14.5</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>4.8</td>
<td>19.4</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>5.6</td>
<td>25.0</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>3.2</td>
<td>28.2</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>6.5</td>
<td>34.7</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>4.0</td>
<td>38.7</td>
</tr>
<tr>
<td>7</td>
<td>19</td>
<td>15.3</td>
<td>54.0</td>
</tr>
<tr>
<td>6</td>
<td>17</td>
<td>13.7</td>
<td>67.7</td>
</tr>
<tr>
<td>5</td>
<td>34</td>
<td>27.4</td>
<td>95.2</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>4.8</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>124</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Note: Data are missing for the Maldives Islands.
positive aspect of the Cox and Jacobson index is its reduced vulnerability to both the random and systematic error that normally intrudes upon aggregate statistics. The assumption here is that error, from whatever source, will generally not be so great as to shift a country from one scale point to another. One final point deserving attention has to do with the reliability of their judgmental scales. In contrast to Spiegel's approach, Cox and Jacobson attempted to insure the replicability of their judgements by explicitly specifying their scoring criteria. Although they did not succeed in removing all ambiguity from the two scales, their method represents a vast improvement over Spiegel's.

The major weaknesses of the Cox and Jacobson procedure can all be discussed under the rubric of validity. Let us begin by examining the validity of the component scales. Perhaps the most controversial aspect of the scales concerns the seemingly arbitrary location of the dividing points for the three quantitative measures. Although arguments could be made for alternative scales, we can think of no definitive criteria that would yield the "best" scale divisions. Because creation of scales such as these is bound to be somewhat arbitrary, in our view Cox and Jacobson's solution is neither better nor worse than alternative methods. A more serious problem concerns their scaling of prestige. The authors admit that their measure lacks sophistication; we would add that it seems to lack validity, as well. For one thing, there is much more to prestige than autonomy, the property their scale actually measures. Moreover,
it is not at all clear that alliance membership necessarily indicates less autonomy than nonalignment. Are there validity problems with the overall scale? Although Cox and Jacobson suggest that in combination the five scales result in relatively broad coverage of the components of power, it is worth pointing out that there is a considerable amount of redundancy in their use of GNP and population and GNP per capita. By using all three measures they are, in effect, introducing an implicit variable weighting scheme into their procedure. We are not arguing against the inclusion of variable weights, only against the failure to acknowledge them when they are included.12

We computed the Cox and Jacobson index for our 125 governments using 1965 statistics for GNP, GNP per capita, and population. The fact that the scale divisions for GNP and GNP per capita were originally developed for 1965 prices greatly facilitated this task. Where possible, nuclear capability and prestige were scored using Cox and Jacobson's own judgements for 1967. Since their scores were available for only thirty-nine countries it was necessary to supplement them with our own estimates for the remaining eighty-six.

Our final indicator of overall structure is considerably more complex, both conceptually and operationally, than any we have examined thus far. It is based on the notion of a government's "capacity to act," a construct developed and operationalized by Maurice East (1975; 1978a). Conceptually, capacity to act refers to "...the resources available for national use in foreign affairs and
the ability to utilize these resources in the service of various foreign policy goals and objectives" (East, 1978a: 134). On this conceptual level, capacity to act can be distinguished from power if we take the latter to mean an ability to exert influence over others. Practically, however, the two concepts overlap a great deal since exerting influence often will require utilizing available resources. The important point is that capacity to act emphasizes control over resources, defined broadly to encompass both human and nonhuman resources, rather than control over others' actions or over the outcome of events. For this reason the construct fits very nicely into the positional structure orientation of this study.

Measurement of capacity to act requires making separate measurements of its three main components. East (1978a:136) has summarized this decomposition:

...a nation's capacity to act is a function of two general factors, size and level of social organization. Size taps the total resources potentially available in a nation. Social organization taps the nation's ability to control, convert, and allocate resources for use in foreign affairs. It is comprised of two dimensions: (1) modernization, which relates positively to social organization and (2) societal stress, which relates negatively to social organization.

Independent measurements of size, modernization, and societal stress are carried out using a weighted scoring procedure derived from separate principal components analyses applied to several raw indicators of each concept. For instance, the indicators chosen to represent the size component include total population, total area and agricultural area, annual energy consumption, gross national
product, government expenditures for defense, and size of the armed forces. Principal components analysis applied to a data matrix arranged from a set of observations on these indicators has the effect of decomposing the matrix by summarizing most of its variation in fewer variables, or "components." The technique is similar to common factor analysis but does not depend upon the causal structure hypothesized by the latter; principal components are revealed as exact mathematical representations of linear combinations of observed data (Kim and Mueller, 1979). Moreover, principal components are uniquely defined such that the first component summarizes a maximum proportion of the total variation, the second component summarizes the next largest proportion, and so forth. To return to our example, East found that the matrix of seven size indicators could be adequately summarized by one principal component. The seven indicators were then replaced by a single component score obtained by combining the scores on all seven indicators weighted in proportion to their loading (correlation) on the principal component. This component score is then treated as a summary measurement of size. Modernization and societal stress were measured in identical fashion, except that for the latter two, principal components were required to achieve an adequate summary of the original indicators.

After having obtained summary measurements for size, modernization, and societal stress, the next stage of the procedure involves the successive combination of component scores until reaching a final,
overall measurement of capacity to act. The first step in this process requires the simple summation of the two stress scores to obtain a single, comprehensive stress index. Because component scores are standardized to have zero means and unit variances, the two stress scores are adjusted so that the minimum score is set equal to zero.\textsuperscript{15} The next step is to arrive at a summary measure of social organization, which, it will be recalled, is a function of both modernization and societal stress. Because modernization and stress work at cross purposes in their effect on social organization, the former contributing positively and the latter negatively, it is necessary to subtract the total stress score from the modernization score. As above, these scores are adjusted so that their minimum values equal zero. The final step in the process involves the summation of size and social organization to yield an overall capacity to act score.\textsuperscript{16} Note that at each successive step the separate scores are given equal weight in determining the value of the combined measures. East conceded that separate scores need not contribute equally to the combined measures; however, he argues that differential weights could not be introduced without some theoretical or empirical justification for doing so. The weighting issue is important and we will want to explore it further in a moment; but, first, let us consider East's construct and measurement procedures against our seven criteria.

We must admit that the complex nature of East's work makes it difficult to know where to properly begin an evaluation. Should our comments be directed at the construct or its operationalization?
Should we consider the raw indicators, the component scores, or the final capacity to act measure? What about the technical aspects of the principal components analysis? We will want to touch upon all of these matters at some point in our discussion, but let us begin with the last since, in our view, that will lead us directly to the most problematical aspect of East's procedures.

The first question we must address is a fundamental one: Is principal components analysis the most appropriate method for obtaining summary measurements of size, modernization, and societal stress? Certainly the principal components technique is suitable for this purpose; whether or not it is better than alternative methods depends, in part, on the conceptualization of capacity to act and its constituent elements. Common factor analysis, the main alternative to the principal components method, presupposes an underlying causal structure that explains covariation among a set of variables in terms of a smaller number of factors. Although he does not say so explicitly, there is a reason to believe that East has just this sort of hypothetical model in mind. For example, the common factor model is suggested by his repeated assertions that certain measures "tap the dimensions" of size, modernization, and stress. Another feature of factor analysis that is particularly advantageous in conjunction with the use of cross-national aggregate statistics is the technique's ability to accommodate certain amounts of measurement error; principal components analysis, on the other hand, simply incorporates this error into its summarization of the data matrix.
East ultimately chose to rely on the principal components method because it permits derivation of unique and precise component scores, in contrast to factor analysis which, at best, can yield only estimates of factor scores. If we are correct about East's theoretical assumptions then common factor analysis would be the more appropriate data reduction technique, its indeterminant scoring procedure notwithstanding. However, because we cannot be certain about East's assumptions and because there are not likely to be major differences between the two approaches for these data, we can accept the principal components method with only minor reservations.

Our acceptance of East's decision to employ principal components analysis leads to a second question: Has the technique been appropriately applied? To answer this question we will need to consider some of the more technical aspects of the principal components method. In the first place we should note that although principal components are interpretable as streams of variation running through a data matrix, the actual extraction of components is performed not on the raw data matrix itself but on a dependence matrix containing either covariances or correlations. Because of its standardization property, a product-moment correlation matrix is appropriate when incommensurable units are used to record the raw measures. For example, the size indicators listed above encompass measurements made in terms of people (population), square kilometers (area), dollars (GNP), and so forth. The point to be emphasized, however,
is that the product-moment coefficient entails certain assumptions about the structure of the raw data matrix, assumptions which, if violated too severely, could have serious implications for the principal components procedure. We refer, of course, to the requirements of linearity and approximate bivariate normality. These assumptions are quite well known; nevertheless, they deserve our close attention since, as we will argue below, they play a major role in our evaluation of East's operationalization of capacity to act.

The product-moment coefficient provides a reasonably accurate representation of linear dependence when the variables are distributed in an approximately normal fashion. However, when the relationship is a nonlinear one, or when the distributions are highly skewed or contain extreme data points the product-moment coefficient can be very misleading. Obviously, it would be ill-advised to place much confidence in the principal components extracted from a matrix of such coefficients. How does all of this relate to East's use of the principal components technique? Time and again it has been demonstrated that cross-national aggregate data in general, and the indicators used by East in particular, are far from being normally distributed about some mean (see Rummel, 1972; Russett et. al., 1964; Taylor and Hudson, 1972). These data commonly assume a right or left-skewed exponential (J-shaped) distribution with most of the observations bunched toward one extreme. Under such circumstances the product-moment coefficient may conceal important relationships involving the bunched end of the
distribution because it is unduly influenced by the presence of a few very extreme outliers at the other end.\textsuperscript{18}

One method for dealing with this problem is to apply some rank invariant transformation to a variable in order to "normalize" its distribution. Not only will the successful application of this procedure reduce the effect of extreme outliers, it also increases the likelihood of the relationships being linear (Rummel, 1972). East did not perform any sort of transformation on his data prior to his principal components analyses; had he done so we would not have pursued this line of argument. How would judicious use of this transformation strategy have altered the principal components results? Would any differences be evident in the final capacity to act index? In order to answer these questions we attempted to replicate East's procedures using transformed data where appropriate.\textsuperscript{19}

Before proceeding let us be absolutely clear about one thing: we are not advocating the indiscriminate transformation of all cross-national aggregate indicators. Any transformation capable of altering a variable's distribution necessarily looses a certain amount of information in the process. In general, then, the transformation issue is one of balancing the effect of possibly misleading correlations against the loss of information resulting from the transformation procedure. "The application of a transformation," writes Rummel (1972: 175) "is one of the many areas in which research intuition and scientific method meet." In order to provide some
guidance for this intuitive process we examined three characteristics of each indicator:

1. The most important consideration has to do with the shape of the indicator's distribution since that is where the need for transformation originates. We inspected three features of each distribution: the number of outliers at or more than an absolute value of three standard deviations from the mean, the direction and degree of skewness, and a measure of goodness of fit to normality. Goodness of fit was indicated by the maximum deviation statistic, D, derived from the Kolmogorov-Smirnov one-sample test. D measures the largest absolute deviation between the cumulative frequency distribution of the observed variable and a specified theoretical (in this case normal) cumulative distribution; thus, the larger the D the poorer the fit. Although D has a known sampling distribution that permits computation of significance levels, it is clearly inappropriate to speak of statistical significance with respect to our population of 125 governments since no sampling model is involved. So long as we are mindful of this fact and cautious in our interpretation, the significance level of D can
be useful as a rough guideline that, in Rummel's (1972: 195, note 1) words, "...indicates whether a particular distribution would be significantly normal were it for a sample selected in some appropriate way, such as randomly, from a universe."

2. In order to assess the value of the information lost through the transformation process it is helpful to examine certain properties of the indicator's scale. One of these properties concerns the theoretical range of the scale—that is, whether or not it contains a meaningful upper or lower bound. For example, a measurement expressed as a percentage is normally confined to values between 0 and 100. Although there may be exceptions, as a general practice it seems advisable to avoid transformations that would alter the meaning of fixed scale bounds.

3. Another scale property that deserves close attention is the meaning given to individual scale units and the distances between observations as expressed in those units. All transformations involve changes of scale; the issue, however, is whether or not such changes alter the meaning of the measured concept. In some cases a transformation may actually, improve the scaling of an indicator by making it
more sensitive to conceptually important distinctions. To take a widely used example, an increase in GNP per capita from $500 to $800 is likely to be much more significant than an increase from $2500 to $2800, even though the amount of increase is the same in both cases. A logarithmic or root transformation of GNP per capita would be consistent with this idea, whereas an exponential transformation, such as the square of GNP per capita, would have just the opposite effect. Hence, before a transformation is applied to an indicator we must consider whether the conceptually important distinctions are adequately reflected in the resulting scale.

There are literally an infinite variety of possible scale transformations that will force a distribution closer to the familiar bell shape of the normal distribution, but the one that seems to have gained widest acceptance is the common logarithm. There appear to be two reasons for this: first, the logarithmic scale is familiar and easily understood; and second, log transformations have been found to work reasonably well for the highly right skewed distributions that are so common among cross-national aggregate data. For distributions exhibiting only moderate right skewness or left skewness, Rummel (1972) has suggested use of the square root and square, respectively. For each raw indicator, Table 4 lists the number of
### TABLE 4
Transformations of Capacity to Act Indicators

<table>
<thead>
<tr>
<th>Indicator</th>
<th>N</th>
<th>Raw Data</th>
<th>Transformation</th>
<th>Transformed Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Number of Outliers</td>
<td>Skewness</td>
<td>K-S Dc</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total population</td>
<td>125</td>
<td>1</td>
<td>6.57</td>
<td>.374</td>
</tr>
<tr>
<td>Total area</td>
<td>125</td>
<td>4</td>
<td>5.53</td>
<td>.350</td>
</tr>
<tr>
<td>Agricultural area</td>
<td>119</td>
<td>3</td>
<td>4.86</td>
<td>.355</td>
</tr>
<tr>
<td>Energy consumption</td>
<td>121</td>
<td>2</td>
<td>7.97</td>
<td>.408</td>
</tr>
<tr>
<td>Gross National Product</td>
<td>124</td>
<td>1</td>
<td>8.17</td>
<td>.403</td>
</tr>
<tr>
<td>Defense Expenditures</td>
<td>121</td>
<td>2</td>
<td>8.16</td>
<td>.355</td>
</tr>
<tr>
<td>Military manpower</td>
<td>121</td>
<td>2</td>
<td>5.16</td>
<td>.423</td>
</tr>
<tr>
<td><strong>Modernization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent urban population</td>
<td>125</td>
<td>2</td>
<td>1.73</td>
<td>.141</td>
</tr>
<tr>
<td>Literacy rate</td>
<td>125</td>
<td>0</td>
<td>-0.02</td>
<td>.119</td>
</tr>
<tr>
<td>Students in higher education</td>
<td>123</td>
<td>2</td>
<td>1.98</td>
<td>.205</td>
</tr>
<tr>
<td>Percent GDP for industry</td>
<td>124</td>
<td>1</td>
<td>0.77</td>
<td>.100</td>
</tr>
<tr>
<td>Newspapers per 1000 pop.</td>
<td>116</td>
<td>0</td>
<td>1.41</td>
<td>.221</td>
</tr>
<tr>
<td>GNP per capita</td>
<td>124</td>
<td>1</td>
<td>1.88</td>
<td>.238</td>
</tr>
<tr>
<td>Energy consumption per capita</td>
<td>121</td>
<td>3</td>
<td>2.55</td>
<td>.262</td>
</tr>
<tr>
<td>Proteins per capita per day</td>
<td>125</td>
<td>0</td>
<td>0.22</td>
<td>.107</td>
</tr>
<tr>
<td><strong>Stress</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractionalization index</td>
<td>125</td>
<td>0</td>
<td>0.26</td>
<td>.139</td>
</tr>
<tr>
<td>Ave. inflation rate (64-66)</td>
<td>117</td>
<td>2</td>
<td>10.75</td>
<td>.456</td>
</tr>
<tr>
<td>Ave. political strikes (64-66)</td>
<td>125</td>
<td>1</td>
<td>10.95</td>
<td>.444</td>
</tr>
<tr>
<td>Ave. political deaths (64-66)</td>
<td>125</td>
<td>1</td>
<td>10.66</td>
<td>.480</td>
</tr>
</tbody>
</table>

---

**a**This column indicates the number of outliers before and after transformation. An outlier is defined as any value ±3 standard deviations from the mean.

**b**This column lists the skewness of the distribution before and after transformation. The measure of skewness is defined as $\frac{N}{(N-1)(N-2)} \sum z_i^3$ where $z_i$ is a standardized variable such that $z = (x - \bar{x})/s$.

**c**This column indicates the maximum deviation, D, resulting from the Kolmogorov-Smirnov one-sample test for a normal distribution. D is defined as $\max|F_0(X) - S_N(X)|$ where $F_0(X) = \text{the cumulative normal frequency distribution function}$ and $S_N(X) = \text{the observed cumulative frequency distribution}$.

Assuming an N size of 122 (the average for all 19 indicators), "significant" departure from normality at $p \leq .05$ is indicated by $D > .124$; for departure from normality at $p \leq .01$, $D > .148$ (See Siegel, 1956: 47-51).

**d**This column lists the type of transformation applied to the raw data distributions. All log transformations are to the base 10.

**e**No transformation was applied in the usual sense; however, we did remove two very extreme outliers (Brazil and Indonesia) by changing their scores to an arbitrary value of 100.
nations for which data were available; certain characteristics of
the distribution, both before and after transformation; and the
type of transformation applied, if any. All log transformations are
to the base 10. Because there is no common logarithm of zero, one
is added to the original score when an indicator was known to con­
tain a zero value. Transformations for several indicators were
chosen on the basis of recommendations reported by Rummel (1972).

The results of these transformations are dramatic. Of the
twenty-eight outliers contained in the raw indicators, only twelve
remain after transformation. Furthermore, all of the transformed
indicators display a sharp drop in skewness and, with the exception
of the stress indicators, improvement to within acceptable limits
in their goodness of fit to a normal distribution. In reporting
these encouraging results we must not lose sight of the equally
significant fact that in each case where a transformation was
applied we are satisfied that the conceptually important variation
has been preserved. Note, too, that no transformations were applied
to the several indicators having fixed upper and lower bounds.21

One other matter that deserves brief comment concerns the peculiar
unresponsiveness of the stress indicators. The average inflation
rate (defined as the average yearly increase in the consumer price
index) was not actually transformed although we did alter its distri­
bution by changing the values of two very extreme outliers. Between
1964 and 1966 Brazil's average inflation rate was 114 percent and
Indonesia's was an incredible 3134 percent; both of these figures were changed to an arbitrary value of 100. Because inflation rate was included as an indicator of societal stress (and not of inflation, per se) we concluded that this procedure was a reasonable method for removing extreme outliers without losing these two observations, since once inflation reaches an annual rate of 100 percent it seems unlikely that further increases would involve concomitant increases in stress. We purposely did not impose a logarithmic transformation on this indicator because the change of scale did not seem justified on conceptual grounds. The ineffectiveness of the logarithmic transformation in the case of the two event based indicators is not surprising given the rarity with which such events occur. We applied a log transformation to these indicators primarily to impose a change of scale since we did not want the middle range of values to be outweighed by the few extreme observations.22

The results of the principal components analyses on the transformed indicators are presented in Tables 5, 6, and 7. So that the reader may get a clear picture of how the transformation process affects the component structures, we have included the results of East's analyses in the tables.23 Looking first at Table 5, we see that the most pronounced effects of transforming the size indicators is to enhance the importance of population and to diminish the loadings of the two area measures. In fact, we found that the area indicators were primary contributors to a second component
TABLE 5
Principal Components Analysis of Size

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Raw Data (N = 136)</th>
<th>Transformed Data (N = 125)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population</td>
<td>.627</td>
<td>.906</td>
</tr>
<tr>
<td>Total Area</td>
<td>.819</td>
<td>.617</td>
</tr>
<tr>
<td>Agricultural Area</td>
<td>.868</td>
<td>.715</td>
</tr>
<tr>
<td>Energy Consumption</td>
<td>.932</td>
<td>.885</td>
</tr>
<tr>
<td>Gross National Product</td>
<td>.916</td>
<td>.938</td>
</tr>
<tr>
<td>Defense Expenditures</td>
<td>.929</td>
<td>.937</td>
</tr>
<tr>
<td>Military Manpower</td>
<td>.942</td>
<td>.886</td>
</tr>
</tbody>
</table>

Percentage of total variance 75.4 72.0

\(^a\)These results were reported in the appendix to East (1975). Signs have been reversed to facilitate comparison with results of analysis on the transformed data.

(not reported in Table 5) that accounted for an additional twenty percent of the total variance. East's second component, on the other hand, was most heavily influenced by military expenditures and manpower. In our view, the component structure of the transformed data provides the better indicator of overall size because of the added weight given to population at the expense of area. The modernization analysis is a bit more difficult to assess since the effects of the scale transformations are more evenly distributed among the eight indicators (see Table 6). What is remarkable, however, is that all but two of the indicators exhibit an increase in their component loadings when transformations are introduced. Because of these
TABLE 6  
Principal Components Analysis of Modernization

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Raw Data (^a) (N = 136)</th>
<th>Transformed Data (N = 125)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent urban population(^b)</td>
<td>.670</td>
<td>.730</td>
</tr>
<tr>
<td>Literacy rate</td>
<td>.855</td>
<td>.921</td>
</tr>
<tr>
<td>Students in higher education</td>
<td>.793</td>
<td>.862</td>
</tr>
<tr>
<td>Percent GDP from industry</td>
<td>.721</td>
<td>.702</td>
</tr>
<tr>
<td>Newspapers per 1000 population</td>
<td>.868</td>
<td>.940</td>
</tr>
<tr>
<td>GNP per capita</td>
<td>.894</td>
<td>.946</td>
</tr>
<tr>
<td>Energy consumption per capita</td>
<td>.848</td>
<td>.957</td>
</tr>
<tr>
<td>Proteins per capita per day(^b)</td>
<td>.829</td>
<td>.762</td>
</tr>
<tr>
<td>Percentage of total variance</td>
<td>66.1</td>
<td>73.6</td>
</tr>
</tbody>
</table>

\(^a\) These results were reported in the appendix to East (1975). Signs have been reversed to facilitate comparison with results of analysis on the transformed data.

\(^b\) For a discussion of modifications made prior to analysis on the transformed data, see note 19.
shifts, and because of the concomitant increase in overall variance accounted for, the transformed component structure seems to be a more satisfactory summary index of modernization than the raw data component structure.

The differences in component structures attributable to the transformations performed on the size and modernization indicators are generally rather modest, and in neither instance require significant reinterpretation of the measured construct. This is by no means true of the stress analysis — here the differences are striking (see Table 7). The first thing to notice about Table 7 is that two distinct stress components have been extracted and rotated according to the varimax criterion. East's first component represents inflationary stress and deaths from political violence, whereas the second indicates political strikes and fractionalization. By contrast, analysis of the transformed indicators yields a more general first component reflecting strikes, deaths, and inflation, and a more specialized second component representing only fractionalization.

How are we to account for this discrepancy? We can state with some assurance that East's results are, at least in part, an artifact of the outliers present in the untransformed data. Two examples should suffice to illustrate the dimensions of this problem. Let us first consider the case of Indonesia, which, it will be recalled, suffered massive inflation during the mid 1960's. In the last three months of 1965, the year of East's data, the army began a systematic effort
TABLE 7
Principal Components Analysis of Societal Stress

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Raw Data (N = 136)</th>
<th>Transformed Data (N = 125)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Ethno-linguistic fractionalization</td>
<td>.051</td>
<td>.698</td>
</tr>
<tr>
<td>Average inflation rate (64-66)c</td>
<td>.925</td>
<td>.070</td>
</tr>
<tr>
<td>Average political strikes (64-66)c</td>
<td>-.094</td>
<td>.751</td>
</tr>
<tr>
<td>Average political deaths (64-66)c</td>
<td>.919</td>
<td>.035</td>
</tr>
<tr>
<td>Average growth rated</td>
<td>-.221</td>
<td>-.362</td>
</tr>
<tr>
<td>Percentage of total variance</td>
<td>35.9</td>
<td>23.1</td>
</tr>
</tbody>
</table>

The components shown have been subjected to orthogonal rotation using the varimax criterion.

These are results reported in the appendix to East (1975).

For a discussion of modifications made prior to analysis on the transformed data, see note 19.

This variable was omitted from analysis on the transformed data. See note 19.
to destroy the Indonesian Communist Party, an effort which was to lead to the deaths of close to 50,000 alleged communists before year's end (Weinstein, 1976). As a result, in 1965 Indonesia surpassed all other countries (even South Vietnam) in the number of recorded deaths from political violence (Taylor and Hudson, 1972). Undoubtedly, Indonesia's very extreme position on these two indicators disproportionately influenced the structure of East's first stress component. The circumstances of our second example are much the same, except that we shift our attention to India and its impact on East's second stress component. In 1965 there were 268 politically motivated strikes in India; the next highest country, South Vietnam, had only eight (Taylor and Hudson, 1972). Because India also ranked third on ethno-linguistic fractionalization it is put in a position to inordinately affect the second component's structure. It is possible that other extreme cases have a similarly disproportionate effect on East's components; we have identified only the two most obvious cases.

Our next task involved computing component scores for size, modernization and the two stress dimensions. Because component scores utilize information from all of the variables in the analysis (see note 14), a single missing datum normally will prevent computation of a score for a given observation. In order to augment the number of countries that could be scored, we followed a procedure that permitted scoring if no more than two indicators were missing for size and modernization, and if no more than one was missing for
stress. When a tolerable amount of missing data was encountered scores were computed by substituting the mean for the missing value. We did not use this procedure indiscriminately, however. A visual inspection was made of each raw data matrix to determine if the mean seemed to be a reasonable substitute for the actual value; if it was not, we attempted to estimate the value on the basis of the values for comparable countries; finally, if we were unable to arrive at a reasonable estimate the case was left unscored. Perhaps it bears reiterating that scores are computed using standardized versions of the indicators. The implication, of course, is that substituting the mean effectively eliminates any contribution to the final score, since zero, no matter how heavily it is weighted, remains zero.

At this point we encountered an entirely unanticipated problem. Recall that the next stage in the operational procedure requires adding the two stress scores and then subtracting this sum from modernization to yield a summary measure of social organization. What was unanticipated was the excessive impact of stress in determining the value of social organization. This impact tended to be concealed by the extreme outliers in the component scores, but when these outliers were removed following the scale transformations it became quite evident. The most salient example involved the United States, which ranked first in modernization by a comfortable margin but fell to tenth place in social organization because its
stress level was considerably higher than most other modernized countries. Had the data been left untransformed the few extreme values on the stress variables would have produced inflated means, thereby lessening the United States' distance from those means and, concomitantly, its stress values. Similarly, the fact that the United States lies at such an extreme position on most of the modernization indicators would be reflected in its untransformed modernization score.

It might be thought that one solution to this problem would be to forgo transformation of the data, but this is unacceptable for two reasons. In the first place, we are convinced that scores derived from the transformed indicators constitute more tenable representations of the theoretical constructs, both statistically and conceptually, than scores from untransformed data. Secondly, such an action would be merely cosmetic since it would provide an adjustment for the United States (and perhaps a few other countries) without addressing the real issue: the relative impact of stress and modernization on social organization. A true solution calls for differential weighting of stress and modernization, a strategy that was rejected by East. Through a process of trial and error, a variety of plausible weighting schemes were examined until one was found that seemed to coincide with our intuitive notion of social organization. Let us state most emphatically that the final weighting scheme is offered, and should be regarded, as no more than an experimental first-cut at this vexing issue. Having expressed this caveat, we can now
relate the quite simple details of the scheme: total stress was
divided by five — that is, discounted by eighty percent — before it
was subtracted from modernization. This solution not only elevated
the United States to second place on social organization (behind West
Germany) but also provided more satisfactory results for highly
stressful countries like India, Indonesia, and South Vietnam. Computation
of the overall capacity to act measure may be summarized by

\[
\text{Capacity to act} = \text{Size}^\ast + [\text{Modernization}^\ast - \frac{\text{Stress 1}^\ast + \text{Stress 2}^\ast}{5}]
\]

where the individual factors are expressed as component scores and
the asterisks indicate that the lowest score has been adjusted to
equal zero.

Before concluding this portion of the chapter with a brief dis-
cussion of our seven measurement criteria, we must address one final
question on the topic of data transformation: How has the transformation
process and the consequent introduction of differential weights altered
the measurement of overall capacity to act? In order to compensate
for the added complexity introduced by our revisions, it is incumbent
upon us to identify some nontrivial differences in the results ob-
tained by the two methods. Furthermore, this effort must be confined
to only thirty-six countries, the number for which East's scores
were available. Table 8 presents adjacent lists of these countries,
rank ordered by East's and our revised measurements of capacity to
act. The rank order (Spearman) correlation between the two lists
is .879 indicating that they are similar, but not so similar as to
TABLE 8

Selected Nations Ranked on Capacity to Act

<table>
<thead>
<tr>
<th>East's Procedure&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Revised Procedure&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 United States</td>
<td>1 United States</td>
</tr>
<tr>
<td>2 Soviet Union</td>
<td>2 Soviet Union</td>
</tr>
<tr>
<td>3 China</td>
<td>3 West Germany</td>
</tr>
<tr>
<td>4 Canada</td>
<td>4 France</td>
</tr>
<tr>
<td>5 West Germany</td>
<td>5 Canada</td>
</tr>
<tr>
<td>6 France</td>
<td>6 Japan</td>
</tr>
<tr>
<td>7 India</td>
<td>7 East Germany</td>
</tr>
<tr>
<td>8 New Zealand</td>
<td>8 Italy</td>
</tr>
<tr>
<td>9 Japan</td>
<td>9 China</td>
</tr>
<tr>
<td>10 East Germany</td>
<td>10 Switzerland</td>
</tr>
<tr>
<td>11 Switzerland</td>
<td>11 Spain</td>
</tr>
<tr>
<td>12 Belgium</td>
<td>12 New Zealand</td>
</tr>
<tr>
<td>13 Israel</td>
<td>13 Belgium</td>
</tr>
<tr>
<td>14 Iceland</td>
<td>14 Norway</td>
</tr>
<tr>
<td>15 Norway</td>
<td>15 Yugoslavia</td>
</tr>
<tr>
<td>16 Italy</td>
<td>16 Israel</td>
</tr>
<tr>
<td>17 Venezuela</td>
<td>17 Mexico</td>
</tr>
<tr>
<td>18 Spain</td>
<td>18 Venezuela</td>
</tr>
<tr>
<td>19 Mexico</td>
<td>19 Chile</td>
</tr>
<tr>
<td>20 Yugoslavia</td>
<td>20 Turkey</td>
</tr>
<tr>
<td>21 Cuba</td>
<td>21 Egypt</td>
</tr>
<tr>
<td>22 Turkey</td>
<td>22 Cuba</td>
</tr>
<tr>
<td>23 Lebanon</td>
<td>23 Uruguay</td>
</tr>
<tr>
<td>24 Egypt</td>
<td>24 India</td>
</tr>
<tr>
<td>25 Zambia</td>
<td>25 Philippines</td>
</tr>
<tr>
<td>26 Costa Rica</td>
<td>26 Iceland</td>
</tr>
<tr>
<td>27 Uruguay</td>
<td>27 Lebanon</td>
</tr>
<tr>
<td>28 Philippines</td>
<td>28 Thailand</td>
</tr>
<tr>
<td>29 Thailand</td>
<td>29 Tunisia</td>
</tr>
<tr>
<td>30 Chile</td>
<td>30 Costa Rica</td>
</tr>
<tr>
<td>31 Tunisia</td>
<td>31 Ghana</td>
</tr>
<tr>
<td>32 Ivory Coast</td>
<td>32 Zambia</td>
</tr>
<tr>
<td>33 Kenya</td>
<td>33 Ivory Coast</td>
</tr>
<tr>
<td>34 Guinea</td>
<td>34 Kenya</td>
</tr>
<tr>
<td>35 Uganda</td>
<td>35 Uganda</td>
</tr>
<tr>
<td>36 Ghana</td>
<td>36 Guinea</td>
</tr>
</tbody>
</table>

(rho = .879)

<sup>a</sup>With the exception of the final summation of size and social organization, East's procedure is described in the appendix to East (1975).

<sup>b</sup>The main revisions include scale transformations of selected raw indicators and application of differential weights to discount stress (also see note 19).
exclude important differences. The most significant changes precipitated by our revisions are the gains made by Italy, Spain, and Chile, and the losses experienced by China, India, Iceland, and Zambia. Although we believe that our revisions lead to a more plausible overall ranking, there are some individual ranks that, from an intuitive standpoint, seem a bit out of line. For example, Switzerland and New Zealand may be ranked too close to the top whereas India may be somewhat underrated. Further on in this chapter we will see how our revised measurement of capacity to act accords with the other three indicators of overall structure.

Our evaluation in terms of the seven measurement criteria is rather mixed. The least ambiguous of the criteria are availability and simplicity. Capacity to act is available for 122 countries, twenty-two of which were scored using our modified procedure because of missing information on one or more indicators. Furthermore, it should be abundantly clear from the protracted length of our discussion that this measurement ranks very low on simplicity. Assessment is more difficult when it comes to the sensitivity and reliability of the measurement procedure, since there are both positive and negative aspects relating to each criterion. Because a unique score is assigned to each country, capacity to act is sensitive enough to produce a strongly ordered representation of international positional structure. Note, however, that the additive and subtractive operations require the component scores to be
sensitive to the exact intervals separating countries. There is no mathematical difficulty in meeting this requirement; conceptually, however, the ground seems a bit less sturdy. Similarly, the reliability of the index is open to some question because of the measurement error that is introduced by such extensive use of aggregate statistics. Although the impact of measurement error cannot be eliminated entirely, it is alleviated to some degree by our reliance on multiple indicators of each concept. Moreover, because the measurement procedure is thoroughly specified, the results should be accurately reproduced by other investigators. The one exception would arise when measurement is carried out on different groups of countries. This is because the procedure utilizes standardized scores that are bound to the distributions of the indicators.

Insofar as the comparability and validity criteria are concerned, the most problematical aspects of the measurement procedure reside with the stress components and the indicators comprising them. We already have had occasion to mention some of the difficulties involved in using inflation rate as an inferred measurement of inflationary or economic stress. It was pointed out, for example, that a given level of inflation should not be expected to elicit the same response from different societies. Inflation induced stress is more likely to be a function of the difference between current inflation and an average rate computed for the preceding five or ten years. Another comparability problem arises from the differences between market and centrally planned economies and their differing
philosophies regarding governmental concern and control over price inflation. Other difficulties with the inflation rate indicator include the considerable definitional error in the specification of consumer price indices and the methodological error surrounding the collection of appropriate statistics. The two event based indicators -- political strikes and deaths from political violence -- have certain comparability problems of their own. Perhaps the most obvious is the underreporting of political events for countries in which the press is strongly influenced or controlled by the government. A thorough discussion of the potential difficulties generally associated with events data will be deferred to the next chapter. Ethno-linguistic fractionalization, the final stress indicator, is particularly important because it so heavily dominates the second stress component. The fractionalization index is expressed as a probability between zero and one that two randomly selected persons will not be members of the same language or ethnic group. Fractionalization thus measures only the potential for societal stress, not its actual occurrence or expression. Moreover, it is as oblivious to the persistence and depth of antagonism between groups as it is to successful political accommodation (as in Lebanon during the 1960's).

These difficulties notwithstanding, we elected to retain capacity to act as one of our provisional indicators of overall structure. The measurement procedure is anchored to a firm conceptual
foundation that accords reasonably well with our notion of positional structure and it enjoys unusually wide coverage of the elements contributing to the size and modernization. The weakest link in the procedure -- the measurement of societal stress -- is, from another standpoint, also one of its principal assets: capacity to act is the only one of our overall measures that essays to tap the potential dampening effect of a strained political or economic climate. Before we can make a final judgement regarding the utility of capacity to act and the other candidate measures of overall structural position we must introduce similar measurement procedures for each of the two issue specific systems. Accordingly, the next part of the chapter presents four potential indicators of position in the international trade structure.

**Trade Structure**

A fundamental premise of the positional orientation outlined in Chapter I states that multiple systems entail multiple structures. The implication of this premise is that if world politics is to be decomposed into coterminous issue systems then it must be possible to identify coterminous structural arrangements as well. We have hypothesized that one such system will consist of governmental activities resulting from political level decisions concerning the regulation and promotion of international trade. The positional structure of this trade system, like that of the overall system, is defined by the cross-national distribution of power, in this case
economic power. For the purposes of this study, "economic power" is used exclusively in reference to international trade, although it should be understood that trade is not the only type of economic issue that is conceivable as the basis of an issue specific system. In what follows we will consider several procedures for measuring economic power position in the international trade structure.

Once again we will begin with the simplest of our candidate measurements: an index delineating the relative magnitude of a state's foreign economic transactions. This index reflects Knorr's (1975: 85) observation that "a country accounting for 30 percent of world exports and imports and of world exports of capital and technical assistance, tends to enjoy far greater leverage than a country accounting for only 3 percent." Because we are interested only in transactions involving trade, the index is expressed as a proportional representation of each country's total trade to total world trade, where total trade is defined as the summation of exports and imports. A country's total trade, like its GNP, is a function of economic development and population. A third factor, international economic specialization, also affects the size of foreign trade. Knorr (1975: 84) elucidates these relationships:

The implications of population size and stage of development are obvious. Even though India is a poor country, it has a far larger foreign trade than Switzerland, which is rich; but Switzerland's foreign trade per capita is a multiple of India's. Clearly also, a country's foreign trade will tend to vary with the extent to which it is engaged in the international division of labor. This variable, in turn, is principally the consequence of
trade policy (e.g., free trade versus protectionism),
of endowment with natural and other resources, and
of size of territory...

The extreme simplicity of the proportion of world trade index
allows us to move directly to an evaluation against our seven
measurement criteria. As with the measures of overall structure,
the principal difficulties lie in the areas of comparability and
validity. For example, because we are interested in trade values
rather than quantities of goods, there is the problem of conversion
to a common currency. It should be noted, however, that except for
those countries with widely fluctuating or multiple exchange rates,
currency conversion generally poses fewer difficulties for foreign
trade than for domestic transactions. Another factor diminishing
the comparability of trade values is the difference in pricing
systems of market and centrally planned economics. The chief diffi-
culty here concerns the valuation of trade between two socialist
states since prices, which are usually established in secrecy, may
deviate from world prices by a wide margin (Marer, 1974). Clearly
the most serious comparability problem is one of definitional error
stemming from the use of different trade accounting systems. The
two most commonly used systems, termed general trade and special
trade, differ mainly in their treatment of reexports and imports
into bonded warehouses and free zones.28 Although there are other
sources of error, such as some countries' practice of combining
imports or exports of foreign aid with their trade statistics, most
governments have a long history of collecting trade data for customs purposes. As a result, "'error' is primarily a function of differences in definition rather than of inaccuracies" (Taylor and Hudson, 1972: 349).

What of the validity of this index? Knorr (1975: 79) tells us that "there are two sides to national economic power. One, the active side, is concerned with what a country can do to other countries; the other, the passive side, concerns a country's ability to limit what other countries do to it" (Knorr's emphasis). A country's proportion of world trade reveals something about the active side of power but it obscures the passive side. In fact, a large share of world trade may actually be disadvantageous if trade is also large relative to GNP, since this would tend to increase a country's vulnerability to outside pressure (Knorr, 1975). Another potentially misleading aspect of this index is its inability to delineate a country's importance with respect to the goods it actually does trade. Similarly, the proportion of total world trade assumes a worldwide market, whereas it may be more significant to consider a country's share in the geographically separate markets in which it does trade (Michaely, 1960). Because the remaining trade position indices pay special attention to these problems we shall postpone fuller discussion until later in this section.

The proportion of world trade index was acquired for 123 countries from Banks (1975). The values of the index are expressed
as percentages and range from .001 for the Maldives to 13.115 for the United States. It should be pointed out that the denominator used in computing this index encompasses only trade by independent states and therefore "...falls somewhat short of being a total summation of world trade" (Banks, 1975: 8). Because we are interested in the relative position of nations rather than their exact proportions, this discrepancy is of little consequence.

The next two indicators were developed by Michael Michaely (1960) in an effort to overcome certain weaknesses in the proportion of world trade index. One of these, the "commodity-weighted share," is designed to compensate for the fact that most countries' trade is not evenly distributed among different types of goods. Michaely's (1960: 307) rationale is quite compelling:

Suppose, for instance, that total exports of a given country in a given year amount to 10 per cent of total world exports in that year. This is a very high figure indeed. But suppose, too, that the commodity composition of the country's exports is exactly the same as it is in world exports as a whole, so that each of the country's respective commodity exports constitutes 10 per cent of total world exports of that commodity. And contrast this country with another, a much smaller trading nation, but one which exports only a few goods, in each of which it accounts for a major or very substantial part of world exports. This latter country -- small as it is if judged by the simple ratio of its total exports to world total exports -- is nonetheless very important in the markets for the goods which it does export, much more so than the former, "bigger" country.

The crux of Michaely's argument is that often it is less important to know a country's share of total world trade than to know its
shares inspecific commodities. By weighting each commodity share according to its importance in the country's total trade we can combine them into an overall index -- the commodity-weighted share.

In Michaely's presentation, commodity-weighted shares were computed separately for exports and imports. Although it would be possible to combine exports and imports into a single index, we chose to follow Michaely's example by treating them individually. This seemed the wiser course because most commodities are not both imported and exported by the same country. Furthermore, we limited our consideration to shares of world commodity exports since most countries are more concentrated, and thus more likely to have a controlling edge, in exports than in imports. Symbolically, the commodity-weighted share in world exports of country j is

$$W_j = 100 \sum_i \left( \frac{X_{ij}}{X_i} \right) \left( \frac{X_{ij}}{X_{ij}} \right)$$

where, for a given period of time, $X_{ij}$ is the total exports of country j over all commodities, $X_i$ represents the total world exports of commodity i, and $X_{ij}$ stands for country j's exports of commodity i. Multiplication by 100 simply allows the index to be expressed as a percentage. This index is functionally identical to the "trade position index" used by Barend de Vries (1967) in his study of the export experience of developing countries.

The commodity-weighted share ranges between a maximum value of 100 and a minimum value equal to a country's simple proportion of total world exports. These limits can be reached only under very
special circumstances: The lower limit indicates that a country's commodity exports are distributed in precisely the same proportions as total world commodity exports; in other words, its export trade is entirely unspecialized. In contrast, a value of 100 is reached whenever a country is the sole exporter of any of its export commodities. Three basic factors of a country's trade are reflected in its commodity-weighted share: (1) the overall size of the country's exports; (2) the degree to which its exports are diversified among a variety of commodities; and (3) the total amount of worldwide trade in its export commodities. "The larger the country's exports, the more concentrated their commodity composition, and the smaller the world trade in these commodities in which the country has concentrated, the higher the commodity-weighted share of the country in world exports" (Michaely, 1960: 308).

Before turning to our seven measurement criteria for an evaluation of this index, let us more fully recount the details of its computation. Perhaps the most critical issue affecting the magnitude of the index is the definition and scope of "commodity". Note that the broader the definition of commodity the more difficult it is for a country to assume a position of importance in its export trade; and, conversely, the more narrow the definition, the easier it is to assume such a position. Note, too, that when very fine distinctions are drawn, commodities tend to become close substitutes of one another. Michaely solved this dilemma by adopting a definition based on the 150 "groups" (three digit code) in the Standard International Trade
Classification (SITC). Shortly, after Michaely's work was first published the United Nations Statistical Office expanded the SITC to include separate categories for 1312 items (five digit code) combined progressively into 625 subgroups (four digit code), 177 groups (three digit code), 56 divisions (two digit code), and 10 sections (one digit code) (United Nations, Statistical Office, 1963).

For the present analysis we elected to focus on the division level of the SITC, Revised. Two reasons governed this decision: first, export data classified into divisions were available for many more countries than were data classified at finer levels; and second, the time and effort required to collect and process data at finer levels were prohibitive given the purposes of this study.

The commodity export data were obtained from the 1967 volume of the Yearbook of International Trade Statistics, compiled annually by the United Nations Statistical Office (1969). Where countries reported their export trade according to the original SITC or according to their national commodity classification, the data were reclassified to the division level of the SITC, Revised. Reclassification must be considered only approximate for trade based on national commodity classifications because similar nomenclature may refer to aggregates of differing composition. Unlike total trade, exports disaggregated into specific commodities frequently exhibit wide fluctuations from year to year. To counteract this tendency each country's commodity exports were averaged over a three year period.
(1964-1966). So that the final index would reflect a constant time frame all other magnitudes were averaged in a comparable fashion. Commodity trade reported in national currencies was converted into U.S. dollars using conversion factors obtained from the Yearbook and from Banks (1975). Total world exports in each of the fifty-six commodities were estimated as the summation of commodity exports across all countries. Exact figures for world commodity exports are not readily available due to differences in nomenclature and classificatory schemes employed by reporting countries.\textsuperscript{32} Our estimates of world commodity exports accounted for approximately eight-eight percent of total recorded world exports, although it is unlikely that we have uniformly captured this high a proportion of trade in each commodity.

Using the procedures described above, we calculated commodity-weighted shares in world exports for a total of 108 countries.\textsuperscript{33} Most of the seventeen countries omitted from our calculations either failed to report the commodity composition of their export trade or reported their commodity exports in quantities rather than values. In addition, we excluded any country whose reported commodity exports accounted for less than seventy percent of its total export trade. Empirically, the index ranges from a minimum value of .06 for Malta to a maximum of 26.63 for Cuba. Although no country reached its theoretical minimum, Malta followed by Lebanon, Rwanda and Jordan all came quite close. The countries exhibiting the largest gains
over their simple shares of world exports were Cuba, Venezuela, Maylasia, and Brazil.

How does this index stand up against our seven measurement criteria? We have already seen that there are problems with its availability, and certainly it is not as simple as the proportion of world trade index. On the positive side, it is amply sensitive for our purposes and, except for the definitional error in the classification of commodities, seems to be adequately reliable. Moreover, because it focuses only on those areas in which a country actively participates in world trade it appears to be a more relevant measure than the simple share, particularly for most of the smaller trading countries. Once more, it is the comparability and validity of the index that are most troublesome. The comparability of this index, like that of the proportion of world trade, is hampered by the differences between centrally planned and market economies. In the former, according to Clark and Farlow (1976: 22-23), "the foreign trade plan is a component of the overall economic plan, so that there is a marked tendency to emphasize quantities rather than prices or total value in trade. Thus, most of the Soviet bloc states apparently possess an 'import first' position whose first concern is imports as necessary domestic inputs with only secondary consideration for the possible exports to pay for them." Clearly, the dynamics governing export practices are very different in the two types of economies. The coarseness of our commodity classification scheme gives rise to a more subtle sort of comparability problem to which there is no
easy solution. It is possible (though devilishly difficult to demonstrate empirically) that some countries may be penalized just because their exports happen to fall into unusually broad or inclusive categories. Unfortunately, space does not permit a thorough analysis of this problem with respect to the SITC.

When we consider the matter of validity it must be remembered that the commodity-weighted share is intended as an inferred measurement of a government's economic power position in the world trade structure. This implies that countries ranking high on this index should have a decided advantage in their particular commodity markets over nations with smaller shares. But is this the case? Speaking primarily of developing countries, de Vries (1967: 11) suggests just the reverse:

Differences in the trade positions [i.e., commodity-weighted shares] of countries may help to explain differences in the growth of exports of major commodities, since the size of a country's trade shares may restrain that country from increasing its exports. The freedom of opportunity that a country has in expanding its exports without upsetting market prices varies inversely with its size of the share in the commodity markets. The larger a country's trade share, the greater is the chance that a proportionate increase in exports will upset the prevailing price structure, disturb the workings of an international commodity agreement, or call forth retaliation by competitors.

Of the twenty-nine developing countries included in de Vries' study, those having below average commodity-weighted shares tended to exhibit greater proportional increases in their export earnings over a ten year period than countries with above average shares. This is an intriguing finding that manages to raise some provocative questions about our use of commodity-weighted shares. We do not believe, however, that the validity of this index is undermined altogether.
For one thing, we do not know whether this same phenomenon would be found among more industrialized countries, nor can we infer anything about absolute levels of earnings. But most important for the present context, we cannot be sure how this finding will affect the relationship between a country's commodity-weighted share in world exports and its government's trade-oriented foreign policy behavior.

At this juncture let us leave consideration of the commodity-weighted share to take up the second of Michaely's two measures -- the "geographically-weighted share." This index is responsive to situations in which a country accounts for only a small part of the world's total trade and yet still plays a leading role in the trading patterns of more circumscribed markets. In Michaely's (1960: 312) words, this measure detects "...how important each country is in the trade of those (geographically separate) markets in which it does trade -- rather than in the 'world market' taken as a single unit."

In a manner that is directly analogous to the commodity-weighted measure described above, the geographically-weighted share represents a weighted combination of a country's share in the trade of each of its partner countries. Once again, the weightings indicate the importance of each share in the country's overall trade. The straightforward relationship between the two indices is even more transparent when they are depicted in symbolic form. The geographically-weighted share of exports in world trade, defined for country j, is:

\[ A_{xj} = 100 \sum_{s} \left( \frac{X_{sj}}{M_{s}} \frac{X_{sj}}{\bar{X}_{j}} \right) \]
where \( X_{j} \) again represents the total exports of country \( j \), \( M_{s} \) stands for the total imports of country \( s \), and \( X_{sj} \) refers to exports by country \( j \) to country \( s \).

Given the similarity of the two measures, it seems reasonable to expect that the theoretical limits and the three basic factors that characterize the commodity-weighted share would have parallel counterparts in the index of geographically weighted share. Indeed, this is the case. A country can receive a geographically weighted share of 100, the upper limit, only if it is the sole exporter to any of its partner countries. The lower limit, which is the ratio of the country's overall exports to total world exports, signifies that the exporting country sells its goods to all other countries in proportion to each customer's share of total world imports. Of the three basic factors reflected in the two indices, only the size of a country's exports is common to both. The two other factors affecting a country's geographically-weighted share are: (1) its concentration of export receiving countries and (2) the overall size of its major partners' international trade.

Thus far we have emphasized the conceptual proximity of Michaely's two measures; it is equally important that we mention the point at which they diverge. A central feature of the commodity-weighted share is that it "assumes away" any geographical specialization in a country's export trade; the geographically-weighted share, on the other hand, focuses exclusively on geographical specialization at
the expense of commodity specialization. Thus, it would seem that one measure's principal strength reflects the other's weakness, and vice versa. Of course, it would be possible to combine the advantages of each into a single index sensitive to both geographical and commodity specialization. Michaely suggested that just such an index could be defined as:

\[ I_{ij} = \sum_i \sum_s \left( \frac{X_{isj}}{M_{is}} \cdot \frac{X_{isj}}{M_{sj}} \right) \]

where \( X_{..j} \) represents the total exports of country \( j \), \( M_{is} \) stands for the imports of commodity \( i \) by country \( s \), and \( X_{isj} \) symbolizes the exports of commodity \( i \) by country \( j \) to country \( s \). Unfortunately, the merits of this index are unable to compensate for the overwhelming practical difficulties involved in obtaining and then processing the requisite massive amounts of information on country-to-country commodity trade.

The geographically-weighted share has certain advantages over its commodity-weighted counterpart when examined against our measurement criteria. In the first place, dyadic trade statistics, the information utilized in the calculation of geographically-weighted shares, tends to be available for more countries than comparable statistics on commodity trade. Moreover, the geographical index virtually eliminates any possibility of definitional error in apportioning trade into appropriate categories. Furthermore, because the relevant categories refer to countries rather than ad hoc commodity classes there is little danger of the distorting effects
caused by nonequivalent categories. Finally, we should point out that de Vries' argument regarding the freedom of opportunity to expand exports is not pertinent to the geographically-weighted share. One problem not alleviated by the geographical index is the noncomparability of market and planned economies.

Our main source of data for the index of geographically-weighted share was Gillespie and Zinnes' (n.d.) compilation from the International Monetary Fund's *Direction of Trade (Annual)*. Where necessary we supplemented these data with information extracted from the *Yearbook of International Trade Statistics*. (Incidentally, the *Yearbook* is the source for all intra-Soviet bloc trade data.) From these two sources we had sufficient information to calculate geographically-weighted shares for a total of 120 countries, ranging from a maximum of 33.45 for the United States to .001 for Outer Mongolia. The countries recording the most significant gains over their simple proportion of world exports were the Soviet Union, the United States, and Japan. As with the commodity-weighted index, no country reached its theoretical minimum; the countries coming closest were Singapore, Outer Mongolia, Albania, and Portugal.

One important feature lacking in both of Michaely's measures is an ability to capture what Knorr called the passive side of national economic power. The last measurement procedure to be examined in this part of the chapter represents an attempt to balance the active and passive sides of power in a single index termed the "economic power
ratio." This measure was developed by James Caporaso (1974) as a way to express the importance of foreign trade in "realpolitik terms." Underlying this ratio is a conception of a two-way relationship between a country's influence on other system members and, conversely, their influence on that country. Caporaso (1974: 106) amplifies this relationship:

The influence of country A on the rest of the system is hypothesized as being equal to its share of the total consumption of every other member of the system. Each country consumes a certain volume of goods and services. Most of the goods and services are produced domestically but not all; some have to be imported from foreign countries. At the same time A is dependent on its imports from the rest of the world since these are responsible for part of its consumption.

Symbolically, the economic power ratio of country i (expressed as a percentage) is

\[ 100 \frac{\text{TE}_i / (\text{WC-TC}_i)}{\text{TI}_i / \text{TC}_i} \]

where, using notation different from Michaely's, \( \text{TE}_i \) represents the total exports of country i, WC is total world consumption, \( \text{TC}_i \) stands for the total consumption of country i, and \( \text{TI}_i \) represents country i's total imports.

Several features of this index deserve comment. Note that the economic power ratio is actually a ratio of two ratios. The numerator of the overall ratio expresses the proportion of total world consumption (less consumption by country i) accounted for by exports from country i and the denominator indicates country i's imports as a proportion of its total consumption. Thus, at any
given level of consumption the economic power ratio will increase as exports increase relative to imports. Interpretation of the index is made clearer if we think of it as a ratio of export influence to import dependence. It can be seen at once that a value of one assumes critical importance as a benchmark of international trade dependence — magnitudes less than one reveal that a country depends more on world imports than the world depends on its exports and, contrariwise, magnitudes greater than one indicate that the world depends more on the country's exports than the country depends on world imports. If all countries were equally dependent on one another — that is, interdependent — then each country's economic power ratio would equal one. Furthermore, the theoretical upper limit of the economic power ratio is equal to the summation of the ratio for all countries which, in turn, is a constant value equal to the number of countries analyzed. In the case of a country having no exports whatsoever, the ratio can reach a minimum value of zero, but at the other extreme, complete autarky, the ratio is mathematically undefined.

The economic power ratio does comparatively well in meeting the standards of our measurement criteria. It is available for most countries, moderately simple, and sensitive enough to yield a strongly ordered representation of international trade structure. Because the measure uses total country aggregates (rather than aggregates subdivided by countries or commodity classes) it is probably less susceptible to measurement error than either of
Michaely's weighted shares. The prime weakness of the economic power ratio is its inability to detect situations in which a country holds a strong position only in relation to selected countries or types of commodities. Put differently, the measure assumes away both commodity and geographical specialization. In this respect the economic power ratio is similar to Michaely's indices: each reflects only one of three main aspects of economic power. Further on in this chapter we will examine the covariation among these three measures to provide some indication of the extent to which these conceptually distinct aspects co-occur empirically.

A country's total consumption serves as a vital component of the economic power ratio, yet it is a statistic not normally recorded in national accounts. Following Caporaso's example, we estimated national consumption as gross national product plus total imports. The resulting consumption estimates were then summed over all countries as an approximation of total world consumption. With these consumption figures and each country's imports and exports we were able to calculate economic power ratios for 121 countries for 1965.36 There is a slight upward bias in these ratios stemming from a minor anomaly in our calculation procedure. Our estimate of world consumption is based on the summation of consumption figures for only 121 countries whereas the import and export data are based on trade with the entire world (including dependent territories). More accurate measurement would require basing trade and world consumption data on the same
countries. We should emphasize that the overall effect of this anomaly is slight and the rank order of countries is unaffected altogether. Empirically, the economic power ratio ranges from a minimum of .0003 for Laos to a maximum of 57.27 attained by the United States.

Military Structure

In this part of the chapter our task turns to the measurement of military power position. Our working assumption will be that the distribution of military power defines the structural arrangement underlying an issue system encompassing negotiation and settlement of intergovernmental military security disputes. Care must be taken not to confuse military power with overall national power, a broader term, or the actual use of military force, which is more narrow. In the following, military power is used in a putative sense to refer to military capabilities available to governmental decision makers without resorting to full scale mobilization of their demographic and industrial resources.

The first two measures of military power, size of armed forces and governmental expenditures for defense, were considered primarily on the basis of their simplicity and face validity. Armed forces refer to all military personnel on active duty in regular army, navy, and air forces, and where significant, in paramilitary forces as well. With two exceptions, reserve and militia forces are not included, even though these account for a major part of many nations'
military strength. Noting the vast cross-national differences in reserve training and armament, Wallace (1973) argued that their exclusion would tend to enhance the comparability and reliability of this indicator. The two exceptions are Switzerland, where a national militia exists in lieu of a standing army, and Israel, where reserves comprise a substantial portion of readily available forces (Taylor and Hudson, 1972). "Military expenditures are defined as current and capital expenditures to meet the needs of the armed forces and to cover all expenditures of national defense agencies other than those expenditures used for civilian projects" (Taylor and Hudson, 1972: 18). These data are generally more inclusive than published military budgets since they encompass military related expenditures in such areas as research and development, atomic energy, and paramilitary forces.

Owing to the special attention given these indicators by governmental agencies, research institutes, and numerous individual scholars, data are available for most countries in reasonably comparable form. This is not to say there are no problems with these data, however. Neither measure, for example, provides any indication of a government's ability to control its military establishment even though the relationship between military leadership and those in political control of the government varies enormously from nation to nation. Another factor contributing to military strength but ignored by these indicators is the government's disposition — or better still,
reputed disposition — to resort to military instruments in the pursuit of foreign policy objectives (Knorr, 1975). Other difficulties emerge when the measures are taken singly. Wallace (1973: 36) has pointed out that "the number of military personnel is not necessarily related to military process, as it does not take into account such factors as weapons, material, and training. It thus tends to inflate the military strength of large, poor nations with sizeable but ineffective armies." Presumably, a measure of armed forces expenditures would partially compensate for differences in military effectiveness. Expenditure data, however, are occasionally subject to outright falsification or, as is more often the case, more subtle forms of manipulation, such as hiding costs under nonmilitary budget items. Then too, expenditure data must be converted into common currency units using exchange rates that do not necessarily reflect real purchasing power. In other words, "...with the same number of U.S. dollars, Turkey can buy 'more army' than the U.S." (Wallace, 1973: 37).

Military personnel and expenditure data were available for 121 countries from Taylor and Hudson (1972) who, in turn, acquired their data primarily from United States Government publications. For ten of these 121 countries expenditure data were reported as estimates in the original source. All expenditures have been converted to millions of U.S. dollars.

Our third indicator of military power is one originally proposed by Quincy Wright (1956) over twenty years ago. Although the index
has not gained wide currency, it has been adopted by Rummel (1972; 1977) for use in several of his factor analytic studies. Proceeding under the assumption that manpower and industrialization are "...the main elements in military potential," Wright (1956: 597) measured military strength as the product of population and energy production. The industrial component is represented by the amount of energy actually produced rather than the total energy resource base because the former better reflects the nation's readily available capabilities. Figures for population were taken from Taylor and Hudson (1972) and energy production data, as conventionally measured in metric tons of coal equivalent, are from Banks (1975). According to the United Nations Statistical Office (1974: 48), the original source of the data, the latter are "...based on the production of coal, lignite, crude petroleum, natural gas and hydro and nuclear energy; where peat used as fuel is important, it is included with coal and lignite."

Computation of the index followed the method recommended by Rummel (1979). This involved standardizing each component measure to a zero mean and unit variance, adding twenty to each measure -- presumably to eliminate negative values -- and then taking their product. 39

It is clear that this measurement procedure is amply sensitive for our purposes and reasonably simple to compute. Even so, it is vulnerable to a degree of measurement error that could affect its reliability and comparability. In the first place, because not all countries maintain records of their energy production, these data
were supplemented where possible with estimates compiled by the United Nations Statistical Office. Thus, even though production figures are available for 120 of our countries, some are based on estimates whereas others have been reported by official sources.\textsuperscript{40}

Furthermore, it must be remembered that virtually all population statistics are themselves only estimates and are subject to varying degrees of methodological error. Three sources of error are germane to most demographic estimation procedures. The first has to do with the error associated with the original data base used in the calculation of later estimates. Base data may be obtained from complete or partial censuses, sample surveys, or even conjecture. A second source of error derives from the method of adjusting for population increase since establishment of the data base. Adjustments may be calculated from continuous population registers of births, deaths, and migration, from rates estimated from two or more actual censuses, or from rates inferred from regional estimates. The third source is the length of time elapsed between the date of the adjustment and establishment of the data base. In addition, some governments define their populations so as to exclude jungle tribes, aborigines, nomadic peoples, and refugees. Although such practices must be considered a source of definitional error, their effects on this indicator are fairly innocuous since such groups are unlikely to contribute in any significant way to a nation's military or industrial capacity.

The validity of the Wright-Rummel measurement procedure is also open to question. Like the measures of military personnel and
expenditures, this index fails to tap the military's responsiveness to political control and the government's reputation for military successes. Another problem is that the index seems to be aimed beyond immediately available military capabilities to a more extreme or long-term notion of military capacity under fully mobilized conditions. This is clearly not the concept we are attempting to measure with this indicator. It might be possible to introduce a greater sense of immediacy to the measure by circumscribing its population component to include only persons of working age or even further, to persons in the military. For present purposes it seemed advisable to forgo such modifications pending our examination of the measure's convergent and discriminant validity.

The final measurement procedure to be considered in this chapter is one developed in conjunction with the 'Correlates of War (COW) Project by Singer, Bremer, and Stuckey (1972; also see Ray, 1979). This index is a bit more complex than the other measures of military power position since it combines multiple indicators of three separate capability dimensions. "The demographic dimension includes, first, total population, and second, the number of people living in cities of 20,000 or larger. The industrial dimension embraces both energy consumption... and...steel production... The third pair of measures are military expenditures and armed forces size, excluding reserves" (Singer, Bremer, and Stuckey, 1972; 26 authors' emphasis). The authors employ a simple but effective method of integrating the disparate
components into a single index. For each component they first derive a total score for the entire system and then compute each nation's percentage share. A nation's score on the final measure is equal to its average share across all six indices and can be thought of as a nation's relative share in the world's available capabilities.

Since many of our earlier comments are also pertinent here, this measurement procedure need not be considered in great detail. For example, we have already discussed some of the measurement difficulties associated with population estimates, energy production, and the two military indices. As for the measure's content validity, the authors state that they "...carefully considered the need for separate indicators of social organization, national unity and motivation, and technical skills, but concluded that each of those was adequately reflected in one or more of the six specific indices." (Singer, Bremer, and Stuckey, 1972: 26). Assuming the authors are correct in their assessment of the breadth of this index, then it may serve better as a measure of overall power than of available military capability. This is a contingency we must keep in mind as we examine the covariation among our twelve candidate measures.

Data were sufficiently available to permit calculation of the COW index for 115 countries. The urbanization measure and the two production indices were taken from Banks (1975) whereas Taylor and Hudson (1972) was the source for total population and the two military measures. Because the composite index is expressed in
percentage form — a fact that renders it adequately sensitive for our purposes — it is confined to values between 100 and zero. Empirically, it ranged from a maximum of 26.61 for the United States to .003 for Iceland. Note, however, that these figures contain a slight upward bias since not all countries were included in the computation of world totals (see note 42).

Summary

This section has been devoted to the presentation and preliminary assessment of twelve procedures for measuring the relative structural positions occupied by 125 contemporary national governments. We have assumed that the structural configuration of an international system conforms to the hierarchical arrangement defined by the distribution of national power capabilities. Accordingly, each measurement procedure focused on some aspect of national power as a means for delineating international positional structure. The first four measures — GNP, Spiegel's rating, Cox and Jacobson's power index, and East's capacity to act — did so under the conventional single system approach. In each case the conceptualization underlying the measure emphasized overall capabilities in foreign affairs undifferentiated by issue or type of problem facing the government. Just the reverse is true of the remaining eight measures. These reflected the multiple systems approach by being more sharply focused on capabilities suited to a limited variety of problems. Four measures were offered in each of two problem areas: The proportion of world trade, Michaely's
commodity and geographically-weighted shares, and Caporaso's economic power ratio were considered appropriate to problems in international trade; armed forces size and expenditures, the Wright-Rummel measure, and the COW Project index were employed for nonviolent military security problems.

As our discussion has made clear, the twelve measurement procedures are not without their difficulties. In particular, we have stressed their susceptibility to various types of measurement error, their frequent lack of cross-national comparability, and their problematical validity. Without wanting to diminish the severity of these problems, let us conclude this section by stating that these measurement procedures provide an acceptable first approximation at this stage in our discipline. The measurement of power and its derivative, positional structure, have proved especially troublesome. This is due, in part, to the elusive and amorphous nature of the construct and, in part, to what Knorr (1975: 3) called the "considerable confusion and disagreement" characterizing the scholarly writing on this subject. Ultimately, however, we must concur with Kaplan's (1964: 176) assertion that "...whether we can measure something depends, not on that thing, but on how we have conceptualized it, on our knowledge of it, above all on the skill and ingenuity which we can bring to bear on the process of measurement which our inquiry can put to use."
Covergent and Discriminant Validity

One vitally important aspect of any measurement procedure is its construct validity. As initially explicated by Cronbach and Meehl (1955), construct validity concerned the extent to which a measured construct actually conforms to theoretically predicted associations with measures of other constructs. With the publication of Campbell and Fiske's (1959) now classic paper on convergent and discriminant validity, the meaning of construct validity was broadened to include the convergence of different measurements of the same construct and, conversely, the divergence of similarly obtained measurements of ostensibly different constructs. In this section we will examine the empirical convergence and divergence of the twelve measurement procedures in order to arrive at a final determination regarding the operationalizations of international positional structure.

Two simple, common sense ideas underlie the notions of convergent and discriminant validity. The former, according to Campbell and Fiske (1959: 83), "...is represented in the agreement between two attempts to measure the same trait through maximally different methods." Surely our confidence in the specification and operationalization of a theoretical construct is justifiably enhanced when two different procedures for measuring that construct yield essentially the same results. Discriminant validity, on the other hand, is demonstrated when a measured construct is shown to be empirically
distinct from measures of other constructs. In Campbell and Fiske's (1959: 84) words, "one cannot define without implying distinctions, and the verification of these distinctions is an important part of the validational process." The degree to which these two basic validation criteria are met can be determined by imbedding the observed measurement in a "multitrait-multimethod matrix." The entries to such a matrix consist of correlations computed among at least two different properties (or traits) each of which has been measured by at least two different procedures. A singular advantage of this approach to construct validation is its ability to decompose a measurement's variance into trait specific and method specific components. Of course, a measure must be considered invalid to the extent that its scores are produced by the method of measurement rather than the property being measured.

The measurement procedures examined in the preceding section lack the symmetry required of a true multitrait-multimethod matrix. To have gained such symmetry we would have had to utilize the same four procedures to measure each of the three variants of positional structure. Without this symmetry we are denied the full power of the multitrait-multimethod technique, though we can still make use of some of its provisions. Specifically, we will be concerned with the congruence of different measurements within each group and the divergence of measurements across groups. For present purposes the latter criterion would seem to be the more important of the two
since such divergence is the key to system differentiation. Earlier 
it was argued that different systems can be distinguished by differ­
ces in their structures. The implications for the present analysis 
are twofold: First, in order to maintain that the military and trade 
problem areas are indeed distinct systems as the multiple systems 
model would have it, then we must be able to demonstrate the structural 
differences between them. In other words, our measurements of 
positional structure must display some noticeable degree of divergence 
between these two putative systems. Second, by the same line of 
reasoning we must also establish the structural differences between 
the overall system and the military and trade systems. This exercise 
actually represents a preliminary and very crude test of the multiple 
systems hypothesis since complete or nearly complete convergence of 
all twelve measurements would entail unequivocal support for the 
single system point of view. Note, however, that even under condi­
tions optimally favorable to the multiple systems model we should 
not expect the differences between the overall system and the two 
issue systems to be as pronounced as the differences separating 
the issue systems from one another. This follows not only from the 
content of the measures but from their conceptualization as well--
overall power is intended to embrace both military and economic 
components.

A triangular matrix of intercorrelations among the twelve struc­
ture measures is presented in Table 9. All entries represent 
Spearman rank correlation coefficients fully adjusted for ties.43
TABLE 9
Intercorrelations Among Structure Measures
(Spearman rho)

<table>
<thead>
<tr>
<th>Overall Structure</th>
<th>Trade Structure</th>
<th>Military Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross national product</td>
<td>0₁</td>
<td>0₂</td>
</tr>
<tr>
<td>Spiegel's power rating</td>
<td>0₂</td>
<td>.938</td>
</tr>
<tr>
<td>Cox and Jacobson's power score</td>
<td>0₃</td>
<td>.948</td>
</tr>
<tr>
<td>Capacity to act (revised)</td>
<td>0₄</td>
<td>.898</td>
</tr>
<tr>
<td>Trade Structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of World trade</td>
<td>T₁</td>
<td>.945</td>
</tr>
<tr>
<td>Commodity-weighted share</td>
<td>T₂</td>
<td>.750</td>
</tr>
<tr>
<td>Geographically-weighted share</td>
<td>T₃</td>
<td>.793</td>
</tr>
<tr>
<td>Economic power ratio</td>
<td>T₄</td>
<td>.959</td>
</tr>
<tr>
<td>Military Structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size of armed forces</td>
<td>M₁</td>
<td>.830</td>
</tr>
<tr>
<td>Military expenditures</td>
<td>M₂</td>
<td>.931</td>
</tr>
<tr>
<td>Population X energy production</td>
<td>M₃</td>
<td>.836</td>
</tr>
<tr>
<td>COW index</td>
<td>M₄</td>
<td>.927</td>
</tr>
</tbody>
</table>
Two reasons lead to our choice of the Spearman coefficient over the more conventional product-moment coefficient: First, for most of our measures we have considerably more confidence in the rank order of assigned scores than in the intervals between them. Second, our immediate purposes call for comparisons on the basis of countries' relative positions rather than proportional differences. Though not of overriding importance, there are also certain technical advantages to the Spearman coefficient — conversion to ranks eliminates outliers and effectively standardizes the variables thus diminishing the likelihood of curvilinear relationships. Although we attempted to collect information for 125 countries, for all but one of the measures we fell somewhat short of that mark. This fact should be kept in mind as we examine Table 9 since the correlations were computed for cases with data jointly present for both measures. The \( N \) size of an individual correlation cannot exceed the smaller of the \( N \)'s for the two variables; over the entire matrix the \( N \) sizes vary between 105 and 125. For convenience much of the information contained in Table 9 has been summarized in Table 10. The entries in Table 10 are each measure's average correlation with the other measures in each group as delineated by the column headings. The darker boxes in the diagonal enclose a measure's average correlation with other measures in its own group. Thus, Table 10 indicates the average convergence and divergence of each measurement procedure.

Surely the most striking feature of Table 9 is revealed in the uniformly high magnitudes of the correlations. Although the
TABLE 10

Average Convergence and Divergence of Structure Measures

<table>
<thead>
<tr>
<th>Overall Structure</th>
<th>Trade Structure</th>
<th>Military Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross national product</td>
<td>.928</td>
<td>.862</td>
</tr>
<tr>
<td>Spiegel's power rating</td>
<td>.885</td>
<td>.821</td>
</tr>
<tr>
<td>Cox and Jacobson's power score</td>
<td>.910</td>
<td>.834</td>
</tr>
<tr>
<td>Capacity to act (revised)</td>
<td>.871</td>
<td>.797</td>
</tr>
<tr>
<td>Trade Structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of world trade</td>
<td>.912</td>
<td>.866</td>
</tr>
<tr>
<td>Commodity-weighted share</td>
<td>.720</td>
<td>.782</td>
</tr>
<tr>
<td>Geographically-weighted share</td>
<td>.768</td>
<td>.775</td>
</tr>
<tr>
<td>Economic power ratio</td>
<td>.914</td>
<td>.866</td>
</tr>
<tr>
<td>Military Structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size of armed forces</td>
<td>.794</td>
<td>.643</td>
</tr>
<tr>
<td>Military expenditures</td>
<td>.894</td>
<td>.784</td>
</tr>
<tr>
<td>Population X energy production</td>
<td>.764</td>
<td>.731</td>
</tr>
<tr>
<td>COW index</td>
<td>.872</td>
<td>.783</td>
</tr>
</tbody>
</table>
Coefficient values range from a notable .959 to .509, only two coefficients in the entire matrix fall below a value of .640. The general picture we get of the overall pattern of relationships is brought into still sharper focus by the summary correlations presented in Table 10. There is no escaping the conclusion that there exists an extraordinary degree of congruence among these measures. Does this high level of congruence obviate the need for additional investigation of the multiple systems hypothesis? After all, if our measures are unable to detect any appreciable structural differences then our putative systems cannot be differentiated and any further analysis through the stratification model is rendered moot.

In our estimation, abandonment of the multiple systems hypothesis is not justified solely on the basis of the correlations in Table 9. Several reasons lead us to this conclusion. In the first place, refutation of the multiple systems model at this stage of our analysis clearly contravenes the Lakatosian methodology outlined early in Chapter I. Lakatos would advise us that it is a greater mistake to relinquish one's theories prematurely than to retain them too long. Secondly, it must be emphasized that our knowledge of systems and system structures is rather meager and that we are still a long way from being in a position to advance an a priori criterion or threshold level specifying the degree of structural divergence minimally required of distinct systems. It may be, for example, that even slight structural differences are sufficient for system differentiation. At any rate, we are hopeful that some clarification
of this issue will result from continuation of our research program. Thirdly, our measurements have been carried out for only a single point in time during the mid 1960's. Of course, we must always beware of conclusions hastily drawn from synchronic analyses, but more than this general methodological maxim lies behind our decision not to abandon the multiple systems hypothesis. Numerous scholars (for example Brown, 1973; Keohane and Nye, 1977) have characterized the early 1970's as a period marked by significant and wide ranging transformation of global politics. Because many of these changes seem to be consistent with tenets of the multiple systems model, a more felicitous research design would have focused on 1975 rather than (or in addition to) 1965; unfortunately, the chronological domain of the present research is circumscribed by the restricted availability of behavioral data (see Chapter IV). Finally, on a more pragmatic level we should note that under close inspection the convergence evinced by Table 9 does not appear so homogeneous as to preclude the possibility of further investigation along the lines suggested by our research strategy. This point requires fuller discussion and interpretation of the information presented in Tables 9 and 10.

We should point out that the matrix in Table 9 divides into six regions. The three small triangular matrices contain correlations indicating the convergence among alternate measurements of the same system structure. The three rectangular arrays comprise correlations across the three structural variants and thus delineate the
divergence of our measures. Let us begin with the middle triangle and
the convergence of the trade structure measures. All of the coeffi­
cients are quite high as we have already indicated but closer examina­
tion reveals the presence of three interrelated dimensions. The
proportion of world trade and the economic power ratio are virtually
interchangeable, sharing about ninety percent of the variance in
their rankings. The two weighted share measures, on the other hand,
are plainly distinct from one another and moderately distinct from
the proportion of trade-economic power dimension.

Strict adherence to the logic of convergent validation would
direct us to the proportion of trade or the economic power ratio as
the most valid measures of trade structure. It bears repeating,
however, that we are chiefly interested in our measures' ability
to discriminate among the three structures and on this criterion the
proportion of trade and the economic power ratio perform rather
poorly. This is particularly evident in their strong relationships
to the overall measures. In fact, both measures display higher
average correlations with the overall measures than with the other
trade indicators (see Table 10). Not unexpectedly, the weighted
share indices behave considerably better. Although the commodity
measure tends to be a bit more discriminatory than the geographical
measure, both seem to fall within tolerable limits. Still, we
must deal with the lack of convergence of these two measures.
Interpretation of the trade system's structural configuration becomes
rather ambiguous if both measures are retained for analysis of
the stratification hypotheses; on the other hand, it is not altogether clear which of the weighted share indices better represents the actual configuration. In the end we elected to utilize both measures on the grounds that they are tapping different but complementary aspects of the power distribution underlying trade structure. Moreover, this approach confirms our earlier remarks concerning the offsetting advantages of these indices and suggests that future efforts might do better to rely on Michaely's composite index where possible.

Having located our indicators of trade structure, let us now examine the military measures. Size of armed forces does an outstanding job of discriminating between the military and trade structures, although the other three measures also perform reasonably well in this respect. We must also consider divergence from overall structure, and on this criterion our choice of indices is narrowed to the size measure and the Wright-Rummel index. Since these two military measures exhibit a relatively low degree of convergence should we treat them as alternative aspects of military structure? We decided against this option, preferring instead to focus exclusively on armed forces size. This decision was based primarily on our earlier criticism of the Wright-Rummel index as too broad or long-term an indicator for our present purposes, and secondarily on the superior discriminatory power of the military size measure. Before turning to the indicators of overall structure it is worth pointing out that the strong convergence displayed by the COW index is
largely a statistical artifact since it incorporates the variables comprising the other three military measures.

The first thing to notice about the measures of overall structure is their very high degree of convergence. Here again, this is no coincidence — the measures are linearly dependent on one another because they are composites of many of the same variables. Gross national product, which in one way or another is built into each of the other indices, offers the clearest illustration of this effect. The pattern of divergence is equally interesting. With the exception of the relationship between Michaely's geographical index and Cox and Jacobson's power score, each of the measures displays greater divergence from the two trade structure indicators than from military size. Moreover, none of the overall measures diverge from the military or trade structure indicators to the extent that these diverge from one another. At this juncture we will omit any further consideration of Spiegel's rating due to its gross insensitivity and low reliability. Despite the fact that it is the least convergent of the three remaining measures, capacity to act was chosen to represent overall structure in our investigation of the stratification hypotheses. We eliminated GNP because it was the least discriminatory and Cox and Jacobson's measure because of its lack of sensitivity.

By emphasizing the divergent qualities of our measurements we have shown that it is indeed possible to differentiate the three system structures, thereby enabling us to continue our investigation of the
multiple systems hypothesis. This is not to deny the overwhelming congruence that exists among these measures. There is a touch of irony in the fact that even the relationships we have characterized as highly divergent would be looked upon as evidence of convergence in most other research contexts. Obviously, our use of "divergence" is meant to be taken in a very relative sense to refer to gaps or discontinuities that texture this high level of congruence. The subtlety of these structural discontinuities implies that any corresponding behavioral differences will be equally subtle. This makes our task more difficult but not insurmountable. It is difficult because we must distinguish differences resulting from measurement error, nonstratification influences, and other types of "noise" from whatever slight differences there may be that genuinely indicate responsiveness to distinct system structures. The details of our research design along with the operationalization of the various dimensions of foreign policy behavior are the topics to be covered in the next chapter.
1 Kaplan (1963) makes a similar distinction using the terms "fundamental" and "derived" measurement.

2 In a strict sense there is no such thing as a true score since entirely error-free measurement is impossible. Nevertheless, as error is reduced successive measurements should begin to converge toward some particular value that may be conceived as a "true" score. (See Kaplan, 1963).

3 The first four criteria are discussed at some length in chapter six of Selltiz, Wrightsman and Cook (1976); also see Galtung (1967).

4 We will not attempt to address the criterion or pragmatic validity of our measurement procedures. Criterion validity requires a reasonably valid and reliable criterion with which the measured scores can be compared. No such criterion exists for international positional structure.

5 For an innovative attempt to deal with such cross-national differences, see Rummel (1972).

6 Data for 1965 are contained in Taylor and Hudson (1972), which was utilized as the major data source for over half of our indicators.

7 For a similar set of rules, see Rummel (1972).

8 For a non-technical nevertheless informative discussion of these issues, see Taylor and Hudson (1972: 287-289) or Russett et.al. (1964: 149-151).

9 These data were acquired from Taylor and Hudson (1972). Details concerning sources, conversion rates, adjustments, and estimates of accuracy are discussed on pages 287-289. GNP was unavailable for the Maldives.

10 For the purposes of our analysis it is only six levels since the seventh is composed of dependent territories and what Spiegel terms "mini states" (e.g., Monaco, Liechtenstein, etc.).
A cursory inspection of the raw statistics for GNP, GNP per capita, and population indicates that this assumption probably holds for most countries. We should note, however, that a number of countries, perhaps as many as ten or fifteen, generally reside close enough to the dividing points to have their scores altered by error alone.

All procedures for constructing composite indices, no matter how simple, contain some sort of weighting mechanism. When the weights are unitary (or some other constant) for all components we often consider the index "unweighted," though this is technically incorrect. The more interesting case referred to in the text involves using different, or variable, weights for the different components.

This procedure is detailed in a technical appendix to East (1975) prepared by Barbara Kay Winters and Joe Hagen.

Component scores are calculated from standardized values of the original indicators. For any given observation, a component score is defined as

\[ S_j = \sum_{i=1}^{k} \left( \frac{b_{ij}}{\lambda_i} \right) z_j \]

where \( b_{ij} \) is the component loading for the \( j \)th variable on \( i \)th component, \( \lambda_i \) is the associated eigenvalue, and \( z_j \) is a standardized variable such that

\[ z_j = \frac{x_j - \bar{x}_j}{s_j} \]

Kim and Mueller (1978: 73) point out that "division by the eigenvalue is cosmetic in that it merely assures that the resulting index has a variance equal to 1."

This adjustment is accomplished by adding the absolute value of the minimum score to each observation.

In the paper documenting the operational procedure (East, 1975) this final step is omitted. Size and social organization were combined in an earlier paper, however (see East and Winters, 1975).

In fact, East and his associates experimented with both procedures and found the results to be quite similar (see appendix to East, 1975).

Perhaps the following passage from Rummel (1972: 171) will help clarify the dimensions of this problem:

...the United States is 3.9 and 7.8 standard deviations from the mean on domestic mail [per capita] and GNP, respectively, while all remaining nations have values on both variables less than 2.0 standard deviations from the mean. Therefore, these extreme values for the United States inflate the means for both variables and cause most nations to bunch on the low
side of distribution. This implies that the correlation found between domestic mail [per capita] and GNP will be highly dependent on these outliers for the United States.

In a strict sense we did not actually replicate East's work since we introduced several minor changes in addition to transforming the data. The most important of these resulted from a decision to limit the analyses to our population of 125 governments (rather than the 136 countries and territories analyzed by East). Five other changes involved individual variables used in the analyses:

a. As one indicator of modernization East used the percentage of a nation's population living in cities of 100,000 or more inhabitants in 1965. Because of the substantial definitional error involved in the demarcation of cities, we chose to drop this indicator in favor of a similar one based on a standardized definition of "urbanized areas" having 100,000 or more inhabitants. Although this indicator applies to 1960 (rather than 1965), we concluded that the significant reduction in definitional error more than compensated for the error introduced by the time difference. See Taylor and Hudson (1972: 200-202) for further discussion of these indicators.

b. Another modernization indicator—proteins per capita per day—was available for only fifty countries from Taylor and Hudson (1972), East's principal data source. In order to retain this variable in his analysis East chose to rely on an alternate source, World Economic Survey, 1973, published by the United Nations. This document contains a table grouping countries into six categories based on specific ranges of proteins per capita. Unlike East, we decided not to use this grouped variable. Instead, we supplemented the data in Taylor and Hudson (1972) with estimates based on information contained in World Economic Survey, 1973.

c. One of East's stress indicators—growth in GNP—was eliminated from our analysis altogether. We did this for two reasons. First, it failed to load above .37 on either stress component in East's analysis. Second, and far more important, we were unable to acquire the data for more than 85 of our 125 governments even though the appendix to East (1975) states that the data were available from the United Nations Statistical Yearbook for 102 countries. It is possible (although we cannot say for certain) that this discrepancy is due to East's use of growth in current GNP rather than GNP in constant prices. The latter, of course, is essential for an accurate picture of real growth.
d. Another of the stress indicators—three year average inflation rate—was calculated over the period of 1964 to 1966 (rather than from 1963 to 1965 as in East's analysis).

e. Two other indicators of societal stress included in East's analysis were the annual number of deaths from political violence and the annual number of political strikes, both for 1965. Because such events are extremely turbulent and are apt to change drastically from year to year, we substituted the average number of deaths and the average number of strikes occurring from 1964 to 1966.

Rummel (1972) recommends the more common chi square test but the Kolmogorov-Smirnov test has been shown to be more appropriate for continuous distributions (See Siegel, 1956).

These variables are percent urban population, literacy rate, percent of GDP originating in industry, and an index of ethno-linguistic fractionalization.

For further discussion of transformations applicable to political events, see Taylor and Hudson (1972: 386).

Once again, we must emphasize that our results are not strictly comparable to East's for the reasons outlined in note 19.

Rotational techniques are applied for the sole purpose of facilitating substantive interpretation of factor and component structures. In the case of the latter, however, rotation also changes its mathematical interpretation since rotated components no longer represent maximum (i.e., unique) proportions of total variance (see Morrison, 1967: 227).

Because of excessive missing data we were unable to compute capacity to act scores for North Korea, North Vietnam, and the Maldives.

For example, many governments routinely report consumer price indices solely upon prices in their capital cities.

So that this statement is not read out of context, we should note that Knorr (1975) considers the volume of economic transactions to be only one of the factors contributing to national economic power. The other major factor, which he labels the "structure" of transactions, encompasses elasticities of demand and supply, commodity composition, and geographical concentration.
For details see Taylor and Hudson (1972: 349) or the introduction to any volume of the *Yearbook of International Trade Statistics* compiled by the United Nations Statistical Office.

Data were unavailable for Nepal, North Korea, and North Vietnam.

A major exception is the trade of industrialized countries in manufactured goods.

There are two minor differences between de Vries' index and Michaely's. The latter expressed all magnitudes as values in a common currency; de Vries, on the other hand, used values for the $X_{ij}$ term and quantities for the $X_{ij}$ term. In principle, the ratios obtained from quantities and values should be identical; in practice, however, there may be some slight discrepancy due to fluctuating prices, exchange rate complications, and so forth. The second difference has to do with the commodities encompassed by the two indices. Michaely chose to include all commodities whereas de Vries included only "major" commodities, where "major" refers to any commodity comprising five percent or more of any developing country's exports. Here again, the difference is negligible since a weight of less than .05 would have little impact on the overall index.

According to United Nations estimates, only about eighty-eight percent of total world trade is reported on the basis of the SITC. (United Nations, Statistical Office, 1969).

Luxembourg's exports were included in the calculation of Belgium's weighted-commodity share. Other countries excluded because of missing or incomplete commodity statistics are Haiti, East Germany, Albania, Bulgaria, Romania, Guinea, Burundi, Yemen, China, Outer Mongolia, North Korea, Maldive Islands, Nepal, Laos, North Vietnam, and Singapore.

Once again, Luxembourg is combined with Belgium. Other countries with missing or incomplete data are Nepal, North Korea, North Vietnam, and the Maldive Islands.

For the economic power ratio to equal one it is not necessary that a country's exports equal its imports. In fact, depending on a country's level of consumption, its ratio may range above or below unity irrespective of whether its trade balance yields a surplus or deficit.

Data were unavailable for Nepal, North Korea, North Vietnam, and the Maldive Islands.
37 There is an effect on certain of the mathematical properties of the economic power ratio, however. In principle, the mean value of the ratio should be one and its sum over all countries should equal the number of countries, in this case 121. Our measurement of the ratio has a mean of 1.05 and a sum of 127.6.

38 Data are missing on both indicators for Malta, Gambia, the Maldives, and Singapore.

39 Inexplicably, Rummel (1979: 73) includes the following sentence in his discussion of this indicator: "The number of people a country has, weighted by the amount of energy it can produce per person -- energy production times population -- is highly correlated with a variety of measures of a nation's power..." (emphasis added). Note that insertion of the words "per person" implies use of energy production per capita which would yield total energy production when multiplied by population. Clearly, this is not what Rummel intended.

40 Since population data were available for all 125 countries, the availability of the index depended entirely on the availability of energy production statistics. These could not be obtained for Luxembourg, the Maldives, Outer Mongolia, North Korea, and North Vietnam.

41 This index was intended to measure military capabilities at regular intervals over a time span of one and a half centuries beginning in 1820. To account for changing technology, the industrial dimension shifts from iron to steel production in 1895.

42 At least one of the six indicators was unavailable for Luxembourg, Malta, Gambia, China, Outer Mongolia, the Maldives, North Korea, North Vietnam, Malaysia, and Singapore.

43 The coefficients should be interpreted as Spearman rhos although the calculations were carried out using the product-moment computational formula applied to the ranks of the values of each variable. This has the effect of adjusting for ties since whenever two or more values of a variable are equal (tied) they are assigned their average rank. This average is equal to the average that would be found if all the values had been slightly different (see Barr et al., 1976).
CHAPTER IV
DATA AND METHODS II:
ANALYSIS OF FOREIGN POLICY BEHAVIOR

This chapter continues the task begun in Chapter III. Recall that our proposed research strategy focuses on several theoretically predicted relationships between positional structure and variously patterned dimensions of intergovernmental foreign policy behavior. Empirical determination of the actual relationships requires operationalization of both sets of concepts. The previous chapter accomplished this for the structural side of these relationships. Twelve candidate measurement procedures—four in each of the three structural variants—were explicated in some detail and evaluated according to a diverse set of measurement criteria. After careful consideration, these twelve were narrowed to four that seemed best able to differentiate the structural arrangements underlying a single, comprehensive system on the one hand, and two issue- or problem-based systems on the other. Of course, whether these alternative structural arrangements are apparent or real is the fundamental question of our inquiry.

Two chores—both addressed in the present chapter—remain before the results of our research can be coherently reported and appraised. The first, which has been adverted to already, involves the operationalization of foreign policy behavior. Although this task
parallels that undertaken in the preceding chapter, our approach here will be somewhat different since we do not have the luxury afforded by multiple measurements of each concept. Procedures for the observation and measurement of governmental behavior are presented in the first section of the chapter. The second section takes up the final task: specification of the methods used to empirically test the fit of our stratification hypotheses to the two system constructs.

Measuring Patterns of Foreign Policy Behavior

If there is one characteristic feature of all social science it is an enduring and deep-rooted concern for the context, sources, and consequences of human behavior. Indeed, contemporary usage treats social and behavioral science as interchangeable designations of the same enterprise. Naturally, this inveterate concern has implied a concomitant need for controlled observation and measurement— that is, operationalization— of behavior. Among the special problems complicating operationalization of behavioral concepts, in contrast to attributes like physical size or strength, is the difficulty of observing, let alone measuring, a phenomenon that is by nature fleeting and insubstantial. Of course, there are laboratory techniques for direct measurement of physiological or psychological responses to induced stimuli, but whenever social research takes to "the field" simple observation of behavior is rendered problematical at best. Furthermore, we should point out that observing behavior of complex organizations like national governments is, if anything, even more formidable than observing individual behavior.
Though formidable, these problems generally have not proved totally intractable for two reasons. The first has to do with the fact that many types of human and organizational behavior, though themselves not easily observed, nonetheless leave behind clearly visible traces of their occurrence. Perhaps the most familiar example of this effect is the "trace" that remains when a roll call vote is taken in a formal, deliberative assembly. In this case, of course, direct observation of behavior is possible provided the observer has physical access to the chamber and the vote is taken openly, but the point is that this sort of behavior is observed more easily and possibly more accurately "secondhand" in the assembly's records. The second factor facilitating observation of behavior has been an almost continuous improvement in what Kenneth Boulding (1963: 22) has termed "instrumentation,... the methods by which information coming from the outside world can be detected, sampled, and processed". In the past two decades these two factors have intertwined with dramatic effect in an area that has come to be known as international events research.

The major premise of event data stipulates that behavior is ultimately comprised of one or more discrete actions each minimally divisible into at least three analytical components indicating "who does what to whom". Such individual actions, or events, have become widely accepted as a fundamental unit of observation in international relations. Development of the event as an observation unit represents a significant advance in instrumentation not because events themselves
are susceptible to direct observation, but rather because identifiable traces or artifacts of events remain in the elite press and other public sources long after their actual occurrence. A fuller discussion of event methodology will be forthcoming shortly since the event approach underlies our operationalization of the behavioral concepts introduced in connection with the stratification hypotheses. First, however, it will be useful to summarize the main stages involved in the event data research process.

Although there are numerous categorization schemes for delineating stages of empirical research, we find Munton's (1978) typology particularly illuminating with respect to event data. The process begins with initial observation and recording of activities by the data source. This observation stage is extremely important in events research even though it is not normally under the investigator's direct control. The second, or datamaking, stage actually involves two distinct sets of procedures: one to extract identifiable events from the source material and another to subsequently encode these isolated events according to various properties or characteristics. Unlike the observation stage, all facets of datamaking are strictly controlled by the researcher. The penultimate stage is where measurement procedures transform newly coded events into indicators of behavior suitable for analysis in the final stage of the process. This simple schema, illustrated in Figure 3, organizes our discussion throughout the remainder of the chapter.
FIGURE 3

Stages of Events Research

Source: Munton (1978: 11)
Observing Foreign Policy Behavior

As noted above, the researcher first becomes actively involved at the datamaking stage of the events research process. In fact, most events researchers forgo even this opportunity, preferring instead to conduct secondary analysis on data originally collected by others. The reason is wholly pragmatic: event datamaking is an extremely laborious and time-consuming enterprise that under most circumstances is beyond the physical capabilities and financial resources of a single individual. Heeding this logic, we resolved in the earliest stages of our inquiry that it would be necessary to rely on an established data base for operationalization of our behavioral concepts.

As our overall research strategy began to crystalize, it became increasingly clear that our behavioral data would have to be acquired on the basis of three seemingly modest requirements. First, they would have to encompass events for actors residing at all levels of the structural hierarchy. Notice that this criterion does not necessarily entail universal coverage of all national actors, albeit that is certainly a desirable property; all that is needed is broad representation of the spectrum of structural positions. Hence, data limited to actors in a particular geographical region such as Africa or the Middle East would not qualify on this count. A second requisite of our behavioral data was that they differentiate events according to issue or problem area. This was essential for the realization of our research strategy since it would provide a basis for
allocating behavior to separate issue-based systems. Finally, we required that our data be multidimensional with respect to the properties or characteristics ascribed to behavior. To summarize the argument proffered in Chapter II, examination of a diverse set of behavioral concepts constitutes explicit recognition of the complex and multifaceted nature of foreign policy, provides some indication of the pervasiveness of systemically induced effects, and insures that whatever conclusions are reached reflect more than the idiosyncrasies of a single concept or its operationalization. The event data judged best able to meet these criteria were collected under the direction of Charles Hermann and his associates (1973) as part of a larger project investigating governmental foreign policy behavior. Although this data base is amply documented elsewhere, several of its features deserve brief elaboration here. In particular, we will touch upon its conceptualization of an event, the nature of its source material, and its scope of coverage with respect to actors and chronological domain.

Insofar as these data are concerned, an event is formally defined as "...a minimally aggregated action resulting from a decision by the political authorities of a state who have the power to commit the resources of the national government" (Hermann et. al., 1973: 19). Furthermore, an event is conceived as having four (not three) analytical components: one actor, an action, one or more direct targets, and one or more indirect objects. The first thing to notice about this formulation is its close congruence with our definition of
foreign policy behavior as discrete official actions intended to influence the behavior of other international actors. This similarity surely is no accident since both definitions are the work of the same author and, more importantly, both are deeply rooted in the decision-making approach. A discrete foreign policy action and a foreign event share a common origin in a decision taken by one or more individuals at the authoritative or political level of a national government.

A second aspect of this definition that requires comment concerns the meaning and significance of the phrase "minimally aggregated action". It is presumed that each distinct action ensues from a different political level decision and thus should be represented as an individual event. Callahan (1979: 35) offers an illuminating example:

Very often, an activity such as a speech, a press conference, or a communique will contain a variety of actions. For example, President Kennedy's speech of 22 October, 1962 contained a number of distinguishable actions: it announced the naval quarantine of Cuba; it asserted U.S. intentions to retaliate against any nuclear attack from Cuba with a nuclear attack on the Soviet Union; it declared U.S. intentions to maintain close surveillance of Cuba; it announced the reinforcement of the Guantanamo naval base; it called for a meeting of the O.A.S.; it called for a meeting of the U.N. Security Council; and it called upon Premier Khruschev to order the removal of the missiles from Cuba.

The minimal aggregation principle requires that each unique action be recorded as a separate event. As a general rule, separate political level decisions warranting distinct events were assumed whenever there occurred a change in "(1) the individual or group making
or announcing the decision, (2) the time frame for the action, (3) the level of commitment, or (4) the kinds of resources used" (Callahan, 1979:35). It should be emphasized that minimal aggregation applies only to actions (or decisions). More specifically, events are not minimally aggregated with respect to the immediate recipient(s) of the action (direct target(s)) or the object(s) of the influence attempt (indirect object(s)). This coding convention is fully consistent with the decision-making perspective since a discrete action emanating from a single decision may be explicitly addressed to multiple recipients or may be intended to influence multiple entities.

A vital aspect of any event data base is the source material from which its events are abstracted. The adequacy of one's source(s) at the initial observation stage of the events research process will have major implications for all other stages, including the final analysis. Source adequacy may be affected by such factors as ideological or cultural bias, overreporting of conflictual or otherwise spectacular actions, or a lack of attentiveness to certain countries or geographical regions. Our data are drawn from Deadline Data on World Affairs, a global chronology compiled from a worldwide sample of journalistic sources including newspapers, magazines, wire services, and radio broadcasts.

Source validity is an exceedingly complex issue which must be approached on two levels. The more immediate level has to do with the relative ability of one or more data sources to produce a sample of events adequately representative of reality. Events researchers
sensitive to this type of validity problem have amassed a sizeable literature investigating such matters as the comparative utility of numerous individual sources; differences in coverage among global, regional, and national sources; the relative benefits of relying on multiple sources; and the advantages of issue-specific documents over more general news sources. 5

At a more fundamental level, the validity issue is less a matter of the relative superiority of competing sources than of the acceptability of public sources altogether. It is frequently pointed out, for example, that a conscious decision not to act produces no new behavior and thus leaves no trace to be recorded by the event source. Many governmental actors routinely engage in covert foreign activities with the same effect: no publically observable trace is immediately available to indicate that some behavior has taken place. On the other hand, a strong argument favoring the use of public news sources can be advanced on the grounds that "...policymakers themselves rely on this type of source for information about the international system, and in fact, have their image of international affairs--their 'reality worlds'--partially molded by the news they read" (Kegley, 1975: 94-95). It is unlikely that any definitive resolution of the various source validity issues will be forthcoming in the near future. As for our data, careful consideration of these issues lead Callahan (1979: 63) to remark that they "...may be used in empirical analysis with fairly high confidence in all but the most subtle distinctions". (also see Salmore and Butler, 1978).
Before turning to specific behavioral dimensions and the procedures used to generate them, let us examine the sampled domain of our data base. At present events have been identified and coded for thirty-eight national governments over a time period of thirty months. The thirty month time frame was established by randomly selecting one quarter from each year during the 1959-1968 decade (Hermann, et. al., 1973). The months chosen to represent each year are displayed in Table 11. The governmental actors are delineated in Table 12 along with the total number of events for each. It is important to note that only the actor component of an event is limited to these thirty-eight nations; the recipient components (direct target(s) and indirect object(s)) may refer to any individual, government, or other definable collectivity. Owing to a rather haphazard selection procedure, this group cannot be designated a scientific sample and although the final mixture of nations is clearly not representative of the world community there is nevertheless wide variation in terms of the basic national attributes of size, development, political accountability, and geographical location.

For present purposes, the more important question is whether this "sample" accommodates sufficient variation on the four structural measures selected in the previous chapter. The information presented in Table 13 indicates that this subset does preserve much of the total variation despite the fact that it is heavily biased toward the top rankings on each indicator. The left-hand side of the table compares the "sample" and population means, ranges, and standard deviations.
Table 11
Time Sample for Behavioral Data

<table>
<thead>
<tr>
<th>Year</th>
<th>Selected Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>1959</td>
<td>October, November, December</td>
</tr>
<tr>
<td>1960</td>
<td>April, May, June</td>
</tr>
<tr>
<td>1961</td>
<td>January, February, March</td>
</tr>
<tr>
<td>1962</td>
<td>October, November, December</td>
</tr>
<tr>
<td>1963</td>
<td>April, May, June</td>
</tr>
<tr>
<td>1964</td>
<td>July, August, September</td>
</tr>
<tr>
<td>1965</td>
<td>January, February, March</td>
</tr>
<tr>
<td>1966</td>
<td>July, August, September</td>
</tr>
<tr>
<td>1967</td>
<td>April, May, June</td>
</tr>
<tr>
<td>1968</td>
<td>October, November, December</td>
</tr>
</tbody>
</table>

Source: Callahan (1979: 40)
Table 12

Governmental Actors

<table>
<thead>
<tr>
<th>Actor</th>
<th>Events</th>
<th>Actor</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>381</td>
<td>Lebanon</td>
<td>150</td>
</tr>
<tr>
<td>Canada</td>
<td>388</td>
<td>Mexico</td>
<td>191</td>
</tr>
<tr>
<td>Chile</td>
<td>209</td>
<td>New Zealand</td>
<td>201</td>
</tr>
<tr>
<td>China (PRC)</td>
<td>505</td>
<td>Norway</td>
<td>225</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>164</td>
<td>Philippines</td>
<td>199</td>
</tr>
<tr>
<td>Cuba</td>
<td>326</td>
<td>Poland</td>
<td>233</td>
</tr>
<tr>
<td>Czechoslovakia</td>
<td>206</td>
<td>Soviet Union</td>
<td>1006</td>
</tr>
<tr>
<td>East Germany</td>
<td>202</td>
<td>Spain</td>
<td>171</td>
</tr>
<tr>
<td>Egypt</td>
<td>454</td>
<td>Switzerland</td>
<td>74</td>
</tr>
<tr>
<td>France</td>
<td>855</td>
<td>Thailand</td>
<td>163</td>
</tr>
<tr>
<td>Ghana</td>
<td>273</td>
<td>Tunisia</td>
<td>276</td>
</tr>
<tr>
<td>Guinea</td>
<td>209</td>
<td>Turkey</td>
<td>317</td>
</tr>
<tr>
<td>Iceland</td>
<td>159</td>
<td>Uganda</td>
<td>146</td>
</tr>
<tr>
<td>India</td>
<td>496</td>
<td>United States</td>
<td>1916</td>
</tr>
<tr>
<td>Israel</td>
<td>348</td>
<td>Uruguay</td>
<td>138</td>
</tr>
<tr>
<td>Italy</td>
<td>425</td>
<td>Venezuela</td>
<td>208</td>
</tr>
<tr>
<td>Ivory Coast</td>
<td>148</td>
<td>West Germany</td>
<td>646</td>
</tr>
<tr>
<td>Japan</td>
<td>271</td>
<td>Yugoslavia</td>
<td>257</td>
</tr>
<tr>
<td>Kenya</td>
<td>96</td>
<td>Zambia</td>
<td>78</td>
</tr>
</tbody>
</table>

Source: Callahan (1979: 39)
All three statistics were calculated on the rank values obtained from the full population. The subset's upward bias is clearly evident in the differences between the two columns of mean rank values. For example, when 122 countries, the number for which data were available, are ranked by their capacity to act scores, the midpoint or mean rank is 61.5; however, when the thirty-eight sample countries are considered by themselves, the mean rank value is only 39.4. This difference indicates that the sample contains a larger proportion of high ranking countries—that is, countries scored with low numerical values—than the population as a whole. Note, too, that this bias appears to have relatively little impact on the two measures of dispersion. A more detailed view of the subset distributions on these indicators is depicted by the quartile percentages listed in the right-hand portion of the table. The entries specify the proportion of subset nations falling into each quartile of the full population. Thus, a perfectly representative sample would distribute twenty-five percent of its cases into each quartile. Here again the upward bias is unmistakable, though in our estimation it is not so severe as to preclude use of these data for operationalizing our behavioral concepts.

Our discussion thus far has focused on the observation and event identification stages of the events research process (see Figure 3). To summarize, we have indicated the necessity of secondary data analysis and enumerated three requisite characteristics of a suitable data base. An appropriate data base was identified and briefly profiled in terms of its conceptualization of an event, its source
material, and its selection of actors and time periods. Special attention was given to the ability of the nation subset to adequately reflect variation in structural position. The next section continues our overview of the data making stage by briefly examining the coding procedures used to delineate the separate behavioral dimensions appearing in our ten stratification hypotheses.

**Dimensions of Foreign Policy Behavior**

An event is a device for observing and recording information about the occurrence of a discrete behavioral act. Occasionally all that is necessary is the initial identification of an event. For example, the number of identified events, indicating the number of observed behaviors that have taken place, can be used as a measure of a government's level of international activity. More frequently, however, event identification is only the preliminary step to an additional set of datamaking procedures designed to codify and record particular characteristics of behavior. This aspect of the datamaking stage involves application of specified coding rules to classify or measure a previously identified event on one or more behavioral properties. Space limitations preclude a fully detailed account of the coding instructions underlying the behavioral dimensions introduced in Chapter II. Nevertheless, we can summarize their main features.

Any discussion of coding procedures inevitably must deal with the question of coding reliability. Can the coding instructions be unambiguously applied to the same material by different coders or by a single coder on different occasions to yield the same results? A number of different methods exist for generating rather specific
answers to this question. The researchers responsible for the collection of our behavioral data have made extensive use of the Krippendorff (1971) coefficient of agreement to determine the degree of inter-coder reliability (Hermann, et al., 1973). The Krippendorff coefficient measures the proportion of agreement between two (or more) coders beyond that which is attributable to chance alone. Its values range from zero, indicating purely random agreement, to one, which is attainable only if there is perfect agreement. Wherever possible in the following discussion we will report the Krippendorff measure of reliability applicable to the specific coding rules under consideration.

A word of caution is in order regarding the interpretation of inter-coder reliability coefficients. Normally, reliabilities are calculated over the full set of categories defined by a particular coding scheme. This practice can be misleading whenever three or more categories are involved since the coding rules leading to any one category may be more or less ambiguous than the rules leading to other categories (Callahan, 1979). In other words, calculation of a single reliability coefficient tends to obscure the fact that individual categories need not—and often will not—be equally reliable. This problem becomes particularly relevant when the researcher makes use of a few selected categories rather than an entire variable. We will return to this point later in our discussion since it is pertinent to some of the behavioral properties examined below.

Let us begin with the two behavioral dimensions associated with a government's level of international involvement. These two properties, the amount of activity and scope of action, deviate from the others introduced in Chapter II in that they cannot be meaningfully
conceived at the level of the individual behavioral act. As a result, neither property is operationally defined by coding procedures other than those applied at the event identification stage. In other words, if we know, first, the number of events identified for a particular governmental actor and, second, the recipients to whom those events were directed (delineated as part of the event identification process) then we are in a position to operationalize that government's amount of foreign activity and scope of action, respectively. This last statement notwithstanding, on occasion it will be desirable to narrow one's conceptualization of international involvement by modifying the operational procedure to encompass only certain types of actions or selected recipients. In fact, we have implemented just this sort of modification with regard to scope of action.

In Chapter II scope of action was defined as the dispersion of an acting government's foreign policy activities across various recipients in the international arena. One rather obvious way to operationalize this concept using event data is to count the number of different recipients identified in the events for each governmental actor. Because our stratification theory focuses exclusively on intergovernmental behavior, we delimited scope of action to include only external governmental entities by eliminating from consideration any recipients classified as international organizations or domestic audiences on the "Recipient Differentiation" variable (reliability = .96) (Hermann, et al., 1973). Further complicating this procedure is the fact that the recipients appearing in a single event can assume several different functions or roles. Some of these roles tend to be rather tangential to the substance of the foreign policy
### TABLE 13

Representatives of "Sampled" Nations

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Descriptive Statistics</th>
<th>Quartile Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Range</td>
</tr>
<tr>
<td></td>
<td>Sample Pop.</td>
<td>Sample Pop.</td>
</tr>
<tr>
<td>Capacity to act</td>
<td>39.4 61.5</td>
<td>108 121</td>
</tr>
<tr>
<td>Commodity-weighted share</td>
<td>38.9 54.5</td>
<td>105 107</td>
</tr>
<tr>
<td>Geographically-weighted share</td>
<td>42.3 60.5</td>
<td>103 119</td>
</tr>
<tr>
<td>Armed forces size</td>
<td>45.8 61.0</td>
<td>118 118</td>
</tr>
</tbody>
</table>

*a* Statistics were calculated on the rank values obtained from the full population (see note 6). Population figures apply to all 125 nations.

*b* Quartile divisions were determined from the full population. Entries indicate the percentage of subset nations located in each quartile.
action. For example, an entity that provides the site for a meeting but otherwise is not a participant is counted as a recipient in the role of "facilitator." In order to preserve the idea that foreign behavior is purposely directed to external entities we used the "Roles of Recipients" variable (reliability = .83) to further delimit scope of action by including only those recipients serving as addressees, as objects of an influence attempt, or as fellow participants in the action.7

Incidentally, we might also point out that our cautionary remarks on the interpretation of reliability coefficients are directly applicable to this selective use of recipient roles. In effect, we have taken the roles defined by the original coding procedures and regrouped them into two sets based on whether or not a particular role involves a recipient in the substance of the behavior. The reported reliability of .83 signifies a moderate amount of inter-coder agreement in applying the original set of role categories; however, in this case a more revealing indicator of reliability would measure agreement between the two regrouped categories. Although we can do no more than speculate, it does seem reasonable to suppose that the dichotomized classification is more reliable than the original coding.

The next cluster of behavioral properties concerns the autonomy of foreign policy behavior. The unilateral-multilateral dichotomy was derived from a variable originally intended to record the actual number of governments participating in an event (realibility = .88) (Hermann, et.al., 1973; Hermann, 1979). Here again, a regrouping
procedure was employed to differentiate occasions on which a government acted alone from actions taken in concert with one or more others. In Chapter II the term "collaboration" was used in a special sense to refer to foreign activities taken jointly with a particular dyadic recipient. The subtle distinction between multilateral and collaborative actions becomes relevant at the level of dyadically patterned behavior and will be elucidated more fully in the next part of this chapter. For the present let us simply acknowledge that the "participant" category of recipient role serves as an event specific indicator of collaboration between the actor and a particular recipient.

Discerning the initiative-reactive aspect of behavioral autonomy is a far more complicated matter. This is because a discrete event can be classified as an initiative—or, conversely, as a reaction—on the basis of several, seemingly unrelated behavioral properties. Hermann (1979) simplifies the task by detailing the properties characteristic of reactions and considering all residual actions to be initiatives. Accordingly, the following six questions are intended to elude the reactive nature of foreign policy behavior:

1. Was the action elicited, that is, is the action a response to the behavior of another government or external entity which explicitly requested a response from the actor?
2. Is the action a public vote by a government representative in an international organization?
3. Does the action involve an actual agreement, announcement of, or transfer of a loan, credit, or other foreign assistance?
4. Does the action involve an actual agreement, announcement of, or execution of an economic transaction?
(5) Is the action a denial, a rejection of or agreement to a prior proposal, or a protest of a prior activity directed at the acting government by another government or external entity?

(6) Does the action involve negotiation, consultation, or a response dictated by a treaty (Hermann, 1979: 420-421)?

An affirmative answer to any one of these questions is sufficient for a foreign policy action to be considered reactive, otherwise it is designated an initiative. It should be emphasized that the six questions are listed solely as an illustrative device and do not represent actual coding instructions or coded variables. Selected categories of eight different variables were used to operationalize the conditions indicated in the questions.

The remaining behavioral property, labeled problem sensitivity in Chapter II, is derived from the substantive problem area classification scheme developed by Hermann and Coate (1979). However, the immediate significance of the problem area classification extends far beyond its role in delineating problem sensitivity since it is this coding scheme that provides the substantive basis for assigning foreign policy behaviors to problem-specific systems. The classification is actually a multidimensional typology designed to characterize both the nature of a problem and the context in which it occurs. For purposes of simplicity and brevity, we will forgo even summary explication of the full classification scheme and concentrate only on those dimensions pertinent to our present research effort.
Application of the substantive problem area scheme begins with identification of the specific problem precipitating the foreign policy event. Recall that a problem was defined as a perceived discrepancy between some preferred state of affairs and the state or condition that actually exists or is anticipated. Problems are subjective and are identified from the perspective of the acting government. The first dimension of the problem area coding scheme seeks to determine what fundamentally desirable quality or condition is (potentially) affected by the problem outcome. In other words, what basic human value is jeopardized by the existence of this problem? Hermann and Coate (1979) advance five general categories of basic values regularly represented in the foreign policy actions of national governments. The value classification provides the first glimpse of the problem's substantive content by broadly categorizing the impeded goal as a desire for security or physical safety, economic wealth, respect or status, well-being or welfare, or enlightenment (reliability = .94). Although these five categories are mutually exclusive, this is not to deny that some problems involve mixed value outcomes. Hence, the coding scheme was designed to accommodate up to three basic values for each discrete event. Despite this contingency, multiple basic values were deemed appropriate in fewer than ten percent of the more than 12,000 identified events.

The second problem area dimension delineates part of the problem's context by indicating from the actor's perspective what entity or collectivity is or will be deprived of the specified basic value(s).
Put another way, this dimension requires the coder to determine who will benefit if the problem is successfully avoided or resolved. The coder then records this information using one of six primary categories: (1) the acting government or its entire society; (2) citizens or groups in the actor's nation; (3) external governments or entire societies; (4) citizens or groups of other nations; (5) the actor and other nations temporarily grouped together; or (6) an alliance, regional, or global organization, (reliability = .81).

This coding scheme was collapsed by merging together the first two, middle two, and last two categories to form, respectively, the internal, external, and communal categories of problem sensitivity. 10

One of the distinctive features of the substantive problem area classification is its decision tree style format in which the coding of certain dimensions is linked to categories chosen on one or more prior dimensions. Thus, even though the next problem area dimension is of no inherent interest to our present inquiry, it nevertheless must be introduced as part of a decision tree that will eventually permit determination of the problem type used to assign behaviors to problem-specific systems. This next dimension also helps to set the context of the problem by indicating what human or nonhuman agent(s) seems most immediately responsible for the present or anticipated value deprivation. In other words, who or what is the source of the problem? The decision tree mechanism keys on three broad categories: governmental sources including the government(s) of the deprived nation(s), one or more external governments, and intergovernmental
organization(s); nongovernmental sources including individuals or groups from the deprived nation(s) or from some external nation(s), and nongovernmental transnational organizations; and a nonhuman category for acts of nature and general societal conditions. These three summary categories represent a regrouping of nine somewhat more specific categories contained in the original coding instructions (reliability = .77).

The final dimension to be considered here undertakes a further classification of the problem's substantive content. At this point in the scheme, the coder is asked to select from a list of categories that reflect moderately detailed characterizations of various types of problems. Following the decision tree structure, choice of problem type is dependent upon the prior specification of basic value and source of deprivation. There are a total of sixty-six types of problems but, owing to the decision tree formula, the coder actually chooses from among a subset of no more than nine possible categories. The exact number of categories in each subset varies but the number of subsets is fixed by the fourteen logically possible combinations of five basic values and three sources of deprivation. The reliability score calculated over all sixty-six problem area categories is .81.

At present our interest centers on two types of problems associated with the two putative systems selected for investigation as exemplars of the multiple systems model. The substantive content of these two problem areas was described in some detail in Chapter II.
Each event located in these categories passes through a decision tree of the sort depicted in Figure 4. The boxes in the center of the diagram signify choice points represented by three problem area dimensions. As indicated earlier, the initial coding decision is made with regard to the basic value(s) affected by the problem. At this point our concern focuses on the security/military and economic wealth categories which Hermann and Coate (1979: 115) characterize as follows:

- **Security/Military - Physical Safety**: The desire to enjoy physical safety free from organized violence to persons, property, or national institutions.

- **Economic Wealth**: The desire to enjoy physical goods and to maintain and promote the institutions and arrangements (e.g., currency) pertaining thereto.

The second decision point categorizes the source of the perceived deprivation into one of three broad clusters. Although different in basic value, both of our selected problem areas are attributable to governmental—rather than nongovernmental or naturally occurring—sources of deprivation. The final coding decision applies a substantively oriented classification of the problem within the previously established confines of governmentally inspired deprivations of security or economic values. On the military/security side this choice is made from among nine problem categories including the conflict negotiation or military settlement category described earlier. With respect to economic values, on the other hand, the coder is faced with only six problem categories. For the purposes of this investigation, two of these—commodity transactions and trade
FIGURE 4
Problem Area Decision Tree

Source: Adapted from Hermann and Coate (1979).
agreements or associations—were combined to form a single problem area dealing with international trade.

The datamaking operations reviewed in this part of the chapter share a common origin in the controlled application of specified coding procedures to characterizations of discrete governmental actions collectively designated foreign policy events. These operations enrich the original characterizations by providing explicit evaluations of certain properties that are apt to vary from event to event. However, we have indicated that it is necessary to move beyond the single event—that is, beyond the individual behavioral act—if we are to detect the kind of sustained effects posited by our stratification hypotheses. The next section takes up the observation of behavioral patterns and in so doing advances our discussion to the measurement stage of the events research process.

Patterns of Foreign Policy Behavior

In Chapter II we stated that an individual behavioral act could be discretely located in time and space whereas a behavioral pattern involved extension through one or both of these dimensions. A central theme in the present chapter has been that a discrete action can be operationally represented as a foreign policy event. We are now in a position to carry this correspondence a step further by proposing that behavioral patterns be operationalized as aggregations of events and that the behavioral properties discussed in the previous section be indicated by an aggregate summary of individual coding decisions recorded on an event by event basis.
In general, monadic patterns are operationally defined as the aggregate behavioral output of a governmental actor over a specified period of time with no reference to whom those behaviors are directed. Discrete event properties expressed in monadic format are typically represented by summary indices that reflect something about how those properties are distributed across the behavioral aggregate. In cases where a coding procedure is based on a nominal classification, such as with the initiative-reactive and unilateral-multilateral dimensions of behavioral autonomy, a simple and effective summary index is the proportion of aggregated events in each of the nominal categories. Similarly, dyadic patterns are operationalized as the aggregate behavioral output of a governmental actor over a specified period of time toward a particular foreign recipient. A single event can involve more than one recipient and therefore can contribute to the dyadic aggregate of more than one actor-recipient unit; however, an event counts toward the dyadic aggregate only if the recipient in question occupies a role involving it in the substance of the behavior. Representation of event properties at the dyadic level follows the same proportional formula applied in the monadic case. We can illustrate this procedure by examining the operational distinction between collaborative and multilateral dyadic behavior. The former, which is based on the recipient role variable, is measured as the proportion of events in the dyadic aggregate in which the recipient actually participates in the action along with
the actor. Multilateral behavior, on the other hand, is the proportion of the dyadic aggregate in which the actor joins with any government—including, but not limited to, the dyad's recipient.12

Thus far our description of the two behavioral patterns has been cast in the conventional mold of the single system perspective. The problem-specific patterns invoked by the multiple systems approach are operationally defined as the aggregate behavioral output of a governmental actor within a designated problem area. The problem area designation is to insure the substantive integrity of behavioral patterns imputed to the multiple systems model. But problem area specialization has the concomitant effect of reducing the number of events represented in the aggregate pattern. Moreover, this diminution of events can be of sufficient magnitude to undermine our confidence in any subsequent analyses of problem-specific behavioral patterns. The remainder of this section examines this diminution of events and its consequences for the present study.

Underlying our discussion has been an assumption that behavioral patterns provide a fairly accurate characterization of a government's behavioral tendencies or dispositions. Our confidence in this assumption is, in part, a function of the number of events comprising the pattern—the more events in the pattern the more likely it is to reflect genuine tendencies rather than merely a few deviant cases. National governments are, for the most part, multipurpose international actors in that over any appreciable period of time they will distribute their foreign activities across a
variety of substantive areas though typically not in equal amounts. Like any taxonomy or classification, as substantive categories of foreign policy behavior become increasingly specialized one finds that they apply to fewer and fewer actual cases. Furthermore, as the number of cases—in this instance, events—diminishes so does the stability of any pattern based on those events, particularly when relative frequencies or percentages are involved. In Galtung's (1967: 189) terminology, there is a "law of large numbers" which can lead to great fluctuations in proportional measures when the base figure is small, but promises a certain stability when the base is reasonably large.

Of course, the number of foreign actions taken by a government in a particular problem area is beyond the researcher's direct control; however, the number of events minimally required by the operational definition of a behavioral pattern is not. By setting a threshold on the minimum number of events necessary for a behavioral pattern we can avoid unwitting analysis of clearly unstable or suspect patterns. On the other hand, disregarding any putative patterns that fail to embrace the requisite number of component events depletes the number of dyadic or monadic cases available for subsequent analysis. Faced with this situation, the analyst must strive to balance the need for representative and stable behavioral patterns with the countervailing, but no less important, need for sufficient cases to achieve replicable and generalizable results.
Having sketched in the general outlines of this dilemma, let us now be more specific about its implications for the data at hand.

What is a reasonable threshold of events for monadically organized patterns of problem-specific foreign policy behavior? An answer to this question must take account of two basic parameters: the number of events comprising a proposed threshold and the corresponding number of monads that qualify as legitimate cases for analysis. Figure 5 illustrates in graphic fashion the close interdependence of these two parameters for the military and trade problem areas. The figure's horizontal axis displays a range of possible event thresholds and its vertical axis lists the number of monadic cases that remain at a given threshold level. We should add that the monadic cases represented in the figure incorporate events from the entire thirty month time frame for which data were available.

Certainly the most prominent aspect of Figure 5 is the marked difference in the composition of trade and military patterns. As expected, both problem areas exhibit a steadily declining number of monadic cases as the event threshold rises, but there the similarity ends. The relative position of the two lines and the difference in their slopes indicate that most of our sample governments act on trade oriented problems much more frequently than on problems related to the negotiation and settlement of military disputes. Intuitively at least this is an encouraging finding since trade concerns are squarely within the realm of normal international relations whereas military settlements, though not uncommon, nonetheless do represent a deviation from conventional modes of diplomacy.
FIGURE 5

Event Composition of Problem Specific Monads
This paucity of military settlement events is less desirable when it comes to the matter of choosing an appropriate event threshold. Note that a minimum of only five events is met by fewer than twenty governmental actors, a minimum of three by fewer than thirty. Somewhat reluctantly, we finally settled on the latter threshold of three events in order to maximize the number of cases included in the analysis of the four monadic hypotheses. Our reluctance was motivated by the highly volatile character of patterns derived from such small aggregations of events. Nevertheless, with a sample of only thirty-eight governmental actors, it seemed advisable to emphasize the number of cases rather than the number of events. Moreover, so that we do not arrive at erroneous conclusions on the basis of unstable or idiosyncratic patterns, analysis of each hypothesis will be replicated several times at higher event thresholds. Should we discover any appreciable differences between our primary analysis and these replications, that fact will be reported. For each of our sample governments, Table 14 delineates the number of military and trade events comprising the pattern aggregates.

Let us redefine our query to consider what event threshold would be appropriate to dyadic patterns of problem-specific international behavior. Figure 6 indicates that both problem areas experience a decreasing rate of decline in the number of remaining dyadic cases associated with increasing threshold levels. The figure also reveals some intriguing characteristics about how governments deal with different types of substantive problems. At a level
TABLE 14

Number of Events Comprising Monadic Patterns

<table>
<thead>
<tr>
<th>Governmental Actors</th>
<th>Number of Events</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trade</td>
</tr>
<tr>
<td>Belgium</td>
<td>84</td>
</tr>
<tr>
<td>Canada</td>
<td>53</td>
</tr>
<tr>
<td>Chile</td>
<td>32</td>
</tr>
<tr>
<td>China</td>
<td>22</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>12</td>
</tr>
<tr>
<td>Cuba</td>
<td>15</td>
</tr>
<tr>
<td>Czechoslovakia</td>
<td>27</td>
</tr>
<tr>
<td>East Germany</td>
<td>23</td>
</tr>
<tr>
<td>Egypt</td>
<td>21</td>
</tr>
<tr>
<td>France</td>
<td>123</td>
</tr>
<tr>
<td>Ghana</td>
<td>16</td>
</tr>
<tr>
<td>Guinea</td>
<td>8</td>
</tr>
<tr>
<td>Iceland</td>
<td>10</td>
</tr>
<tr>
<td>India</td>
<td>27</td>
</tr>
<tr>
<td>Israel</td>
<td>14</td>
</tr>
<tr>
<td>Italy</td>
<td>87</td>
</tr>
<tr>
<td>Ivory Coast</td>
<td>15</td>
</tr>
<tr>
<td>Japan</td>
<td>50</td>
</tr>
<tr>
<td>Kenya</td>
<td>5</td>
</tr>
<tr>
<td>Lebanon</td>
<td>4</td>
</tr>
<tr>
<td>Mexico</td>
<td>21</td>
</tr>
<tr>
<td>New Zealand</td>
<td>25</td>
</tr>
<tr>
<td>Norway</td>
<td>30</td>
</tr>
<tr>
<td>Philippines</td>
<td>6</td>
</tr>
<tr>
<td>Poland</td>
<td>34</td>
</tr>
<tr>
<td>Soviet Union</td>
<td>57</td>
</tr>
<tr>
<td>Spain</td>
<td>13</td>
</tr>
<tr>
<td>Switzerland</td>
<td>24</td>
</tr>
<tr>
<td>Thailand</td>
<td>4</td>
</tr>
<tr>
<td>Tunisia</td>
<td>16</td>
</tr>
<tr>
<td>Turkey</td>
<td>21</td>
</tr>
<tr>
<td>Uganda</td>
<td>13</td>
</tr>
<tr>
<td>United States</td>
<td>82</td>
</tr>
<tr>
<td>Uruguay</td>
<td>20</td>
</tr>
<tr>
<td>Venezuela</td>
<td>24</td>
</tr>
<tr>
<td>West Germany</td>
<td>108</td>
</tr>
<tr>
<td>Yugoslavia</td>
<td>24</td>
</tr>
<tr>
<td>Zambia</td>
<td>3</td>
</tr>
</tbody>
</table>

*A null entry indicates that the government engaged in too few military settlement events to yield a monadic pattern.*
FIGURE 6
Event Composition of Problem Specific Dyads
of one event the number of trade dyads surpasses the number of military dyads by more than two to one. This ratio shrinks dramatically at the two and three event levels until, at four events and beyond, the ratio is reversed with the number of military dyads exceeding the number of trade dyads by a slight margin. The overall pattern suggests that trade issues are of more widespread concern but involve less sustained interaction than military settlement questions.

Analysis of the dyadic hypotheses was carried out on patterns comprised of four or more events, a threshold level yielding upward of fifty dyadic cases for each problem area. Beyond a threshold of four events the number of trade dyads declines rather quickly, with fewer than thirty remaining at a threshold of five and fewer than twenty at a threshold of six. Moreover, anything less than four seemed unwise in view of the aforementioned shift in the ratio of military to trade dyads. Here, as with the monadic analysis, we will follow a replication strategy to provide a check on the stability of the four event threshold level. Not all of the thirty-eight governmental actors are sufficiently involved in trade or military settlement problems to surpass the four event level in any of their bilateral relationships. Table 15 identifies the actors represented in one or more dyads in either problem area. It is worth pointing out that of the twenty-six governments listed in Table 15, only eight engage in sustained dyadic behavior in both problem areas.
TABLE 15

Representation of Governmental Actors in Dyadic Patterns

<table>
<thead>
<tr>
<th>Governmental Actor(^a)</th>
<th>Number of Dyadic Patterns</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trade(^b)</td>
<td>Military(^c)</td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>2</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>4</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Chile</td>
<td>3</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>-</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Cuba</td>
<td>-</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Czechoslovakia</td>
<td>-</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>East Germany</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Egypt</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>6</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Ghana</td>
<td>-</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>2</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Israel</td>
<td>-</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>3</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>5</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>New Zealand</td>
<td>1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Philippines</td>
<td>2</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td>2</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Soviet Union</td>
<td>4</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>-</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Tunisia</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Turkey</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Venezuela</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>West Germany</td>
<td>4</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Total 53 70

\(^a\)The following twelve governments engaged in too few events to yield a dyadic pattern in either problem area: Costa Rica, Guinea, Iceland, Ivory Coast, Kenya, Lebanon, Mexico, Norway, Spain, Uruguay, Yugoslavia, Zambia.

\(^b\)A null entry indicates that the government engaged in too few international trade events to yield a dyadic pattern.

\(^c\)A null entry indicates that the government engaged in too few military settlement events to yield a dyadic pattern.
To this point our concern has focused on the number of events comprising problem-specific pattern aggregates. But, regardless of problem area, national governments do not distribute their activities in equal amounts across all potential foreign recipients—just as they do not distribute behaviors equally across all problem areas. This suggests, and Figure 7 confirms, that there is wide variation in the number of events representing dyadic configurations of overall system behavior. The changing rate of decline in the number of overall dyadic cases is strikingly similar to that of the problem-specific dyads in Figure 6. First there is a sharp decline with over half of the 2000 plus cases involving fewer than four events, then the rate levels off so that more than 200 cases remain at a threshold as high as twenty events. The more important characteristic of Figure 7, however, is the magnitude of the vertical axis scale factors used to express the remaining number of overall dyadic cases at various threshold levels. With this relatively abundant supply of overall system dyads we were able to impose a ten event threshold and still retain more than 400 cases for subsequent analysis. The ten event criterion was chosen somewhat arbitrarily, subject only to the constraint that it be considerably larger than the four event level used for problem-specific dyads. Since overall system behavior is by definition more substantively heterogeneous than problem-specific behavior, we reasoned that its accurate representation as a behavioral pattern would require a larger event base than for the more homogeneous behavior occurring within a single problem area.
MINIMUM EVENTS FOR DYADIC PATTERN

FIGURE 7
Event Composition of Overall System Dyads
Summary and Evaluation

This completes the first of the two tasks set forth in the opening paragraphs of this chapter. We have provided a moderately detailed account of how certain hypothesized behavioral effects of the stratification process have been operationalized through the use of foreign policy event data. These operational procedures were broken down into separate stages dealing with the identification of events from public source material, the classification or coding of identified events according to relevant behavioral characteristics, and the aggregation of events into behavioral patterns. Before concluding this part of the chapter let us review the central elements of these procedures with specific reference to the seven measurement criteria introduced in Chapter III. To recapitulate, the seven criteria are relevance, sensitivity, reliability, validity, comparability, availability, and simplicity. We will consider each in turn.

Are the operationalized measurements relevant to each governmental actor or dyadic unit for which they were obtained? We would argue that relevance is not an issue with respect to either the assignment of behavioral properties to particular events or the aggregation of events to monadic or dyadic patterns. The relevance criterion is less easily dismissed when used to evaluate our procedures for including actors and dyads in the problem-specific systems. In other words, is it reasonable to assume that a problem system is relevant to an actor or a dyadic unit just because its
pattern has the requisite number of events to surpass the threshold? Contrariwise, are we justified in excluding an actor or dyad because it fails to meet the requisite number of events? These questions are unsettling, in part because of the arbitrariness in our choice of threshold levels and in part because of our data's limited coverage of sample quarters rather than the entire ten year period. In sum, the relevance criterion leads to the much more fundamental issue of what determines membership in a problem system. We have relied on the event threshold as an operational expedient but the issue remains a conceptual loose end for the multiple systems approach.

In principle, the sensitivity of our behavioral measurements—that is their ability to discriminate among observations—is unexceptionable. Amount of activity and scope of action are operationalized as the incidence of events and the count of distinguishable governmental recipients, respectively. The other behavioral properties are represented as percentages of events in the pattern aggregate assigned to distinct nominal categories. In practice, however, the sensitivity of the percentagized measures tends to degenerate when only a very few events comprise the pattern aggregate. Percentages based on a small number of events may appear to be more sensitive than they really are. Moreover, if the sensitivity of a measurement procedure can vary from observation to observation depending on the base number of events then the comparability of the measures also becomes problematical.
All of the difficulties mentioned thus far are related in one way or another to the number of events coming together to define a behavioral pattern. They can also be interpreted as problems of internal consistency or reliability. Let us move away from this issue to touch upon another aspect of reliability more commonly associated with event data: the degree of intercoder agreement. In general, the reliability scores measuring the proportion of agreement beyond that expected from chance alone were found to be in the upper seventies or higher, a range fairly typical of this type of data. We should reiterate, however, that a score calculated over a full coding scheme does not necessarily reflect the reliability of individual categories within that scheme.

Do our operational procedures succeed in measuring what we intend them to measure, that is, are they valid? Earlier in the chapter we mentioned some general problems of source validity that have been of long standing concern to event data researchers. Here we will consider some implications of these validity issues with respect to our particular behavioral concepts and their operationalizations.

Perhaps the most serious validity problem concerns our inferred measurement of foreign activity as the number of events attributed to a particular actor or dyadic unit. Underlying this operational procedure is an implicit assumption that each country and region of the world has approximately the same proportion of its foreign policy behaviors reported in public news sources. Some researchers, however,
have taken the contrary position that news gathering and reporting are so heavily biased in favor of elite nations and nations geographically or culturally proximate to the news organization that the number of identified events may be a better indication of a nation's "newsworthiness" than of its government's international activity. The truth of the matter probably lies somewhere between these two extremes.

Assuming, then, that there is some systematic error biasing our measurement of foreign activity, what will be the effect on our analysis? It is likely that the consequences will be more severe for the monadic hypothesis since the direction of the bias—that is, toward overreporting of elite or powerful nations—parallels the direction of the predicted relationship. The dyadic hypothesis, on the other hand, predicts foreign activity on the basis of positional proximity. In this case the potential effects of news selection bias are more likely to result from overreporting of nations geographically or culturally proximate to the "home" country of the news organization. Since our data were derived from a global compilation of news sources this type of bias may be mitigated to some degree, though it will not be eliminated entirely. The more important point, however, is that in the dyadic analysis the effect of reporting bias will not necessarily overlap the hypothesized effect of the stratification process. In other words, the systematic error of biased source coverage may behave as though it were random error in the analysis of dyadic cases.
Another type of suspected source bias concerns the quality of reported events rather than their quantity. Some researchers have maintained that news services customarily emphasize dramatic or conflictual actions at the expense of routine or cooperative ones (see Galtung and Ruge, 1965). How might this sort of bias influence the content of our data? In our view, the most likely possibility involves the two dimensions of behavioral autonomy. It could be, for example, that multilateral behaviors are disproportionately reported simply because they tend to be more visible than unilateral activities. Furthermore, if we take into account the aforementioned cultural and elite nation biases then we would expect this overrepresentation of multilateral behavior to be particularly acute for governments located toward the bottom of the structural hierarchy since their actions might otherwise go unnoticed. Unfortunately, the monadic level effect of this combination of measurement errors again overlaps the hypothesized effect of the stratification process. Biased coverage of elite or culturally proximate nations may also influence the initiative-reactive quality of our behavioral patterns. For instance, it is possible that many Western news services treat Third World reactions aimed at industrialized countries as more significant or more "newsworthy" than other kinds of Third World behaviors. Clearly this sort of bias would lead to an overreporting of reactive behaviors toward positively distant recipients, a result that once again parallels the expected effect of the stratification process.
The remarks contained in the preceding paragraphs by no means constitute an exhaustive survey of source validity issues; rather, they were intended merely to highlight a few of the areas in which source bias could have some noticeable and potentially damaging consequences for our event-based measurements of foreign policy behavior. We would argue, however, that the ultimate significance of these difficulties must be assessed within the overall context of our inquiry and not with respect to any single behavioral property or hypothesis. In this regard the following observations deserve close consideration.

First, the extent and direction of bias in event data sources cannot be empirically demonstrated with any certainty since there exists no unbiased "universe of events" to serve as a criterion. It is largely for this reason that source bias research has produced such mixed results (see note 13). But, having stated this caveat, let us also mention that one source validity study conducted on our data base concluded that "the results clearly do not support the 'common assumptions' about the distortions inherent in newspaper reportage..." (Salmore and Butler, 1978: 101). Second, it is essential that we not lose sight of the primary objective of our investigation, which is to compare and evaluate the single system and multiple systems models of world politics. Although our inquiry employs and hypothesis testing strategy, the hypotheses themselves and the theory in which they are embedded are of ancillary importance.
Our main task is to uncover a discernable pattern of empirical results across the various hypotheses that will permit inferences regarding the mode of systemic organization overlaying contemporary intergovernmental relations. Thus, the overall conclusions of our study will not depend on the analysis of any single hypothesis or behavioral property. Finally, it is worth pointing out that our research design is unusually resistant to moderate levels of systematic measurement error so long as there is no appreciable variance in the severity or direction of error across the three systems selected for investigation. If this error remains fairly constant—and there is no empirical or speculative research indicating otherwise—then it cannot be a major factor underlying any differences between the military and trade analyses on the one hand and the overall system analysis on the other.

These observations suggest that the source validity problems outlined above are actually less troublesome than they might first appear. This is not to say that source bias is completely absent from our behavioral measurements, nor does it imply that our study is totally immune to the distorting effects of such bias. Although we can be reasonably confident in our overall conclusions regarding the systemic configuration of world politics, we are less certain about the accuracy of subsidiary conclusions dealing with individual hypotheses or the stratification process as a whole.
The two remaining measurement criteria, availability and simplicity, require only a few brief comments. From all that has been said thus far regarding the stages of the events research process, it should be clear that our operational procedures are not about to receive high marks for their simplicity. Nor are the behavioral measurements readily accessible for governments or time periods other than those identified in Tables 11 and 12 at the beginning of the chapter. We should add, however, that the limited availability of these measurements is solely a function of the time and resources involved in the identification and coding of foreign policy events.

Analysis Design

To this point, discussion of the empirical side of our inquiry has emphasized the many operational and measurement issues germane to the structural and behavioral concepts explicated in Chapter II. Now that these concepts have been operationalized, we may turn to an examination of the procedures used to empirically test the hypothesized relationships between positional structure and patterns of governmental behavior. Although there are numerous techniques for delineating bivariate association, simple regression analysis was judged most appropriate to the requirements of the present study. The remainder of this chapter presents the rationale for this choice and then details some technical considerations involved in applying the regression model to the stratification hypotheses.
To begin let us review the purposes of this analysis within the context of the overall research strategy outlined in Chapter I. We have assumed, first, that in any international system governmental behavior will be constrained or influenced in various ways by the prevailing system structure and, second, that a priori knowledge of these influence relationships can serve as a benchmark for evaluating the actual relationships discerned from empirical analysis. Furthermore, if several such analyses are carried out under the assumptions of the single system and multiple systems models, then by comparing results across models it would be possible to determine which of the two competing system constructs better approximates its intended empirical referent. Thus, our research strategy imposes two requirements on our choice of analysis technique: first, it must provide some indication of how well the predicted relationship fits the data; and, second, it must allow for comparisons across subpopulations, that is, across the different models of systemic organization. A third requirement that is operative under any research strategy links the choice of analysis technique to the scale properties of one's data. This constraint deserves brief comment before we proceed to an examination of alternative procedures.

The properties inherent in the various scales or levels of measurement are detailed in virtually every introductory statistics text and therefore need not be recounted here. What is needed,
however, is an explicit statement identifying the scale properties of our structural measures. In the previous chapter we recorded our preferences for the use of ordinal rankings rather than actual scores on the grounds that, first, there is probably less error associated with the ranks than with the intervals between scores and, second, our theoretical formulation emphasizes the relative positions of countries and not the absolute distances separating them. Does this imply that only ordinal level techniques are appropriate for the analysis of these data? In this case, it probably makes little difference if ordinal measures are given an interval interpretation for the purpose of applying a more powerful analysis technique. There is evidence to suggest that the interval assumption can be violated with relative impunity under most circumstances provided that: (a) the ordinal measure is monotonically related to the "true" scale and (b) there are a large number of ordinal categories (Bohrnstedt and Carter, 1971; Labovitz, 1970). The large number of unique (untied) ranks in our structural measures is especially encouraging since Labovitz (1970: 521) has found that "...the greater the number of ranks (N), the greater the confidence in assigning an interval scoring system to ordinal data."16

The most elegant and simple method for specifying a bivariate relationship is in symbolic form as a mathematical equation. Under ideal conditions, the ten stratification hypotheses which constitute our set of benchmark expectations would have been expressed mathematically. Since every mathematical equation can be represented as
a geometrical curve, it would then be possible to test the hypotheses by examining the squared error or "residuals" from the curves of the theoretical relationships. The most straightforward method for accomplishing this would involve correlational analysis of theoretically predicted and observed values on the behavioral measures. Unfortunately, neither our knowledge base nor our stratification theory have evolved to the point where the hypothesized relationships can be stated in precise mathematical terms. In fact, as the hypotheses now stand they tell us nothing about the functional form of the relationships other than their direction. Without prior knowledge of the relationships' functional form we must turn to more conventional analysis techniques.

Perhaps the simplest and most widely used method for exploring the relationship between two variables is by means of a contingency table or a cross-classification. Actually, this rubric houses a large and diverse family of techniques ranging from simple percentages to ordinal measures of association such as the taus and Goodman and Kruskal's gamma. The usual starting point for this type of analysis is a two-way table displaying the joint frequency distribution of cases over the categories of two variables. Clearly, such a tabular arrangement of our data would be of unmanageable proportions without some regrouping of the values into a smaller set of summary categories. For example, our measures of structural position might be redefined to differentiate only among high, medium, and low ranks. Although this procedure has been used by some stratification
researchers (for example, Galtung, 1966c; Gleditsch, 1967) it necessarily involves an enormous loss of information. A less drastic approach leading to an entirely different set of statistical techniques would involve regrouping of only the structural measures, leaving the behavioral measures intact. This formulation, which resembles the familiar analysis of variance model, would allow comparison of behavioral means and variances across the summary categories of structural position. Since only the structural measures are transformed the loss of information is substantially reduced. Nevertheless, we would contend that both designs have several important disadvantages.

In the first place, it is not always clear just where the lines should be drawn between the regrouped categories of our measures. This is an important consideration since manipulation of the cutting point could conceivably alter the results of our study. Second, regrouping of the structural measures severely limits our ability to investigate the dyadic hypotheses where the emphasis is on the relative proximity of structural positions. Finally, the loss of sensitivity entailed by a regrouping procedure seems particularly unwise given the close congruence of the structural measures demonstrated in the previous chapter. So long as we maintain a positional interpretation of system structure it is vital that we preserve the structural distinctions between systems as a basis for differentiating the two models of systemic organization. At this point we should qualify these remarks by stating that regrouping procedures can be a
useful analytic tool under certain conditions, as we will see further on in this section. However, any regrouping during the initial phase of the design does not seem appropriate to the needs of the present inquiry.

Another common technique for examining bivariate relationships is the product-moment correlation or Pearson's r. Correlational analysis does not require regrouping of either variable and is sensitive to more of the information contained in the data than the techniques described above. Moreover, the correlation coefficient has a straightforward intuitive interpretation as a measure of association since its square is equal to the proportion of total variation in one variable explained (or accounted for) by the other. Despite these advantages, correlational analysis by itself is not a suitable means for implementing our research strategy since it deals exclusively with determining the strength of a relationship. The more important question, however, concerns the relationship's form and is in the domain of regression analysis.

Regression is a general technique for determining the magnitude of change expected in a dependent variable given a specified change in one (or more) independent variable(s). Because there is this clear distinction between the independent and dependent variables, regression analysis is particularly appropriate for investigating hypotheses that are given a causal interpretation. In other words, regression preserves the asymmetry between cause and effect whereas correlation merely indicates the degree of joint variation. For
the present study, this means that regression can be used to determine the size or potency of the hypothesized structural effects on the designated patterns of behavior within each system relative to those in other systems. We must caution that the results of such analyses are for comparative purposes only and are not to be construed as estimates of the causal paths linking positional structure and behavior. Path analysis and other causal modeling techniques are appropriate only if it can be assumed that within reasonable limits all pertinent causal factors have been incorporated into the model under consideration (Asher, 1976; Blalock, 1964). Our exclusive focus on systemic effects means that such an assumption, even as only a first approximation, is quite out of the question insofar as the present inquiry is concerned. However, if we assume that the cumulative effects of extraneous influences are about the same in each of the systems chosen for analysis, then regression results can be compared across systems to indicate the relative impact of system structure.

We will discuss some of the more technical aspects of regression analysis in a moment, but first we must be a bit more precise about the kinds of comparisons that will enable us to assess the relative utility of the single system and multiple systems models of world politics. For each stratification hypothesis, two types of evidence will be examined. The first requires a comparison of regression results obtained from within each of the three systems. These results will be considered supportive of the single system
configuration if the appropriate measure of overall structure has a greater impact on overall behavior than both problem-specific structures have on their respective patterns of behavior. Contrariwise, support will be indicated for the multiple systems model if both problem structures prove more potent than overall systems structure. If neither of these standards is met, that is, if overall structure has a greater impact than only one of the problem-specific structures, then our results will be considered ambiguous with regard to mode of systemic organization.

Although our investigation centers on only three systems, the comparison will involve four separate regression results—an anomaly that arises from our decision to retain two measures of international trade structure. In the event that these two measures lead us to different conclusions when compared to overall structure, our final determination will be based exclusively on the performance of military structure vis-a-vis overall structure. In other words, our findings will be deemed supportive (rather than ambiguous) even if this is indicated by only one of the trade structure measures. Admittedly, this is a lenient decision rule but it is not an unreasonable one given that the two weighted share indices differ on more than half of their total rank variation (see Table 9).

The second type of evidence is derived from a comparison of results obtained when trade and military behavior patterns are regressed on overall structure as well as their respective problem structures. This comparison is designed to test the single system
model's presumption that overall capabilities are easily transferred from one issue to another and, hence, that the systemic effects of overall structure are congruent across issue areas. Our criteria for interpreting these results are the same as those outlined above. If overall structure appears to have a greater impact on trade behavior than trade structure and a greater impact on military behavior than military structure, then support for the single system model is clearly indicated. On the other hand, the results will be considered supportive of the multiple systems construct if both problem-specific structures do better than the overall structure. As before, mixed or ambiguous results are not deemed supportive of either model.

Taken together, these two types of comparisons provide a convenient framework for organizing our results into a coherent appraisal of the two systemic constructs. Figure 8 juxtaposes the possible outcomes of each comparison to reveal nine permutations containing five distinct evaluative categories for classifying the results on each hypothesis. The parenthesized values in the cells and margins represent the probabilities with which these outcomes should occur by chance alone. Let us suppose that our structural and behavioral measures are simply random variables. Under these conditions we would expect overall system structure to be a more potent regressor than problem-specific structure about half of the time and, likewise, problem-specific structure to be more potent than overall structure about half of the time. Simple probability theory tells us that a comparison between overall structure and two problem structures would entail four possible outcomes (overall superior to both,
**TYPE I COMPARISON**

<table>
<thead>
<tr>
<th>Supportive of Single System</th>
<th>Mixed Support</th>
<th>Supportive of Multiple Systems</th>
<th>Outcome Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong Support for Single System (.141)</td>
<td>Weak Support for Single System (.094)</td>
<td>Ambiguous (.141)</td>
<td>(.375)</td>
</tr>
<tr>
<td>Weak Support for Single System (.094)</td>
<td>Ambiguous (.063)</td>
<td>Weak Support for Multiple Systems (.094)</td>
<td>(.25)</td>
</tr>
<tr>
<td>Ambiguous (.141)</td>
<td>Weak Support for Multiple Systems (.094)</td>
<td>Strong Support for Multiple Systems (.141)</td>
<td>(.375)</td>
</tr>
</tbody>
</table>

**Outcome Probabilities**

| (.375) | (.25) | (.375) |

**FIGURE 8**

Design for Evaluating Systemic Constructs
overall inferior to both, overall superior to problem structure A only, and overall superior to problem structure B only) each of which would occur with about the same probability. According to the criteria specified earlier, two of these permutations would be collapsed into a single "mixed" category thus leaving only three distinct outcomes, one of which would occur about twice as often as either of the others. The marginal probabilities in Figure 8 were calculated according to these same principles, although the actual procedure was modified slightly to allow for the presence of two trade structure measures and our special decision rule for evaluating them.²⁰

Notice that under the random variable assumption the results of the two comparisons are statistically independent, that is, the outcome of one comparison in no way prejudices the outcome of the other comparison. It is well known that the probability of the joint occurrence of two statistically independent events is equal to the product of the probability of one event times the probability of the other. Accordingly, the joint probabilities displayed in the cells of Figure 8 represent the product of the probabilities of the two outcomes occurring separately. These joint probabilities can then be summed over the five evaluative categories to indicate the likelihood of obtaining each outcome purely by chance:

- **Strong Support for Single System** = .141
- **Weak Support for Single System** = .187
- **Ambiguous** = .344
- **Weak Support for Multiple Systems** = .187
- **Strong Support for Multiple Systems** = .141
So there is no misunderstanding, we should point out that it is definitely not our intention to treat these probabilities as an expected baseline in any sort of formal statistical test of our results. What, then, is the benefit of knowing these probability values? For one thing, the symmetry of the distribution tells us that our design is not inherently biased toward either of the two system constructs. In addition, the values themselves provide a rough indication of the severity of our design. Since an ambiguous outcome is more than twice as likely as either of the strongly supportive outcomes we can regard our study as a reasonably severe test of the two system models.21

Technical Considerations

Having presented the broad outlines of the research design, we may now turn to a more detailed discussion of the regression model and its application to the ten hypotheses enumerated in Chapter II. We will consider two issues relating to regression analysis in general—the use of standardized versus unstandardized coefficients and the assumption of linearity—as well as some problems dealing with the particular hypotheses under investigation.

Conventional regression techniques assume that a relationship between two variables can be adequately depicted by a straight line or, more formally, by a linear model. The mathematical representation of a linear relationship takes the following familiar form:

\[ Y = a + bX \]

Although there are several kinds of nonlinear relationships, the
linear model has been found to be a fairly good approximation to the true functional form in many research situations, particularly those in which the degree of fit is not too exact. Given our measurements' susceptibility to error and the extraneous factors that are bound to be influencing behavior, it is virtually certain that the empirical relationships examined here will reveal only a moderate degree of fit at best. Does this suggest that we should settle on the linear model as a satisfactory approximation? More to the point, can an a priori assumption of linearity have a significant impact on the overall conclusions of our study?

In general, linear analysis of nonlinear data will produce a distorted view of the actual relationship by underestimating the change in the dependent variable associated with a change in an independent variable. Setting aside the effects of extraneous factors, this means that linear analyses of our data will reflect the structural measures' impact on the various behavioral properties plus some decrement owing to the true relationships' departure from linearity. Thus, any differences in individual regression results could be due to differential effects on nonlinearity. It may turn out upon investigation that the empirical relationships are indeed linear in form or that the effects of nonlinearity are the same across all of the paired comparisons called for in our design; in either case, the linearity assumption would be unproblematical. Nevertheless, these are empirical questions and should
not be treated as untested assumptions. Accordingly, we will
examine the scatter plots of our data for evidence that the
relational forms are other than linear. If a nonlinear model is
indicated we will attempt to "linearize" the relationship by
applying an appropriate transformation to the data (Hilton, 1976;
Neter and Wasserman, 1974). 22

Another issue that deserves brief comment concerns the use of
standardized or unstandardized regression coefficients. As a general
rule, standardized coefficients are preferable if one's purpose is
to compare the relative importance of variables that are measured
on different units. On the other hand, because unstandardized
coefficients are unaffected by differences in variances they are
the more appropriate choice for making comparisons across subsets
of data, provided that the variables are measured on comparable
scales. The relationship between the two coefficients is illust­
trated in the following equality:

\[ B_{yx} = b_{yx} \left( \frac{s_x}{s_y} \right) \]

where \( B \) is the standardized coefficient, \( b \) the unstandardized
coefficient, and \( s \) represents the standard deviation. Although the
standardized coefficient adjusts for differences in measurement
units, this adjustment is made in terms of a quantity (the ratio of
the standard deviations) that is not invariant across samples or
subsets of data.

Several features of our design bear upon the choice of coef­
ficient. In the first place, the design requires comparisions of
regression results across different systems (type I comparisons)
and within the same systems (type II comparisons). In all but two cases these comparisons are between results involving variables measured on identical scales: the structural indicators are based on each country's rank position and many of the behavioral measures are prestandardized to percentages. The two exceptions, which will be discussed more fully later, arise in type I comparisons involving amount of foreign activity and scope of action. Another point to keep in mind is that our regression analyses in the two problem-specific systems will be performed on different subsets of the cases contained in the overall system. Since the sample of governmental actors and intergovernmental dyads varies from system to system, the standard deviations of the structural and behavioral measures will also vary. Thus far, all indications point toward the unstandardized coefficient as the appropriate choice for making the comparisons encompassed by our design.

Now let us consider the two exceptions mentioned earlier. Even though amount of foreign activity and scope of action are measured in the same units across the three systems, they are not measured on the same scales. Because both measures record raw frequencies rather than percentages they are directly affected by the number of events aggregated to form the behavioral patterns. Moreover, because the events comprising the problem system patterns represent only a fraction of the events in the overall system pattern, it is unreasonable to expect these variables to have comparable scales.
Hence, for analyses of these two behavioral measures the standardized coefficient is not only preferable, it is absolutely essential. Of course, these remarks apply only to type I comparisons made across systems since type II comparisons involve only the problem-specific behavioral measures.

In the end we elected to deal with the standardization issue by following a mixed strategy. Standardized coefficients will be used in those few cases that require adjustment for scale incomparabilities, otherwise our comparisons will be based on unstandardized coefficients since they are immune to the effects of different variances across subsets of cases. It will be interesting to see if the final outcome of our study is affected by our decision to concentrate on the unstandardized coefficient. We will be able to answer this question even though we do not plan to report the standardized coefficients directly. In addition to the unstandardized coefficient, which indicates how two variables are related, we will also report the coefficient of determination, which measures the amount of spread about the regression line and tells us how well two variables are related. Since the standardized coefficient is equal to the square root of the coefficient of determination, it will be an easy matter to determine the extent to which the two coefficients lead to parallel conclusions.

Because our overall conclusions will be derived from comparisons of relative magnitudes, it is appropriate to consider how small an observed difference between coefficients can be interpreted with
any degree of assurance as a meaningful difference in terms of the systemic representation of structural effects. Put more simply, how much of a difference actually makes a difference?

Ordinarily one might address this question by constructing a confidence interval to formally test the difference (or equality) of two regression coefficients. This test procedure is a straightforward extension of the technique used to calculate a specific confidence interval around a single regression coefficient. In this case the observed regression slope, $b$, is assumed to be a point estimate of an unknown population slope, $B$. The confidence interval is calculated as a function of the standard error of $b$ and a value chosen from the $t$ distribution to reflect the desired confidence level and the appropriate degrees of freedom. Likewise, a confidence interval for the difference in population slopes, $B_1 - B_2$, is a function of the combined error in the slope estimates, $b_1$ and $b_2$, and a $t$ value representing the combined degrees of freedom at a preselected level of significance (see Neter and Wasserman, 1974: 166). Just as the confidence interval for a single slope estimate can be used to implicitly test the null hypothesis $B=0$, so can the confidence interval for the difference in slope estimates be used to implicitly test the null hypothesis $B_1 - B_2 = 0$. In other words, if zero lies outside the range of values defined by confidence limits calculated as some predetermined level of significance and if all appropriate assumptions are met, then one would not normally infer that the population slopes are the same.
They key to this procedure, of course, is the qualification that "all appropriate assumptions are met." For present purposes the most important of these assumptions is that of random sampling. Note that a confidence interval represents a span on both sides of a point estimate defined as a certain multiple of standard errors corresponding to the desired significance level. Standard error refers to the standard deviation of a sampling distribution for the point estimate and can be thought of as the amount of error that would occur over the long run if the estimation process were repeated for a very large number of random samples. Thus, whenever the standard error is used in a statistical procedure there is a built-in assumption of random sampling.

From what has been said in the first part of this chapter it is clear that our data do not represent a random sample and, therefore, that it is inappropriate to construct a confidence interval for the purpose of formally testing the difference in regression slopes. An alternative procedure not requiring the randomness assumption would be to impose some arbitrary minimum value as a lower limit on what can be designated a conceptually significant difference in slopes. By this method, any comparison involving regression coefficients that differed from one another by at least as much as the criterion value would be evaluated as supportive of one of the two system constructs, whereas a comparison of coefficients separated by a margin less than the criterion would be deemed ambiguous and placed in the mixed category.
The chief weakness of this procedure is the arbitrariness in the choice of criterion value. Notice that whatever value is ultimately selected, it represents a trade-off between the desire for an overall pattern of results that clearly discriminate between the two system constructs—an outcome that is more easily obtained the lower the difference criterion—and the desire to avoid reaching an apparent conclusion that is in fact false—a goal that favors a higher criterion. There are other problems with this procedure, as well. For example, once a criterion value is chosen there is no straightforward method for its direct translation between standardized and unstandardized coefficients. Moreover, by definition, a fixed criterion cannot take into account variations in one's intuitive confidence that might occur across analyses involving different proportions of explained variance or different N sizes.

Given the uncertainties surrounding the arbitrary criterion method and the inappropriateness of the formal statistical test, it was decided that for the limited purposes of this inquiry we would utilize whatever relative differences emerged from the comparisons without any attempt to reinforce our confidence in these differences. This procedure involves a fairly high risk of error when the observed differences are small, nevertheless, it does have two distinct advantages: First, although there is some bias to this method of comparing slopes, at least we know that it is the direction of what statisticians refer to as Type II error. Second, this method will be sensitive to even the very small differences that would be expected due to the high degree of congruence among the structural indicators.
The final matter to be taken up in this part of the chapter concerns the translation of the structural measures into a form that is representative of the effects indicated by the stratification hypotheses. Recall that the first four hypotheses deal with monadic behavior patterns and the effects of absolute structural position. Stated differently, these hypotheses postulate that certain characteristics of a government's foreign policy behavior will depend upon where that government is located in the international structural hierarchy. Since the structural measures record each country's rank position they are already in an appropriate form for investigating monadic level effects.

They dyadic hypotheses introduce a complicating factor into our analysis by stipulating that the acting government's position must be considered in relation to the recipient's position. The first three dyadic hypotheses concern the effects of structural proximity, that is, the effects of the distance separating the positions occupied by the actor and recipient. Our measure of structural proximity will be formally defined as the absolute difference between the actor's rank position and the recipient's rank position. Symbolically, this measure of dyadic proximity is

\[ P_{ij} = |R_i - R_j| \]

where \( R_i \) denotes the rank score for acting government \( i \) and \( R_j \) stands for the rank score of recipient government \( j \).

The three remaining hypotheses go beyond simple distance to deal with the more sophisticated notion of structural directionality as an influence on dyadic patterns of behavior. Perhaps the most
obvious way to investigate the effects of directionality is by means of a signed measure of proximity, that is, a measure of the positive or negative difference between dyadic rank positions. Although each hypothesis makes reference to two behavioral properties, the analysis would require that only one of these be regressed on the signed measure of structural distance. This is because the two behavioral measures in each hypothesis are compliments of one another and, hence, are linearly dependent. For illustrative purposes let us suppose that Figure 9 represents the least squares regression line resulting from our analysis of hypothesis 10. Remember that hypothesis 10 asserts that multilateral behavior will tend to occur in positively distant dyads whereas unilateral behavior will be more prevalent in negatively distant dyads. Because unilateral and multilateral behavior are mutually exclusive and exhaustive categories their percentage scores must sum to 100. As a consequence, the imaginary results depicted in Figure 9 are consonant with both parts of the hypothesized relationship because the high degree of multilateral behavior on the right-hand side of the curve implies a correspondingly low degree of unilateral behavior and vice versa.23

The analysis procedure illustrated in Figure 9 has two important disadvantages. First, the relationships predicted by at least two of the last three hypotheses are probably not linear in form when analyzed using the signed measure of structural distance. This nonlinearity is due to the interaction between the effects of structural proximity and the signed measure indicating directionality. For instance, hypothesis 7 asserts that close proximity
FIGURE 9

Hypothetical Relationship Between Multilateral Behavior and Signed Structural Distance (Hypothesis 10)
will be accompanied by a high level of collaborative activity, and since collaboration is actually a subset of the larger class of multilateral behaviors (see note 12) we would expect the effects of proximity to be evident in our analysis of hypothesis 10. The same kind of interactive relationship is found between hypothesis 5, which states that proximate dyads will tend to act on communal problems, and hypothesis 8, which associates positive distance with internal or communal problems and negative distance with external problems.

Second, and more important, use of the signed measure for analysis of the last three hypotheses precludes independent assessment of what are really two, conceptually distinct hypothesized relationships. Although we have postulated that positive and negative structural distance will have contrasting effects on dyadic behavior, it is not necessary to assume that these relationships are symmetrical in form or magnitude. In fact, it is to our advantage to examine these relationships separately since by doing so we will get a clearer picture of the stratification process as well as a broader basis for evaluating the models of systemic organization. The obvious way to implement this strategy is to perform separate regression analyses on positively distant and negatively distant subsets of dyads. Unfortunately, this method involves a serious reduction in the number of problem specific dyads available for analysis. Moreover, it is only partially effective in dealing
with the nonlinearity problem outlined above. We can avoid these difficulties and still distinguish the effects of positive and negative distance through an alternative procedure called effect coding.

Effect coding is normally used as a method for conveying interval level properties onto nominal level measurements. It is similar to dummy variable coding in that it yields a set of variables corresponding to the categories in a nominal scale; the difference is that in dummy coding one group is assigned 0 in all the variables whereas in effect coding one group is assigned -1 in all the variables. When a dependent variable is regressed on effect coded independent variables the resulting coefficients are directly interpretable as the effects associated with the original nominal categories. The special advantages of effect coding for analysis of the last three stratification hypotheses will be made clear in what follows.25

Earlier we alluded to the potential utility of collapsing data values into a smaller number of groups. As a prelude to effect coding we will make use of this procedure by collapsing the signed measure of structural distance into three groups depending upon whether a dyadic recipient is positively distant, negatively distant, or proximate in relation to the acting government. One of the reasons given earlier for rejecting such a regrouping procedure was because it involved an intolerable loss of information. There is still some loss of information but because it is the signed
distance measure that is collapsed the loss at this stage of our analysis is far less severe than if the regrouping were performed on the original measure of rank position. Referring to the second of our earlier objections, we should further note that there are still no clear divisions or cutting points for this regrouping of data values. We are not without some guidance, however. For statistical reasons to be explained shortly, it is better if the three groups of dyadic cases are approximately equal in size. Since our purpose is merely to isolate the influence of structural proximity from the contrasting effects of positive and negative distance, this criterion seemed amply sufficient. Accordingly, a dyad is considered structurally proximate if the recipient is positioned within ten ranks above or below the actor; if the recipient is more than ten ranks above the actor the dyad is designated positively distant and if the recipient is more than ten ranks below the actor it is deemed negatively distant.

The application of effect coding to the three categories of signed distance is illustrated in Table 16. As in dummy variable coding, effect coding requires the generation of k variables (where k equals the number of groups minus one). Since the signed distance measure has been collapsed into three groups, the effect coding process results in two effect variables. Furthermore, since our primary interest is in positive and negative distance we have retained these categories as the two effect variables. The columns
TABLE 16

Effect Coding for Directional Hypotheses

<table>
<thead>
<tr>
<th>Definition</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Distance</td>
<td>Distance</td>
</tr>
<tr>
<td>Upward behavior</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(Ri-Rj &gt; 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximate behavior</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>(/Ri-Rj/ ≤ 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Downward behavior</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(Rk-Rj &lt; -10)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Ri = rank score for acting government i; Rj = rank score for recipient government j; Ranks are scored beginning with 1 so that a larger number indicates a lower position in the hierarchy.
of Table 16 list the coding formulas used to construct these variables. Note that proximate dyads are assigned -1 on both variables, whereas positively and negatively distant dyads are assigned either 1 or 0.

If one of the behavioral measures is regressed on both effect coded variables the resulting multiple regression equation has some rather unique properties. In the simplist case when the groups are the same size the intercept, or $a$, will equal the grand mean of the dependent variable and each $b$ will represent the deviation of its group's mean from this grand mean. For example, if we regress multilateral behavior on the effect variables shown in Table 16, the $b$ coefficient associated with positive distance would reflect the average proportion of multilateral behavior between positively distant dyads by indicating the direction and magnitude of this group's deviation from the overall proportion of multilateral behavior and, likewise, the $b$ coefficient for negative distance would reflect the average level of multilateral behavior between negatively distant dyads. Effect coding is so named because these deviations represent the effects of the original categories.

When the groups are not of identical size, as is the case with all of our analyses, the intercept will only approximate the grand mean of the dependent variable. In this situation, the intercept will equal the unweighted mean of the group means. To obtain the grand mean, each group mean has to be weighted by the number of
dyadic cases in the group. (Obviously, then, when the groups are the same size these weights will be identical and the intercept will equal the grand mean.) The b coefficients will have the same properties discussed above, except that the deviations will be from the unweighted mean of the group means rather than from the grand mean.

Although the effect coding procedure does involve some loss of information and a somewhat arbitrary collapsing of data values, it nevertheless has some noteworthy advantages for analysis of the last three stratification hypotheses. In the first place, it utilizes all of the dyadic cases for which there are data in the three systems. Second, it neutralizes the confounding influence of structural proximity by isolating proximate dyads from those which are positively and negatively distant. And finally, it differentiates between the contrasting effects of positive and negative distance to facilitate independent evaluation of the two proposition comprising each hypothesis. This last point is especially important because it has the effect of augmenting the number of comparisons on which our overall conclusions will be based.

This completes the development of our research design. The way is now prepared for presentation of the regression results to be used in comparing the single system and multiple systems models of systemic organization.
1There are exceptions, of course. For one attempt to distinguish behavioral from social science, see Senn (1971).

2Social scientists are not the only ones to make use of such trace phenomena. High energy physicists, for example, follow an analogous procedure when they examine trails left by dismembered subatomic particles in an acceleration chamber.

3The best overall description of these data is Hermann et. al. (1973). More detailed information can be obtained from the various coding manuals used in the project's datamaking phase (see Brady, 1973; Hermann, 1970; Hermann, 1971; Hermann and Swanson, 1972; and Salmore and Brady, 1972). In addition, a concise and very useful overview is found in Callahan (1979).

4Another dimension on which events are not minimally aggregated has to do with the problem or problems addressed by the actor. This aspect of an event is considered more closely in a later section.

5A thorough review of these studies is well beyond the scope of this chapter. For a comprehensive treatment of these issues and an annotated bibliography of source validity studies see the excellent volume edited by Don Munton (1978). Other important references include Burgess and Lawton (1972), Azar (1970), and various contributions to Rosenau (1974), and Azar and Ben-Dak (1975).

6To illustrate this procedure, let us suppose that nation A ranks first in the thirty-eight nation subset but ranks only sixth when all 125 nations are taken together. In this case, nation A would be assigned a rank value of six in the computation of both sets of statistics.

7Except for the disposition of coparticipating recipients, our treatment of recipient roles follows that of East and Hermann (1979). We elected to retain the participant role in scope of action because it is indicative of collaborative behavior and, therefore, is an important element in the stratification theory proposed in Chapter II.

8The variables and operational procedure are described by Hermann (1979). The average reliability of all eight variables is .79; individual reliability scores range from .75 to .84.
We should underscore Hermann and Coate's (1979) caveat that a coder or analyst can never be entirely certain of the actual problem perceived by an acting government's policymakers. The initial identification of the problem, like all other aspects of an event, must be inferred from publicly available information.

Hermann and Coate (1979) present a similar clustering scheme to facilitate discussion of this problem area dimension. In fact, our clusters reproduce theirs with but one exception: the disposition of the international organization category. They interpreted IOs as external entities whereas we included them in the communal grouping. At first glance our approach may appear incongruous with the fact that IOs are clearly outside the political jurisdiction of any one government, however, we assumed that a basic value attributed to an IO could also be considered a basic value of its member governments. In light of this assumption, the communal designation seems justified provided that (1) the IO in question is an intergovernmental organization and (2) the acting government holds membership in this IO. After careful inspection of numerous coded events, we became convinced that theses two conditions do indeed characterize the vast majority of cases in which a basic value is attributed to an IO. This is not to say that the Hermann-Coate interpretation is incorrect but rather that it is not the only valid interpretation.

The combination of five basic values and three sources of deprivation results in fifteen different permutations, however, one of these—nature or societal conditions combined with the security/military-physical safety basic value—is not logically possible by definition (Hermann and Coate, 1979).

This implies that multilateral behavior, expressed as a proportion of the dyadic aggregate, will always be at least as great as the proportion of collaborative behavior.

Ostgaard (1965) and Galtung and Ruge (1965) were among the first to argue that the news selection process is routinely distorted by these and other sources of bias. Empirical studies of how news selection bias affects the reportage of foreign policy events have yielded ambiguous results because the real "universe of events" can never be positively known. For example, compare the findings of Salmore and Butler (1978) and Smith (1969) with those of Peterson (1978) and Hoggard (1974).

As indicated in the cited passage, the purpose of this study was limited to an examination of some "common assumptions" about reporting bias which were derived from the work of Ostguard (1965) and Galtung and Ruge (1965). Obviously, it should not be taken as the definitive statement on the source validity of event data in general or of our data in particular. In addition, we should note that the study has not been without its critics; see the exchange between Salmore and Munton in Munton (1978).
Since our behavioral measures are based on raw frequencies and percentages of events we consider their interval level properties to be self-evident.

We should point out that not all researchers share our sanguine position regarding the treatment of ordinal data as interval. For opposing views see Mayer (1970) and Wilson (1971).

We use the term "residual" in its conventional sense as the difference between an observed value and a predicted value. Normally, predicted values are obtained from a curve fitting procedure such as least squares regression. When, as in this case, predicted values are obtained from some other method, residuals will not have their usual mathematical properties (see Neter and Wasserman, 1974: 98).

Of course, this condition can never be met with any degree of certainty. Nevertheless, the analyst relying on causal techniques must be prepared to simplify reality by disregarding any factors not explicitly brought into his or her model. The plausibility of this kind of assumption is always a matter of degree and must be weighed against the benefits of conducting the analysis. As Asher (1976: 12) writes, "If we are not willing at some point to proceed on an 'as if' basis—as if confounding variables presented no problem—then we will be paralyzed in our efforts at data analysis."

We considered two other procedures for dealing with this problem. One would have limited supportive decisions to those cases in which both trade measures and the military measure were in agreement vis-à-vis the overall measure. Under this rule the results would be deemed ambiguous whenever the two trade measures lead to different conclusions when compared to overall structure. We regarded this rule as much too harsh in light of the empirical differences between the rankings on the two trade indices. A less obvious alternative would have been to incorporate both trade measures into a multiple regression analysis on the grounds that they represent complementary aspects of trade structure. This procedure was rejected because the presence of two moderately correlated independent variables would virtually guarantee the trade system a more impressive relationship than the overall system.

We can illustrate this procedure rather easily by listing the possible permutations of results and then classifying them according to their support for the system constructs. For notational convenience we will use the greater than (>) and less than (<) symbols to express the relationship between regression results using overall structure (O), military structure (M), and the two measures of trade structure (T1 and T2).
Supportive of  | Mixed      | Supportive of
Single System | Support    | Multiple Systems

| Supportive of | Mixed       | Supportive of |
|Single System  | Support     | Multiple Systems |
|   O>M         | O>M         | O<M          |
|   O>T         | O<T         | O>T          |
| P = 3/8 or .375 | P=2/8 or .25 | P = 3/8 or .375 |

Each column represents a unique outcome and all possible outcomes are listed. Since we are still assuming that our measures are random variables, the probability of any one outcome will be equal to the probability of any other. Thus, the marginal probabilities presented in Figure 8 result from the uneven grouping of these outcomes. It may be useful to contrast this procedure with the simpler case involving only one measure of trade structure.

Supportive of  | Mixed       | Supportive of
Single System  | Support     | Multiple Systems

| Supportive of | Mixed       | Supportive of |
|Single System  | Support     | Multiple Systems |
|   O>M         | O>M         | O<M          |
|   O<T         | O<T         | O>T          |
| P = 1/4 or .25 | P= 2/4 or .50 | P = 1/4 or .25 |

The implications are clear and significant; by including two trade measures and by choosing a lenient decision rule for evaluating them we have increased the probability that our results will be supportive of one of the two system constructs. We should point out that had we chosen the harsher rule described in note 19 the probability of supportive results would have decreased by the same magnitude.

We should point out that a simpler design using only one trade measure would constitute an even more severe test. The summary probabilities of such a design are listed below.

Strong Support for Single System = .063
Weak Support for Single System = .250
Ambiguous = .375
Weak Support for Multiple System = .250
Strong Support for Multiple Systems = .063

Note that an ambiguous outcome is almost six times more likely than either of the strongly supportive outcomes.

One common technique for dealing with nonlinear relationships is to fit an equation containing higher order polynomials. Since this method does not easily lend itself to the comparative purposes of this exercise, particularly if comparisons are to be made between polynomials of different degree, it seems prudent to exhaust other possible remedies before turning to the polynomial technique. Another method for accomplishing this task employs two or more linear segments to approximate a curvilinear form. This technique was rejected because we had no a priori criterion for imposing the splits between segments and because of the problems that would be involved in comparing magnitudes of two or more segmented relationships.
If only to point out the obvious, we might also mention that the regression coefficients for the measures of multilateral and unilateral behavior will have identical values but reversed signs. The regression equation for the hypothetical example in Figure 9 is $Y = 50 + .5X$; had we used unilateral behavior as the dependent variable the regression equation would have been $Y = 50 - .5X$.

At a threshold level of four events our analyses of the military and trade systems will involve only seventy and fifty-three dyadic cases, respectively. Even if the subsets were of about equal size—which they are not—the number of cases included in the analyses would be quite small.

For a thorough treatment of the mathematics underlying regression using effect coded variables, see Kerlinger and Pedhazur (1973).

Ten was chosen after an examination of the empirical distributions associated with cutoff points ranging from five to twelve ranks on each of the structural measures.
CHAPTER V
RESULTS

This chapter presents results of the regression analyses for the ten stratification hypotheses developed in Chapter II. In order to perform the comparisons described in the previous chapter it will be necessary to report six different regression results for each hypothesis. Recall that what we have termed a type I comparison involves an examination of four empirical relationships: overall system structure and overall behavior; military system structure and military behavior; and the two trade structure measures and trade behavior. Type II comparisons require computation of two additional regressions between overall structure and trade behavior and between overall structure and military behavior.

In those cases where the empirical realtionships appear to be nonlinear in form we have attempted to linearize the relationship through appropriate remedial actions. To provide the reader with some indication of the kind of nonlinearities found in these data, we will also present scatter plots of several relationships. Furthermore, if the outcome of either comparison changes as a result of an upward shift in the event threshold level used to operationalize the behavioral patterns, we will trace these changes through the full set of regression results found at different threshold levels.
The following discussion is presented in four sections. We begin by considering the first four hypotheses concerning the impact of a government's position in the structural hierarchy on its patterns of monadic behavior. The second section deals with hypotheses five through seven and the effects of structural proximity on dyadic behavior. The third examines the countervailing effects of structural direction expressed in the final three hypotheses. The chapter concludes with a brief summary of the findings.

Structural Position and Monadic Behavior

To recapitulate, the first hypothesis maintains that governments well endowed with problem-solving capabilities—that is, those positioned toward the top of the hierarchy—will tend to engage in high levels of foreign activity. Since structural position is indicated by a country's rank score (where the highest rank is scored 1) the hypothesis predicts an inverse relationship between structural position and amount of foreign activity. The results obtained from a linear analysis of this hypothesis are displayed in Table 17. Several features of this table deserve comment before we can evaluate the performance of the two systemic constructs.

Although our primary interest focuses on the relative magnitudes of the regression coefficients (b), for completeness we have also reported the values of the intercepts (a), the coefficients of determination (r^2), and the N-sizes. For each analysis, the N-size, which reflects the number of cases included in the analysis, is
### Table 17

**Hypothesis 1:**

**Structural Position and Amount of Foreign Activity**

**(Linear Analysis)**

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Intercept $a$</th>
<th>Slope $b$</th>
<th>Coefficient of Determination $r^2$</th>
<th>$N^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall behavior</td>
<td>Capacity to act</td>
<td>423.11</td>
<td>-3.25</td>
<td>.25</td>
<td>37$^b$</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Military manpower</td>
<td>25.38</td>
<td>-.28</td>
<td>.34</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>18.08</td>
<td>-.12</td>
<td>.05</td>
<td>28</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Commodity share</td>
<td>55.97</td>
<td>-.62</td>
<td>.28</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Geographical share</td>
<td>55.47</td>
<td>-.58</td>
<td>.38</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>53.65</td>
<td>-.58</td>
<td>.38</td>
<td>38</td>
</tr>
</tbody>
</table>

$^a$N-sizes are determined by an event threshold level of 3 or more events.

$^b$The United States was omitted from this analysis since it is an outlier by 4.8 standard deviations on the behavior measure.
normally determined by two factors: (1) the event threshold level used to define the patterns of behavior and (2) the presence of missing data on the structural measures. Missing data plays a minor role in analysis of the monadic hypotheses since only one structural measure, the commodity-weighted share of world exports, was unavailable for any of our sample of governmental actors. The affected governments are East Germany, Guinea, and China (see note 33, Chapter III).

One other matter concerning the N-size of our analyses should be mentioned here. The first row of Table 17 indicates that the analysis of overall foreign activity was performed on only thirty-seven cases despite the fact that capacity to act scores were available for all thirty-eight governments. The United States was omitted from this particular regression analysis because it is an extreme outlier on the measure of overall behavior. A total of 1916 events were identified for the United States, a level that is 4.8 standard deviations above the mean and almost twice the magnitude of the next highest country, the Soviet Union. Two reasons justify removal of this case from the analysis. The first is entirely statistical and has to do with outliers' well known effects on placement of the least squares regression line and the resulting degree of fit. The second reasons concerns the likely presence of considerable measurement error on this case due to recording bias of the data source.
Let us now examine the coefficients in Table 17 that are relevant to type I comparisons. The first thing to notice about this table is that all coefficients indicate an inverse relationship as predicted by the hypothesis. Amount of foreign activity, it will be recalled, is one variable that is measured on incomparable scales and, hence, requires use of the standardized coefficients for type I comparisons. This difference in scale is clearly visible in Table 17. Note that the intercept values—that is, the points at which the regression lines intersect the dependent variable axes—range from about 18 to 55 for problem-specific behaviors whereas the overall behavior intercept is over 400. The same scale differences is evident in the adjacent column of unstandardized slopes. Although the standardized coefficients are not reported in the table, they are easily derived from the coefficients of determination. In fact, since squaring preserves the monotonic ordering of coefficient magnitudes we can perform type I comparisons directly on the $r^2$ values without taking square roots.

Type I comparison of results displayed in Table 17 indicates support for the multiple systems model. Two steps are involved in arriving at this conclusion. The first compares $r^2$ values found for the overall system measures (top row) and the military system measures (second row). Since .34, the coefficient for military behavior/military structure, quite clearly exceeds .25, the coefficient for overall behavior/overall structure, we can regard this result as supportive of the multiple systems model. The second step compares $r^2$
values for the overall system measures and the trade system measures (fourth and fifth rows). Once again, the comparison proves favorable to the multiple systems construct since both trade behavior/trade structure coefficients surpass the overall coefficient. Furthermore, since both comparison steps lead to the same outcome, we are able to conclude that type I evidence on the first hypothesis is supportive of multiple systems.

The type II comparison of results in Table 17 can be performed on the unstandardized slopes since no scale incomparabilities are involved. Here again the comparison is carried out in two steps. The first compares b values found in the two analyses of military behavior (second and third rows). The slope for military behavior/military structure is of greater magnitude than the slope for military behavior/overall structure indicating further support for the multiple systems model. The second step then compares b values from the trade behavior analyses (rows four, five, and six). Even though the slopes for geographically-weighted share and capacity to act are the same, the slightly higher value of the commodity-weighted share slope allows us to conclude that this comparison step also supports the multiple systems construct. (We should reiterate that according to the decision rule for dealing with the two trade structure measures, results for only one measure are sufficient to lend support to one of the system models.) Since both type I and type II comparisons support the multiple systems model, the empirical relationships on the first hypothesis seem to be strongly supportive of multiple systems.
Thus far we have presented a step-by-step illustration of the procedures necessary for type I and type II comparisons of empirical relationships relevant to the first stratification hypothesis. Note, however, that all of these relationships have been presumed to be linear in form. Are there appreciable nonlinearities in these data? And if so, do they have any effect on our interpretation of the evidence contained in Table 17? To answer the first question we inspected scatter plots of the six relationships in Table 17. Two of the plots are reproduced in Figures 10 and 11. All of the plots examined revealed on inverse curvilinear form quite similar to the curved relationships displayed in the figures. Clearly there is sufficient departure from linearity to warrant reinvestigation of the first hypothesis. Before discussion the details of this reinvestigation it is appropriate that we consider the theoretical implications of this relational form.

According to the explanatory logic outlined in Chapter II, high structural position implies a relative abundance of problem-solving capabilities necessary for setting goals, monitoring problems, and taking actions in the international arena. To investigate this proposition we assumed a linear relationship between structural position and amount of foreign activity. Put differently, we assumed that a given change in structural position would be associated with a constant change in amount of activity. The evidence presented in Figures 10 and 11 suggests that a change in position at the top of the hierarchy will tend to have a more dramatic impact on foreign
FIGURE 10

Scatter Plot of Capacity to Act and Amount of Overall Foreign Activity
FIGURE 11

Scatter Plot of Military Manpower and Amount of Foreign Military Activity
activity than an identical change occurring in the middle and lower ranks. It may be, for example, that governments in the topmost positions of the hierarchy attract a disproportionate amount of attention from other actors in the system and as a result tend to engage in even higher levels of activity than might otherwise be expected. Another possible explanation is that as available problem-solving capabilities increase their effectiveness as a constraint on international involvement tends to diminish. The point to be emphasized here is that identification of a nonlinear relational form is not inconsistent with our previously established explanatory rationale.4

To linearize the relationship between rank position and amount of foreign activity we applied a logarithmic transformation to each of the structural measures. This transformation expands the distances between the highest rank scores so that a change in position at the top of the hierarchy will appear to have about the same impact on foreign activity as an identical change occurring farther down in the structure. The results of regression analyses on these log transformed data are reported in Table 18. It should be understood that the analysis procedures underlying the results displayed in Tables 17 and 18 are identical in every respect except that the analyses reported in Table 18 used the common logarithm of the rank score as the independent variable. The column of r² values indicates that the log transformation improved the degree of fit in five of the six relationships analyzed; only trade behavior regressed on commodity-weighted share failed to respond to this transformation.
### TABLE 18

**Hypothesis 1:**
**Structural Position and Amount of Foreign Activity**
**(Nonlinear Analysis)**

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Intercept ($a$)</th>
<th>Slope ($b$)</th>
<th>Determination ($r^2$)</th>
<th>N $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall behavior</td>
<td>Capacity to act</td>
<td>760.36</td>
<td>-324.41</td>
<td>.48</td>
<td>37$^c$</td>
</tr>
<tr>
<td></td>
<td>Military manpower</td>
<td>45.16</td>
<td>-22.27</td>
<td>.60</td>
<td>28</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Capacity to act</td>
<td>26.99</td>
<td>-9.82</td>
<td>.10</td>
<td>28</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Commodity share</td>
<td>72.53</td>
<td>-28.55</td>
<td>.21</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Geographical share</td>
<td>89.92</td>
<td>-41.47</td>
<td>.53</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>94.89</td>
<td>-45.52</td>
<td>.56</td>
<td>38</td>
</tr>
</tbody>
</table>

---

$^a$ All relationships were linearized by taking the common logarithm of the independent variable. The coefficients reported here were derived from these logged data.

$^b$ N-sizes are determined by an event threshold level of 3 or more events.

$^c$ The United States was omitted from this analysis since it is an outlier by 4.8 standard deviations on the behavior measure.
Application of a logarithmic transformation will alter the meaning of regression slopes and intercepts so these figures should be interpreted with care. Since the logarithm of one equals zero, the intercepts listed in Table 18 can be read directly as the predicted activity score for the highest ranking actor. Interpretation of the slopes is somewhat more complicated since they are stated in logarithmic units of the independent variable. For example, the slope for overall behavior/overall structure in Table 18 indicates that a one unit increase in the logged structural variable will involve an expected increase of about 325 total events.

Since the log transformation does nothing to compensate for the scale differences in the measures of foreign activity, type I comparisons of results in Table 18 must be carried out on $r^2$ values. Although the outcome is mixed for the trade structure measures, the type I comparisons again support the multiple systems model. The type II comparisons of unstandardized slopes reveal a more significant change brought about by the log transformation: neither trade structure measure exceeds overall capacity to act. Since the military behavior analyses continue to be supportive of multiple systems, the type II evidence on the first hypothesis indicates mixed support. When both comparisons are taken together the overall result for this hypothesis is weak support for the multiple systems model.

The second monadic hypothesis asserts that rank position will be inversely related to scope of action. Because scope of action is
measured in terms of raw frequencies—that is, the number of
different governmental recipients of an actor's monadic behavior—
type I comparisons on this hypothesis must be based on the standardized coefficients. Furthermore, since scope of action represents a
second aspect of international involvement, it is likely to produce
the same kind of nonlinear relationships as those encountered in
the analyses of foreign activity. In other words, we would expect
to find a constant change in structural position associated with a
decreasing change in predicted scope of action.

Tables 19 and 20 present results of linear and nonlinear analyses,
respectively. The column of slopes indicates that all six relationships are in the predicted direction. The regression analyses listed
in Table 20 were performed on the common logarithm of the rank position scores. In five of the six relationships the degree of fit
($r^2$) was noticeably improved by the logarithmic transformation; once again, trade behavior regressed on commodity-weighted share proved intransigent under this transformation. Scatter plots displayed in Figures 12 and 13 provide additional evidence that the relational form expressed in the second hypothesis is nonlinear.

Type I comparisons for this hypothesis will be carried out on
the $r^2$ values listed in Table 20. The comparisons of military behavior/military structure with overall behavior/overall structure clearly is supportive of multiple systems. Once again, the logged results provide only mixed support when overall behavior/overall structure is compared to the two trade system analyses. Based on
TABLE 19

Hypothesis 2:
Structural Position and Scope of Action
(Linear Analysis)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Intercept $a$</th>
<th>Slope $b$</th>
<th>Coefficient of Determination $r^2$</th>
<th>$N^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall behavior</td>
<td>Capacity to act</td>
<td>64.18</td>
<td>-.21</td>
<td>.17</td>
<td>38</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Military manpower</td>
<td>13.54</td>
<td>-.13</td>
<td>.42</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>11.34</td>
<td>-.09</td>
<td>.15</td>
<td>28</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Commodity share</td>
<td>21.00</td>
<td>-.19</td>
<td>.35</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Geographical share</td>
<td>20.16</td>
<td>-.16</td>
<td>.38</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>19.22</td>
<td>-.15</td>
<td>.33</td>
<td>38</td>
</tr>
</tbody>
</table>

$^aN$-sizes are determined by an event threshold of 3 or more events.
TABLE 20
Hypothesis 2:
Structural Position and Scope of Action
(Nonlinear Analysis)\textsuperscript{a}

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Intercept ( a )</th>
<th>Slope ( b )</th>
<th>Coefficient of Determination ( r^2 )</th>
<th>( N ) \textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall behavior</td>
<td>Capacity to act</td>
<td>83.60</td>
<td>-19.67</td>
<td>.36</td>
<td>38</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Military manpower</td>
<td>21.92</td>
<td>- 9.72</td>
<td>.67</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>17.07</td>
<td>-6.56</td>
<td>.26</td>
<td>28</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Commodity share</td>
<td>26.09</td>
<td>-8.88</td>
<td>.25</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Geographical share</td>
<td>29.18</td>
<td>-11.21</td>
<td>.49</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>30.27</td>
<td>-12.13</td>
<td>.50</td>
<td>38</td>
</tr>
</tbody>
</table>

\textsuperscript{a}All relationships were linearized by taking the common logarithm of the independent variable. The coefficients reported here were derived from these logged data.

\textsuperscript{b}N-sizes are determined by an event threshold level of 3 or more events.
FIGURE 12

Scatter Plot of Geographically-Weighted Share and Scope of (Trade) Action
FIGURE 13

Scatter Plot of Capacity to Act and Scope of (Trade) Action
the decision rule for interpreting split results on the measures of trade structure, we are able to conclude that type I evidence on the second hypothesis supports the multiple systems model. Although the type II comparison between unstandardized slopes for military behavior/military structure and military behavior/overall structure also supports the multiple systems construct, the slopes for both trade structure measures fall below the overall structure slope. Hence, type II evidence on this hypothesis provides mixed support for the systems models. The overall result from the scope of action analyses is weak support for the multiple systems model.

Before turning to the next hypothesis, we should offer a brief word of caution regarding the interpretations given to hypotheses one and two. Based on the evidence reported in Tables 18 and 20, we concluded that both hypotheses provide weak support for the multiple systems model. Moreover, both hypotheses were split on type II comparisons and both responded in remarkably similar ways to the same logarithmic transformation. These and other similarities between the results for hypotheses one and two suggest a very high correlation between the measures of foreign activity and scope of action. As it turns out, the product-moment coefficients between these two measures are .79 for overall behavior (.82 when the U.S. is omitted), .78 for military behavior, and .91 for trade behavior. The empirical correspondence between these variables coupled with the fact that both are recorded as raw frequencies suggests the possibility that they are not products of independent measurement procedures. This
would have serious implications for the analyses just completed be­
cause the two hypotheses could not stand as independent sources of
evidence about the models of systemic organization.

Two kinds of arguments can be offered in response. The first
simply takes note of the conceptual similarity between foreign activity
and scope of action—because both are aspects of international involve­
ment we should expect them to be highly correlated. Even so, the
coefficients reported above indicate that they are not measures of the
same concepts. Second, if the measurement procedures are examined
closely it becomes evident that they are not dependent on one another.
Put another way, we can find no procedurally induced correspondence
between the number of different recipients and the number of events
for monadic behavior patterns. In short, we are prepared to accept
hypotheses one and two as independent sources of evidence.

The third stratification hypothesis maintains that the higher a
government's rank position, the more initiatory its foreign policy
behavior. Regression results for this hypothesis are presented in
Table 21. Because of the scoring system used to record rank position,
the inverse relationships in the table correspond to the direction
predicted by the hypothesis. It should also be noted that, on the
whole, the coefficients of determination for hypothesis three indicate
a substantially poorer fit to the data than do the coefficients for
either of the previous hypotheses.

Since initiatory behavior is measured as a percentage of
events aggregated to form a behavioral pattern, type I comparisons
TABLE 21

Hypothesis 3:
Structural Position and Initiatory Behavior

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Intercept $a$</th>
<th>Slope $b$</th>
<th>$r^2$</th>
<th>$N^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall behavior</td>
<td>Capacity to act</td>
<td>54.35</td>
<td>-.11</td>
<td>.10</td>
<td>38</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Military manpower</td>
<td>66.14</td>
<td>-.30</td>
<td>.18</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>58.33</td>
<td>-.14</td>
<td>.03</td>
<td>28</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Commodity share</td>
<td>48.72</td>
<td>-.26</td>
<td>.22</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Geographical share</td>
<td>40.40</td>
<td>-.07</td>
<td>.02</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>41.41</td>
<td>-.10</td>
<td>.04</td>
<td>38</td>
</tr>
</tbody>
</table>

$^aN$-sizes are determined by an event threshold level of 3 or more events.
for this hypothesis will be carried out on unstandardized slopes. The slope magnitudes for both military behavior/military structure and trade behavior/commodity share are well above that for overall structure/overall behavior, thus indicating type I support for the multiple systems model. The results of type II comparisons are also supportive of multiple systems. Due to the poor performance of the geographically-weighted share, both types of comparisons involved mixed results on the trade behavior analyses. Nevertheless, since inspection of the scatter plots for these six relationships revealed no evidence of a consistent nonlinear pattern, we conclude that hypothesis three provides strong support for the multiple systems model.

The last of the monadic hypotheses asserts that the higher a government's position in the structural hierarchy, the more likely it is to engage in unilateral foreign behavior. The results displayed in Table 22 indicate that all six relationships are in the hypothesized direction. Here again, type I comparisons can be performed on unstandardized coefficients since unilateral behavior is measured on a percentage basis. Although the multiple systems model is supported when the military behavior/military structure slope is compared to the overall behavior/overall structure slope, this outcome is reversed for the two trade structure measures. Thus, type I evidence from this hypothesis is mixed. The type II comparison on the military behavior analyses is clearly favorable to the multiple
<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Intercept  ( a )</th>
<th>Slope  ( b )</th>
<th>Determination  ( r^2 )</th>
<th>( N^a )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall behavior</td>
<td>Capacity to act</td>
<td>36.01</td>
<td>-.23</td>
<td>.14</td>
<td>38</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Military manpower</td>
<td>53.86</td>
<td>-.55</td>
<td>.33</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>40.18</td>
<td>-.27</td>
<td>.06</td>
<td>28</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Commodity share</td>
<td>17.64</td>
<td>-.15</td>
<td>.14</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Geographical share</td>
<td>16.62</td>
<td>-.11</td>
<td>.10</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>16.60</td>
<td>-.12</td>
<td>.12</td>
<td>38</td>
</tr>
</tbody>
</table>

\( ^a \)N-sizes are determined by an event threshold level of 3 or more events.
systems model since the slope for military structure is about twice that for overall structure. Slopes for the trade behavior relationships are much closer together but based on the slightly larger value for the commodity share measure we will regard type II evidence as supportive of multiple systems. Admittedly, this is a rather tenuous conclusion in light of such a small difference in slopes; nevertheless, given the exploratory nature of this study it seems wiser to err toward supporting one of the system constructs than toward the ambiguous category. For this reason, hypothesis four is presumed to offer weak support for the multiple systems model.

To summarize our findings thus far, all four monadic hypotheses indicate some degree of support for the multiple systems model. Only one of these, hypothesis three, could be placed in the strong support category. Of the weakly supportive hypotheses, only the last entailed mixed results on type I comparisons. Examination of the scatter plots revealed that the linear form was a satisfactory approximation for hypotheses three and four but not for hypotheses one and two. Moreover, we found that our conclusions on all four hypotheses proved to be quite stable at higher event threshold levels.

Structural Proximity and Dyadic Behavior

This section marks a shift in the patterns of behavior presumed malleable to stratification influences. The three hypotheses examined here deal with relationships between dyadic behavior and the vertical proximity of structural positions occupied by the actor and its
recipient. Proximity, it will be recalled, is measured as the absolute difference between rank scores.

The first dyadic hypothesis, or the fifth overall, states that the closer the structural proximity of two governments, the more their mutual behavior will be motivated by communal problems. Table 23 displays regression results for this hypothesis when performed on dyadic cases meeting the initially specified ten event threshold level for overall behavior and the four event threshold for problem-specific behavior. N-sizes for three of the six analyses reported in the table were reduced by an absence of structural data for a dyad's actor, recipient, or both. For example, even though a total of 440 overall behavior dyads surpassed the ten event threshold level, eleven of these had to be dropped from the analysis because capacity to act scores were unavailable for their recipients. For similar reasons, five and four dyadic cases were dropped from the analyses of military behavior/capacity to act and trade behavior/commodity share, respectively.

The results presented in Table 23 are interesting in several respects. In the first place, results from the overall behavior/overall structure and military behavior/military structure analyses demonstrate that the choice of coefficient can have an effect on the outcome of the comparisons. Although the differences are slight in either case, the unstandardized slopes favor the multiple systems model whereas the standardized coefficients (indicated by $r^2$ values)
TABLE 23
Hypothesis 5:
Structural Proximity and Communal Problems
(4 Event Threshold)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Intercept $a$</th>
<th>Slope $b$</th>
<th>Coefficient of Determination $r^2$</th>
<th>Na</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall behavior</td>
<td>Capacity to act</td>
<td>43.15</td>
<td>-.33</td>
<td>.09</td>
<td>429</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Military manpower</td>
<td>29.48</td>
<td>-.35</td>
<td>.05</td>
<td>70</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Capacity to act</td>
<td>33.26</td>
<td>-.37</td>
<td>.09</td>
<td>65</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Commodity share</td>
<td>65.66</td>
<td>.35</td>
<td>.04</td>
<td>49</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Geographical share</td>
<td>68.59</td>
<td>.29</td>
<td>.04</td>
<td>53</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Capacity to act</td>
<td>77.62</td>
<td>-.35</td>
<td>.02</td>
<td>53</td>
</tr>
</tbody>
</table>

$^{a}$N-sizes are determined by event threshold levels of 10 or more events for overall behavior and 4 or more events for problem-specific behavior.
favor the single system construct. Since communal problems are measured as a percentage of aggregated events, we regard the unstandardized slopes as the more appropriate basis for comparison. A second notable feature of these results is that only four of the six relationships are in the predicted direction. Contrary to expectations, both trade behavior/trade structure analyses indicate that communal problems tend to increase as structural proximity decreases. Consequently, neither of these relationships can be interpreted as supportive of multiple systems. The type I comparisons result in mixed support and the type II comparisons support the single system model. Thus, the overall evidence on hypothesis five offers weak support to the single system model.

Before moving on to the next hypothesis we should mention one other aspect of the results contained in Table 23; they are sensitive to increases in the problem-specific event threshold. When the regression analyses are repeated on problem dyads comprised of five or more events, the results are as shown in Table 24. The only difference between Tables 23 and 24 having any bearing on the comparisons is a shift in the relative magnitudes of the slopes on military behavior. Note that the slope for military behavior/overall structure exceeds that for military behavior/military structure at the four event level, but just the reverse is true at the five event level. Although this change does not affect the overall conclusion of weak support for the single system model, it does transpose the outcomes of the two comparison types.
TABLE 24

Hypothesis 5:
Structural Proximity and Communal Problems
(5 Event Threshold)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Intercept (a)</th>
<th>Slope (b)</th>
<th>Coefficient of Determination ($r^2$)</th>
<th>N&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall behavior</td>
<td>Capacity to act</td>
<td>43.15</td>
<td>-.33</td>
<td>.09</td>
<td>429</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Military manpower</td>
<td>27.09</td>
<td>-.31</td>
<td>.05</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>29.81</td>
<td>-.28</td>
<td>.08</td>
<td>43</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Commodity share</td>
<td>64.16</td>
<td>.59</td>
<td>.07</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Geographical share</td>
<td>68.88</td>
<td>.53</td>
<td>.08</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>79.40</td>
<td>-.49</td>
<td>.03</td>
<td>29</td>
</tr>
</tbody>
</table>

<sup>a</sup>N-sizes are determined by event threshold levels of 10 or more events for overall behavior and 5 or more events for problem-specific behavior.
Hypothesis six asserts that close structural proximity will be associated with high levels of dyadic foreign activity. Dyadic foreign activity, like its monadic counterpart, is measured as a frequency of events and thus entails a scale difference that necessitates use of the standardized coefficients for type I comparisons. Regression analyses for hypothesis six are reported in Table 25. One trade behavior dyad, France → United Kingdom, proved to be an extreme outlier at almost five standard deviations above the mean and was removed from the analysis. The removal of this case had no effect on the outcome of either comparison.

None of the relationships exhibited in Table 25 are very impressive; in fact, two of the capacity to act analyses fail to account for any variation whatsoever and the third is not much better. Probably the most striking feature of these results is the improved performance of the trade structure measures. Since the outcomes of both type I and type II comparisons are supportive of multiple systems, we can conclude that hypothesis six entails strong support for the multiple systems model.

The relative magnitudes of the relationships between structural proximity and dyadic activity proved to be quite stable when examined at more stringent event threshold levels. Furthermore, given the non-linearities found for the monadically patterned version of foreign activity, we took special care in inspecting scatter plots of these relationships. Are there theoretical reasons for believing that structural proximity might also have a nonlinear effect on foreign
TABLE 25

Hypothesis 6:
Structural Proximity and Amount of Foreign Activity

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Intercept $a$</th>
<th>Slope $b$</th>
<th>Coefficient of Determination $r^2$</th>
<th>$N^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall behavior</td>
<td>Capacity to act</td>
<td>34.23</td>
<td>.05</td>
<td>.00</td>
<td>429</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Military manpower</td>
<td>11.59</td>
<td>-.09</td>
<td>.06</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>8.65</td>
<td>.02</td>
<td>.00</td>
<td>65</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Commodity share</td>
<td>8.09</td>
<td>-.08</td>
<td>.14</td>
<td>48$^b$</td>
</tr>
<tr>
<td></td>
<td>Geographical share</td>
<td>7.51</td>
<td>-.07</td>
<td>.13</td>
<td>52$^b$</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>6.78</td>
<td>-.05</td>
<td>.02</td>
<td>52$^b$</td>
</tr>
</tbody>
</table>

$^a$N-sizes are determined by event threshold levels of 10 or more events for overall behavior and 4 or more events for problem-specific behavior.

$^b$The France-United Kingdom dyad was omitted from this analysis since it is an outlier by 4.9 standard deviations on the behavior measure.
activity? One possibility that comes to mind involves what Rummel (1977a) has termed the "principle of diminishing salience," that is, the inability to discriminate distances of large magnitudes while accurately perceiving or even exaggerating distances of small magnitudes. Figures 14 and 15 plot two of the relationships between structural proximity and dyadic activity. There is a suggestion of nonlinearity in these figures but it is much more subtle than in similar plots of hypotheses one and two. In both figures the data points tend to fall within a triangular pattern emanating from the origin in the bottom left corner. We would expect this kind of distribution to be fairly unresponsive to the logarithmic or other linearizing transformations and, indeed, our efforts along these lines failed to improve the degree of fit in any significant way.

The more interesting characteristic of the patterns in Figures 14 and 15 is that they seem to bear out some of our previous speculations about the sources of intergovernmental activity. The triangular form of the pattern suggests that structural proximity may not be a sufficient condition for dyadic interaction but, rather, that it may serve to enhance levels of activity initially precipitated by other factors such as, for example, geographical proximity or ideological similarity.

The seventh hypothesis maintains that the relative frequency of collaborative foreign activity will be highest between structurally proximate governments. The results presented in Table 26 testify that once again proximity on the trade structure measures seems to influence behavior in the opposite direction. To reiterate, the
FIGURE 14

Scatter Plot of Military Manpower Proximity
and Amount of Dyadic Military Activity
FIGURE 15

Scatter Plot of Commodity-Weighted Share Proximity and Amount of Dyadic Trade Activity
<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Intercept $a$</th>
<th>Slope $b$</th>
<th>Determination $r^2$</th>
<th>N$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall behavior</td>
<td>Capacity to act</td>
<td>29.20</td>
<td>-.32</td>
<td>.09</td>
<td>429</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Military manpower</td>
<td>11.78</td>
<td>-.02</td>
<td>.00</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>20.92</td>
<td>-.25</td>
<td>.06</td>
<td>65</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Commodity share</td>
<td>38.26</td>
<td>.62</td>
<td>.09</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Geographical share</td>
<td>41.80</td>
<td>.47</td>
<td>.07</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>51.21</td>
<td>-.08</td>
<td>.00</td>
<td>53</td>
</tr>
</tbody>
</table>

$^aN$-sizes are determined by event threshold levels of 10 or more events for overall behavior and 4 or more events for problem-specific behavior.
implication of these contrary findings is that neither trade behavior/trade structure relationship can be cited as evidence for the multiple systems model. Examination of the scatter plots between proximity and collaboration revealed no indication of nonlinearities in these relationships.

With the trade structure analyses eliminated from consideration and the b value for overall behavior/overall structure exceeding that for military behavior/military structure by a relatively wide margin, the type I comparisons are clearly supportive of the single system model. At first glance, it would appear that the type II comparisons lead to the same conclusion and, thus, to an overall designation of strong support for the single system model. A closer look at the results in Table 26 reveals the speciousness of this interpretation. Although the trade behavior/capacity to act relationship gives an impression of supporting the hypothesis, the \( r^2 \) value indicates a situation of virtual statistical independence. Under these circumstances, it seemed prudent to withhold support from either systems construct by relegating type II results to the mixed category. Since the outcomes of the comparisons are unaffected by changes in the event threshold level, our overall conclusion is that hypothesis seven provides weak support for the single system model.

The three dyadic hypotheses considered in this section leave us with some conflicting signals regarding the relative utility of the single system and multiple systems models. Even though hypothesis six was strongly supportive of multiple systems, our analyses of hypotheses
five and seven indicated weak support for the contrasting image of a single, comprehensive international system. Furthermore, on both of these hypotheses the trade structure measures performed counter to our expectations. Although the reason for these contrary findings is not entirely clear, the fact that in both instances the trade behavior/capacity to act analyses revealed weak or virtually nonexistent relationships suggests that it may be at least partially a function of the problem area's substantive content. Finally, we should mention these hypotheses' rather poor fit to the data -- across all of the analyses examined in this section, structural proximity managed to account for a maximum of only twelve percent of the variation in dyadic behavior.

Structural Direction and Dyadic Behavior

Although still focused on behavior in the dyadic unit, the final three hypotheses are less concerned with the proximity of structural positions than with their directional configuration. Direction, it will be recalled, is defined in terms of the acting government's position vis-a-vis its recipient -- a dyad is considered positive in direction if the actor ranks below its recipient and negative if the positions are reversed. Each of the hypotheses is composed of two analytically and empirically distinct assertions regarding the countervailing influences of positive and negative direction. To facilitate independent evaluation of these two assertions we have implemented an effect coding procedure that
accentuates the contrasting effects of positive and negative structural distance and simultaneously neutralizes the confounding effects of proximity within ten rank positions in either direction.

The use of effect coding requires some modification of the analysis techniques applied to the first seven stratification hypotheses. The most conspicuous change is the shift to multiple regression analysis in order to accommodate both effect coded directional variables. For obvious reasons, the coefficient of determination will no longer correspond to the square of the stand-"ized regression coefficient; however, this is of little consequence since all of the dependent variables examined in this section are percentagized measures and, hence, are comparable in scale. Furthermore, since effect coding acts as a kind of linearizing transformation its counteracts any disturbances caused by the presence of nonlinearities and thereby eliminates the need to inspect scatter plots of the relationships. Finally, it must be emphasized that these changes in analysis procedures, however significant, will in no way affect the previously established procedures for carrying out and evaluating the type I and type II comparisons of results.

Hypothesis eight deals with directional differences in the kinds of problems motivating dyadic foreign policy behavior. It states that actions directed toward positively distant recipients — that is, upward in the hierarchy — are likely to concern internal or communal problems and, contrariwise, actions toward negatively distant recipients are likely to concern external problems. This
hypothesis was investigated by regressing the combined percentage of internal and communal problems on the effect coded indicators of positive and negative structural distance. The results of these analyses are presented in Table 27.

Special notice should be given to the discrepancies in the column of intercept values. Effect coding produces an intercept that is equal to the unweighted mean of group means; the grand mean, on the other hand, equals the weighted mean of group means, where the weighting factor is the number of cases in each group. The grand means of the behavioral measures represented in Table 27 are 51.29 for overall behavior, 32.65 for military behavior, and 91.72 for trade behavior. These figures typify the close correspondence between intercept values and actual means. The more important point, however, is that both means and intercepts disclose some striking differences among the overall, military, and trade behavior measures, implying that variation in substantive content can be a powerful influence on problem sensitivity. For reasons explicated more fully in Chapter II, this finding must be viewed as somewhat disturbing because it raises the possibility that our results are tapping differences due to problem content rather than differences due to the level of problem specification.

The two columns of b coefficients in Table 27 separately denote the effects associated with positive distance and negative distance. For example, the analysis of overall behavior (top row) reveals that the rate of attention to internal and communal problems is about one
TABLE 27
Hypothesis 8:
Structural Direction and Internal/Communal Problems

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Intercept a</th>
<th>Positive Distance b</th>
<th>Negative Distance b</th>
<th>Coefficient of Determination R²</th>
<th>Na³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall behavior</td>
<td>Capacity to act</td>
<td>52.90</td>
<td>1.00</td>
<td>-11.57</td>
<td>.08</td>
<td>429</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Military manpower</td>
<td>36.06</td>
<td>7.33</td>
<td>-13.71</td>
<td>.06</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>38.15</td>
<td>5.57</td>
<td>-16.24</td>
<td>.08</td>
<td>65</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Commodity share</td>
<td>91.24</td>
<td>-.04</td>
<td>2.35</td>
<td>.02</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Geographical share</td>
<td>93.19</td>
<td>-.50</td>
<td>4.31</td>
<td>.06</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>90.49</td>
<td>-8.62</td>
<td>6.38</td>
<td>.12</td>
<td>53</td>
</tr>
</tbody>
</table>

³N-sizes are determined by event threshold levels of 10 or more events for overall behavior and 4 or more events for problem-specific behavior.
percentage point higher than average among positively distant dyads and about eleven points lower than average among negatively distant dyads. Each column of b coefficients will be evaluated independently according to the established procedures for performing type I and type II comparisons of results. To facilitate summary presentation of two sets of findings on this and the other compound hypotheses, the designations A and B will be used to differentiate between positive distance and negative distance, respectively. We begin our evaluation of hypothesis eight by examining the effects of positive distance.

Table 27 indicates that the effects of positive distance run counter to our expectations in all three trade behavior analyses. Since we cannot infer support for either systems model on the basis of these contradictory findings, evidence from type II comparisons falls into the mixed category regardless of the outcome on the military behavior analyses. Type I comparisons also provide mixed support, although for different reasons. When overall behavior/overall structure is compared to military behavior/military structure, the result is clearly favorable to the multiple systems model, however, when compared to the negative values of the two trade behavior/trade structure relationships, even a b value of 1.00 is sufficient to be deemed supportive of the single system construct. With mixed outcomes on both types of comparisons, the results for hypothesis eight-A can support neither system model and therefore must be regarded as ambiguous.
Turning now to the effects of negative distance, we find that once again all three trade behavior analyses contradict the hypothesized relationship. As before, this guarantees a mixed outcome on type II comparisons. Also as before, evidence from the type I comparisons is partially supportive of both system models. In fact, apart from the rather sizeable differences in the magnitudes of the coefficients, the negative distance results are practically a mirror image of the results for positive distance. Since neither system model is supported by these findings, hypotheses eight-B will be considered ambiguous.

The ninth hypothesis asserts that behavior occurring within positively distant dyads will tend to be more reactive than otherwise and, conversely, behavior within negatively distant dyads will tend to be more initiatory than otherwise. Results for this hypothesis are displayed in Table 28. The column of intercept values indicates that on the whole and in matters specifically related to international trade dyadic interactions are fairly evenly divided between reactive and initiatory types of behavior, whereas on questions dealing with the negotiation and settlement of military security disputes dyadic behavior appears to be about four times more initiatory than reactive. The differences here are neither so large as those found for problem sensitivity nor do they follow the same pattern, nevertheless, they can be taken as further evidence that problem content does have some impact on behavioral output. We will begin evaluation of this hypothesis with part A—the effects of positive distance.
### Hypothesis 9: Structural Direction and Reactive Behavior (4 Event Threshold)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Intercept a</th>
<th>Positive Distance b</th>
<th>Negative Distance b</th>
<th>Coefficient of Determination $R^2$</th>
<th>N&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall behavior</td>
<td>Capacity to act</td>
<td>48.23</td>
<td>8.78</td>
<td>-6.67</td>
<td>.06</td>
<td>429</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Military manpower</td>
<td>19.28</td>
<td>-2.44</td>
<td>- .10</td>
<td>.01</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>20.82</td>
<td>.08</td>
<td>-4.62</td>
<td>.04</td>
<td>65</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Commodity share</td>
<td>50.78</td>
<td>6.86</td>
<td>-2.57</td>
<td>.03</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Geographical share</td>
<td>50.91</td>
<td>- .32</td>
<td>-4.91</td>
<td>.02</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>46.00</td>
<td>6.58</td>
<td>-19.13</td>
<td>.15</td>
<td>53</td>
</tr>
</tbody>
</table>

<sup>a</sup>N-sizes are determined by event threshold levels of 10 or more events for overall behavior and 4 or more events for problem-specific behavior.
The results in Table 28 prove contrary to the hypothesized effects of positive distance in two instances. The more significant of these is the military behavior/military structure relationship because it is automatically eliminated from any further consideration as a basis of support for the multiple systems model. With this contrary finding in the military behavior analysis, we need only compare overall behavior/overall structure with trade behavior/trade structure to conclude that type I evidence supports the single system model. Type II comparison of the military behavior analyses also supports the single system construct but the comparison of trade behavior analyses favors the multiple systems model by a slight margin. Our overall conclusion for hypothesis nine-A is weak support for the single system model.

In a departure from what seems to be a recent trend, the effects of negative distance are all in the predicted direction. The column of b values indicates unanimous agreement in support of the single system construct. Therefore, based on the results in Table 28, hypothesis nine-B yields strong support for the single system model.

When hypothesis nine is reanalyzed using problem dyads comprised of five or more events we find some dramatic changes in the values of the coefficients. Results of this reanalysis are presented in Table 29. Looking first at the effects of positive distance, we see that both analyses of military behavior produce results counter to our expectations. Since neither system model can be supported by these findings, we consider type II evidence to be mixed. Moreover,
TABLE 29

Hypothesis 9:
Structural Direction and Reactive Behavior
(5 Event Threshold)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Intercept $a$</th>
<th>Positive Distance $b$</th>
<th>Negative Distance $b$</th>
<th>Coefficient of Determination $R^2$</th>
<th>$N^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall behavior</td>
<td>Capacity to act</td>
<td>48.23</td>
<td>8.78</td>
<td>-6.67</td>
<td>.06</td>
<td>429</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Military manpower</td>
<td>15.23</td>
<td>-10.23</td>
<td>3.86</td>
<td>.06</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>19.90</td>
<td>-3.69</td>
<td>-4.49</td>
<td>.08</td>
<td>43</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Commodity share</td>
<td>39.94</td>
<td>9.52</td>
<td>-14.60</td>
<td>.14</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Geographical share</td>
<td>35.67</td>
<td>-9.13</td>
<td>-8.17</td>
<td>.26</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>41.24</td>
<td>24.00</td>
<td>-31.24</td>
<td>.40</td>
<td>29</td>
</tr>
</tbody>
</table>

$^aN$-sizes are determined by event threshold levels of 10 or more events for overall behavior and 5 or more events for problem-specific behavior.
since the b value for trade behavior/commodity share surpasses that for overall behavior/overall structure, the type I comparisons also provide mixed support. Thus, when the problem-specific event threshold is set at five, hypothesis nine-A becomes ambiguous.

The major change affecting our evaluation of negative distance occurs in the trade behavior/trade structure analyses. At the five event level, both trade behavior coefficients exceed the coefficient for overall behavior/overall structure, resulting in mixed support on the type I comparisons. Since the type II evidence supporting the single system model does not change, we can infer that at the five event threshold hypothesis nine-B provides weak support for the single system model.

The final stratification hypothesis has to do with directional effects on the multilateral-unilateral dimension of behavioral autonomy. According to this hypothesis, behavior directed toward positively distant recipients is likely to be more multilateral than otherwise and, conversely, behavior toward negatively distant recipients is likely to be more unilateral than otherwise. Results for this hypothesis at the initial four event threshold are reported in Table 30. Once again, the intercept values reveal rather sharp differences in dyadic behavior across the different systems.

The type I comparisons for positive distance are split in their support for the two system constructs. Type I support for multiple systems is indicated by the fact that the b value for military behavior/military structure outdistances that for overall behavior/overall structure by a substantial margin. Obviously
TABLE 30

Hypothesis 10:
Structural Direction and Multilateral Behavior
(4 Event Threshold)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Intercept</th>
<th>Positive Distance</th>
<th>Negative Distance</th>
<th>Coefficient of Determination</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall behavior</td>
<td>Capacity to act</td>
<td>63.95</td>
<td>8.28</td>
<td>-10.69</td>
<td>.07</td>
<td>429</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Military manpower</td>
<td>35.69</td>
<td>24.46</td>
<td>-11.16</td>
<td>.22</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>36.91</td>
<td>.12</td>
<td>-11.35</td>
<td>.07</td>
<td>65</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Commodity share</td>
<td>80.22</td>
<td>4.38</td>
<td>1.19</td>
<td>.04</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Geographical share</td>
<td>79.95</td>
<td>3.90</td>
<td>.05</td>
<td>.02</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>79.81</td>
<td>-8.74</td>
<td>10.82</td>
<td>.05</td>
<td>53</td>
</tr>
</tbody>
</table>

aN-sizes are determined by event threshold levels of 10 or more events for overall behavior and 4 or more events for problem-specific behavior.
then, comparison of the overall behavior and trade behavior analyses favors the single system model. Since the one contradictory finding for positive distance involves the trade behavior/capacity to act relationship, the type II comparison of trade behavior analyses confers support to the multiple systems model. The extraordinarily large coefficient for the military behavior/military structure relationship adds further type II support for multiple systems. Based on the results in Table 30, we conclude that hypothesis ten-A provides weak support for the multiple systems model.

The findings for negative distance replicate the pattern of results for problem sensitivity in Table 27. The most prominent feature of this pattern is the contrary performance of all three trade behavior analyses leading to a mixed type II outcome. Since the overall behavior/overall structure relationship is in the expected direction, its type I comparison to the trade behavior/trade structure relationships supports the single system model. On the other hand, it is the multiple system model that is supported by the type I comparison to the military behavior/military structure relationship. Because this combination of mixed outcomes cannot support either system construct, the evidence from hypothesis ten-B must be deemed ambiguous.

As with several of the previous hypotheses, our initial evaluations were not invariant under alterations in the problem-specific event threshold. Table 31 presents results for hypothesis ten on problem dyads comprised of five or more events. Although there are
TABLE 31
Hypothesis 10:
Structural Direction and Multilateral Behavior
(5 Event Threshold)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Variable</th>
<th>Intercept a</th>
<th>Positive Distance b</th>
<th>Negative Distance b</th>
<th>Coefficient of Determination R²</th>
<th>N&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall behavior</td>
<td>Capacity to act</td>
<td>63.95</td>
<td>8.28</td>
<td>-10.69</td>
<td>.07</td>
<td>429</td>
</tr>
<tr>
<td>Military behavior</td>
<td>Military manpower</td>
<td>32.81</td>
<td>27.90</td>
<td>-12.30</td>
<td>.22</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>23.47</td>
<td>-11.97</td>
<td>-1.58</td>
<td>.10</td>
<td>43</td>
</tr>
<tr>
<td>Trade behavior</td>
<td>Commodity share</td>
<td>78.44</td>
<td>5.72</td>
<td>-1.77</td>
<td>.04</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Geographical share</td>
<td>82.74</td>
<td>5.25</td>
<td>4.75</td>
<td>.10</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Capacity to act</td>
<td>82.26</td>
<td>-9.40</td>
<td>17.74</td>
<td>.17</td>
<td>29</td>
</tr>
</tbody>
</table>

<sup>a</sup>N-sizes are determined by event threshold levels of 10 or more events for overall behavior and 5 or more events for problem-specific behavior.
changes in the magnitudes of the positive distance coefficients, the overall conclusion remains weakly supportive of multiple systems. There are changes in the negative distance results that do have some bearing on the comparison outcomes. One of these is the sharp decrease in the military behavior/overall structure coefficient and another is the sign reversal of the trade behavior/commodity share relationship. Together these two changes imply type II support for multiple systems. Since the type I outcome does not change, we may conclude that at the five event threshold level, hypothesis ten-B offers weak support for the multiple systems model.

The results encountered in this section of the chapter are, if anything, even more ambivalent than those for structural proximity. On several occasions investigation of directional hypotheses produced purely ambiguous results, though in each case this was at least partially attributable to contradictory findings for certain hypothesized effects. In addition, we found that evidence for or against a particular systemic construct often depended on the size of the problem-specific event threshold. At the initially specified four event level, hypotheses nine-A and nine-B provided weak and strong support for the single system model, respectively; however, with the increase to five events, hypothesis nine-A became ambiguous and nine-B moved from strong to weak support for the single system. The other change accompanying an increase in the event threshold was, of course, hypothesis ten-B's shift from the ambiguous category to weak support for multiple systems.
Summary

This chapter has presented a detailed examination of the empirical findings for each stratification hypothesis. Aside from our perfunctory evaluations on the type I and type II comparisons, interpretation of these findings has been deferred to the next, and final, chapter. Before taking up that task, however, it will be useful to reexamine these results on a more macro level for discernable patterns or trends that may be indicative of systemic organization.

A logical place to begin this overview is by inspecting the distribution of hypotheses within the three-by-three matrix defined by cross-classifying type I and type II comparison outcomes. This matrix was first presented in Chapter IV (Figure 8) as a means for consolidating evidence from both comparisons into a summary evaluation of systemic organization. Figure 16 locates each hypothesis, by number, into the cell designated by the results of our empirical analysis. Note that Figure 16 depicts only those results found at the four event threshold level. If one thing is clear about this configuration, it is that neither system model has gained overwhelming empirical support from our investigation. Results from analysis at the five event threshold are similarly displayed in Figure 17. Here we can see a slight shift in the direction of the multiple systems model.
### Type I Comparison

<table>
<thead>
<tr>
<th>Supportive of Single System</th>
<th>Mixed Support</th>
<th>Supportive of Multiple Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong Single System</td>
<td>Weak Single System</td>
<td>Ambiguous</td>
</tr>
<tr>
<td>9-B</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Weak Single System</td>
<td>Ambiguous</td>
<td>Weak Multiple Systems</td>
</tr>
<tr>
<td>7</td>
<td>8-A</td>
<td>1</td>
</tr>
<tr>
<td>9-A</td>
<td>8-B</td>
<td>2</td>
</tr>
<tr>
<td>10-B</td>
<td>10-A</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Supportive of Multiple Systems</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguous</td>
<td>Weak Multiple Systems</td>
<td>Strong Multiple Systems</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 16**

Summary of Results
(4 Event Threshold)
### FIGURE 17

**Summary of Results**

(5 Event Threshold)
Two aspects of the hypotheses' distribution along the "ambiguous" diagonal require further comment. In the first place, the clustering of hypotheses in the middle cell appears to be both unexpected and counterintuitive given the very low probability for this outcome cited in Figure 8. We should reiterate that for each of these hypotheses results from either the trade behavior or military behavior analyses unanimously contravened the predicted direction of the relationship. Under these circumstances, which, incidentally, were unanticipated in the calculation of probabilities, we elected to withhold support from either system model by automatically designating the type II outcome as mixed. Thus, it should be understood that the hypotheses located in the middle cell of Figures 16 and 17 resulted in mixed support in the conventional sense on type I comparisons, and mixed support as a means for dealing with contradictory findings on type II comparisons.

Second, we should take note of the conspicuous absence of hypotheses from the diagonal's two extreme cells. This distribution appears to suggest that the two comparison procedures are not independent of one another and, therefore, cannot be cited as separate sources of evidence regarding systemic organization. In fact, there is a lack of independence but it is more statistical than procedural in origin. Since our analyses have been carried out on a set of highly interrelated variables (see Table 9), it would be most unreasonable to expect a random distribution of results.
Although a useful summarizing device, the placement of hypothesis in Figures 17 and 18 provides a very coarse-grained overview that may obscure more finely textured, and possibly more conclusive, patterns of results. In the remainder of this section, our review takes a less broad-gauged approach in an attempt to winnow out such patterns. For instance, it may be that there are noticeable differences in the systemic implications of our findings depending on whether the hypothesized relationships concern the monadic level effects of structural position or the dyadic level effects of proximity or direction. In this particular case, such differences are easily ascertained in the concluding remarks of the three previous sections. It will be recalled that the monadic hypotheses (one through four) provided unequivocal support for multiple systems whereas both sets of dyadic hypotheses produced more ambivalent results.

A similar pattern of systemic support emerges when our findings are examined across different types of behavioral properties. All three hypotheses (one, two, and six) dealing with aspects of international involvement were supportive of the multiple systems model. This is particularly interesting because international involvement stands out in certain other respects, as well. For one thing, it is the only behavioral property included in our study that traditionally has been of concern to stratification researchers. Moreover, the measures of international involvement are the only ones recorded as raw frequencies and, perhaps concomitantly, their
scatter plots were the only ones to indicate the presence of significant nonlinearities. The analyses of behavioral autonomy (hypotheses three, four, seven, nine, and ten) were evenly divided in their support at the four event dyadic threshold level but evidenced a slight shift toward multiple systems at the five event level. It is more difficult to generalize about problem sensitivity because only one of its hypotheses (five) produced evidence that could be taken as supportive of either construct.

Thus, far, results have been considered supportive of one of the system models only if this was indicated by concurrent outcomes in both problem systems. Adherence to this criterion makes for a more robust test but, by the same token, it masks any differences between the trade and military problem areas. On those occasions in which type I and type II comparisons were in agreement regarding the appropriate model of systemic organization, we found that comparisons involving military behavior favored the single system construct three times and multiple systems six times; comparisons involving trade behavior, on the other hand, were evenly split with two supporting each of the system models. Note, however, that instances of agreement between the two types of comparisons include only half of all comparison outcomes. What is the distribution of supportive results if all comparisons are considered individually? By this method of accounting, comparison outcomes involving military behavior support the multiple systems model sixteen times and the single system model ten times. This pattern of support is
reversed for trade behavior comparisons with thirteen outcomes favoring the single system construct but only nine favoring multiple systems. Finally, let us delve even deeper beneath the surface of these results to examine differences in type I and type II comparison outcomes across the two problem systems. Table 32 summarizes the distribution of supportive outcomes along each of these dimensions. Arraying our findings in this way reveals some interesting patterns concealed by the coarser approach in Figures 16 and 17. For example, it is evident from Table 32 that type I comparisons do a considerably better job of discriminating between the systemic constructs than type II comparisons. Moreover, when type I comparisons are examined separately we find that military behavior comparisons favor the multiple systems model by a wide margin but that trade behavior comparisons just as clearly support the single system model. Of course, we must avoid reading too much into this pattern of type I outcomes. Taken at face value -- that is, without reference to inferred support for either model -- these results merely indicate that structural measures based on military manpower generally have a greater impact on military behavior than structural measures based on capacity to act have on overall behavior and, conversely, the capacity to act structural measures have a greater impact on overall behavior than the weighted share structural measures have on trade behavior.
TABLE 32

Summary of Comparison Outcomes Differentiated by Problem Area

<table>
<thead>
<tr>
<th></th>
<th>Military</th>
<th>Trade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I Comparison</td>
<td>Multiple system=10</td>
<td>Multiple system=4</td>
</tr>
<tr>
<td></td>
<td>Single system=3</td>
<td>Single system=9</td>
</tr>
<tr>
<td>Type II Comparison</td>
<td>Multiple system=6</td>
<td>Multiple system=5</td>
</tr>
<tr>
<td></td>
<td>Single system=7</td>
<td>Single system=4</td>
</tr>
</tbody>
</table>
NOTES

CHAPTER V

1 Space does not permit inclusion of all of the scatter plots examined during the course of this investigation. Accordingly, we have limited presentation of scatter plots to a few that are representative of the nonlinear forms present in the data. Scatter plots not included in this chapter are available from the author.

2 The intercept is the predicted value of the dependent variable when the independent variable is equal to zero. Obviously, it makes little sense to think of a rank score of zero. To avoid any confusion, the intercepts should be read as a base value for computing the predicted score of the highest ranking actor. For the linear form, this is accomplished merely by adding the unstandardized slope to the intercept value. In the top row of Table 17, for example, the slope (-3.25) would be added to the intercept (423.11) to yield a predicted score of about 420 total events for the government ranking highest on capacity to act.

3 So there is no mistake about what is intended by this comparison of $r^2$ values, we repeat that a type I comparison carried out on differently scaled variables must be based on the standardized regression coefficients. In the bivariate case, the standardized coefficient is equal to the product-moment correlation coefficient and, therefore, is also equal to the square root of the coefficient of determination. Since an inequality relation between two values remains true of the squared values as well, it makes no difference whether the comparison is between $r^2$ values or their square roots — the outcome will be the same in either case. We have elected to use the $r^2$ values simply as a matter of convenience.

4 We should point out that the nonlinear forms evident in Figures 10 and 11 might also be attributable to measurement error in either variable. For example, it could be argued that the highest ranking governments attract a disproportionate amount of attention from news organizations as well as from other actors. This would result in a systematic overreporting of events that could account for the greater than linearly expected levels of activity among highly positioned governments. Alternatively, the nonlinearities may have resulted from our decision to analyze rank positions rather
than actual values. This is because our measures of foreign activity are skewed to the right, but the use of ranked structural measures masks whatever right skew (left skew in terms of Figures 10 and 11) there may be in the distribution of actual values.

5It should be noted that we did examine scatter plots using the signed structural distance measure prior to the application of effect coding.

6We could just as easily have performed the analysis on the measure of external problems. The results would be the same except for a reversal in the signs of the $b$ coefficients. For the sake of uniformity, all regression analyses presented in this section used the behavior associated with positive direction as the dependent variable. Thus, each hypothesis predicts a positive coefficient for positive distance and a negative coefficient for negative distance.

7It may be helpful to illustrate how statements about the average rate of attention to internal and communal problems are derived from the regression results displayed in Table 27. The multiple regression equation underlying these results may be depicted as

$$Y = a + b_1(X_1) + b_2(X_2)$$

Where $X_1$ and $X_2$ stand for the effect coded positive and negative distance variables, respectively. Looking back at the effect coding formula in Table 16, we see that a positively distant dyad is coded 1 on $X_1$ and 0 on $X_2$. These figures along with the coefficients from the overall behavior analysis can then be inserted into the preceding equation to yield.

$$Y = 52.90 + 1.00(1) - 11.57(0) = 53.90$$

Since the negative distance coefficient is canceled out by the zero value, the result is determined by the coefficient for positive distance and, hence, may be interpreted as the average rate of attention to internal and communal problems among positively distant dyads. For negatively distant dyads the effect codes are reversed and the relevant equation becomes

$$Y = 52.90 + 1.00(0) - 11.57(1) = 41.33$$

Here the positive distance coefficient drops out and the result represents the corresponding rate of attention among negatively distant dyads.

8The discrepancy in the total number of comparison outcomes for the two problem systems (twenty-six for military, twenty-two for trade) arises because certain of the trade behavior results contravened the hypothesized relationship. In each case, comparisons among contradictory findings were omitted from consideration.
CHAPTER VI

CONCLUSION

We began our study by delineating the main elements of two competing models of systemic organization and devising a research strategy for evaluating them based on a presumed functional relationship between system structure and patterned foreign policy interactions. A set of ten stratification hypotheses were then introduced along with methods for measuring the key concepts in each. We then designed procedures for conducting the analysis and interpreting the results as supportive of either the single system or multiple systems model. Finally, in the last chapter, the empirical findings were reported and summarized. The next major step is to spell out the implications of these findings and to assess what they add to our knowledge of the systemic configuration of international politics.

To begin with, it would be less than honest to proceed as though our investigation had fulfilled its principal objective of indicating which system construct provides the better approximation of its intended empirical referent. On the whole, the results presented in Chapter V are not particularly supportive of either model. This outcome is somewhat anomalous when viewed against the conceptual and methodological parameters of our design, although this does not necessarily mean that
the findings themselves are either uninteresting or somehow invalid. What it does mean, however, is that to account for these findings we must reevaluate one or more of our study's underlying assumptions. Let us examine this point a little more closely.

In accordance with Lakatosian falsificationism, our research strategy was designed as a three-cornered struggle among two competing theoretical formulations and the results of empirical observation. Because the theoretical formulations— the two system models— were depicted as logically inconsistent with one another, it was assumed that at most only one of them could be consistent with the observational data. As it turns out, however, our findings indicate that both of them are partially consistent (or inconsistent) with the data. There would seem to be two ways to account for this anomaly. The first is to retain the inconsistency of the two system perspectives and instead focus suspicion on the integrity of the empirical findings. Put differently, this approach treats the findings as an artifact of the design. The second approach is to search for an explanation on a more theoretical level by revising one or both systemic constructs in light of the empirical findings. Note, however, that both approaches call into question certain necessary assumptions of our inquiry: the former concentrates on design assumptions concerning such matters as the effects of measurement error or the specification of problem systems whereas the latter focuses on the logical assumption that the two system models constitute mutually exclusive images of the international system.
Findings as Artifact

We have seen that our appeal to empirical observation has failed to demonstrate the superiority of either system construct. As stated above, this result may be more a function of observational error or other design difficulties than an accurate reflection of the systemic configuration overlaying world politics. In this section we will examine some plausible sources of design error that could account for the pattern of results exhibited in the previous chapter. Moreover, since it is likely that different modes of systemic organization are vulnerable to different sources of error, this examination will proceed in two parts: first, under the assumption that the traditional image of a single, comprehensive international system is the more appropriate representation and, second, under the competing assumption that the world more closely resembles the model of multiple, problem-centered international systems.

Before we engage in this two fold examination, special notice should be given to a particular class of design errors that, if present, would entail equally devastating consequences under either organizational scheme. We refer to the possibility of error within the "hard core" of assumptions common to both system variants. To recapitulate from Chapter I, the fundamental precept of Lakatosian falsificationism states that theories are not disconfirmed by facts alone; rather, they are superseded by other theories within the context of a research program that coalesces around a core of assumed truths. As the defining
characteristics of a research program, these core assumptions are insulated from the constant modification and adjustment of succeeding theories by a second, less fundamental set of "protective belt" assumptions. It is only when a program fails to progress - that is, when a prevailing theory successfully withstands all efforts aimed at replacing it - that the program's core assumptions are subjected to critical appraisal.

In defining the central core of the present study, we have assumed, among other things, that international systems are structured hierarchically by their distribution of problem-solving capabilities, that there is a causal process linking system structure and recurring patterns of interaction, and that the empirical implications of this process are captured, in part, by the ten stratification hypotheses. All of these assumptions are important but the latter is most visibly so because it allows the hypotheses to be used as a standard for evaluating the empirical relationships observed under alternative representations of systemic organization. Obviously, if one or more hypothesized relationships are mistaken then the paired comparisons on those hypotheses become utterly meaningless for the purposes of this investigation.

The aim of this brief digression is to acknowledge the possibility that our results may be merely a capricious product of erroneous core assumptions. Having done that, we must hasten to add that there are several reasons why this particular source of error should not be given
serious consideration as a potential explanation of our findings. In
the first place, the vast majority of empirical results cited in the
previous chapter tend to corroborate the predicted direction of the
hypothesized relationships. Of course, many of these relationships are
quite weak and others are undoubtedly somewhat inflated due to news
reporting bias; nevertheless, the extent to which the prima facie
evidence supports the hypotheses is an encouraging sign. Secondly, we
must take note of Lakatos' "positive heuristic", the maxim which states
that criticism and revision ought to be centered on a theory's "protective
belt" assumptions. Thus far, our study has focused on alternate forms
of just one "protective belt" assumption dealing with the specificity
of systems and the rules for demarcating their boundaries. It should
be recognized, however, that there are numerous other assumptions
comprising the "protective belt" whose modification can provide a
satisfactory explanation of our findings without recourse to the "hard
core". Finally, we can reiterate the Lakatosian prescription that
critical appraisal of core assumptions ought to be based on the
progression of theories within a research program and not on the
findings of a single study.

Assuming a Single International System

Having decided to abide by the Lakatosian formula and eschew
critical scrutiny of our study's core assumptions, we are left to
search elsewhere for an explanation of our findings. As noted above,
we will begin this task by assuming there is but a single,
overall system and then exploring ways that this "true" configuration could have been distorted to produce the mixed bag of results presented in Chapter V.

Certainly one factor that would help to account for these results is our research design's subtle bias against the single system model. This bias stems from our decision in Chapter III to emphasize the divergence of putative system structures as a criterion for choosing structural measures to be used in the subsequent analyses. It will be recalled that we initially proposed twelve candidate measures of positional structure, four for each of the three putative international systems. We examined the congruence of different measurements within each group as well as the divergence of measurements across groups, but when it came time to select the appropriate indicators for analysis the latter property was considered paramount. This was by no means a neutral decision since by accentuating positional differences the divergence criterion was instrumental in establishing a necessary precondition for the multiple systems approach. Even more importantly, the divergence criterion may have steered us away from a more appropriate measure of overall system structure.

Although beyond the scope of the present research, it would be possible to investigate this matter more systematically by repeating the analyses using structural measures chosen according to alternative criteria. For example, a subsequent study might impose a similar sort of bias in the opposite direction by selecting structural measures on
the basis of their congruence across putative systems. Another approach would combine elements of both criteria in order to strike a balance between the biases of each. Under this method the problem structure measures would be chosen so as to maximize their divergence from one another; the overall measure, on the other hand, would be selected on the basis of its congruence with the problem measures.

Other possible sources of design error that may have beclouded the underlying single system arrangement are to be found in the procedures adopted for the observation and measurement of overall system structure, governmental behavior, or both. As outlined in Chapters III and IV, virtually all of the data used in this study are susceptible to multiple sources of distortion or contamination. Moreover, because such error can be compounded across succeeding phases of a complex operational procedure, its ultimate effect on the magnitude of the observed relationships may be quite unpredictable. Nevertheless, we can identify several potential sources of measurement error that could sufficiently attenuate the overall structure relationships to significantly reduce the likelihood of our finding empirical support for the single system model.

Three principal sources of error are likely to affect the measurement of capacity to act, the concept selected to represent overall system structure. One of these is, of course, the presence of inaccurate, misleading, or incomparable information in the sources for the aggregate indicators of size, modernization, and societal stress. In Chapter III
we suggested that the use of multiple indicators was to be considered a strength of the capacity to act measure because it would alleviate the impact of this type of error. It is possible, however, that reliance on multiple indicators actually aggravated this problem for some countries, particularly the lowest ranking ones, by compounding the error effect across indicators. A second possible source of error concerns the choice of indicators to operationalize the component concepts of size, modernization, and stress. In other words, even if the indicator itself is not marred by erroneous information, it still may be an invalid measurement of a given concept. Finally, we must also mention the considerable room for error in the set of procedures used in distilling a summary capacity to act score from the mass of information contained in the raw indicators.

Similar kinds of problems are apparent in the event-based procedures used to operationalize governmental behavior. Here, however, our discussion must be expanded to include possible errors of measurement for problem-specific behavior as well as for overall behavior. Two potential sources of error seem especially relevant to these measures. The first has to do with the restricted temporal domain of the event data. Although the behavioral patterns analyzed in this study are supposed to represent broad behavioral tendencies occurring over the course of an entire decade, the raw events actually cover only a quarter of this period. Moreover, if the temporal sampling of events does have a distorting effect, it is likely to be more significant for patterns of overall behavior than for the more homogeneous problem-
specific patterns. The second source of error is one already discussed in some detail. We refer to the instability of percentagized measures computed for patterns composed of only a few events. This source of error is not unrelated to the first since augmenting the temporal coverage of the data would have the concomitant effect of increasing the base number of events thereby reducing the degree of instability in the percentagized measures.

In this part of the chapter we have identified a variety of potential difficulties that together or singly may have prejudiced our inquiry against an outcome favorable to the single system perspective. Before we extend this discussion to errors having the opposite effect, we should point out that one aspect of our findings remains somewhat incongruous with the single system construct even after these design problems are taken into account. This aspect is the slight shift toward support for the multiple systems model when the hypotheses were reanalyzed at the five event threshold. To be sure, this shift involved no more than three hypotheses and could be explained as an artifact of random measurement error. Nevertheless, it is curious that we found no comparable shift in the opposite direction when the analysis was restricted to behavioral patterns that may be less prone to error than those examined at the four event level.

Assuming Multiple Problem Systems

At this juncture let us modify the basic premise of our discussion by assuming that intergovernmental political relations are conducted in
a manner that more closely parallels the multiple systems model than the conventional image of a single overall system. Although the findings reported in Chapter V are somewhat more supportive of this assumption than the last, they are still a long way from providing convincing evidence in favor of the multiple systems approach. Therefore, we will proceed much as before by reexamining our design for potential sources of error or bias that would account for these results.

First of all, we cannot discount the possibility that the multiple systems configuration has been obscured by excessive amounts of measurement error in the problem-specific variables. For example, the limited temporal coverage of events and the instability of percentagized behavioral measures, both counted as potential sources of error for the single system model, are no less problematical under the multiple systems assumption. Since both factors are likely to result in random (rather than systematic) error, their effect would be to attenuate the observed relationships and thereby lessen the chances of our finding empirical support for multiple systems. Furthermore, since these sources of error are likely to be most debilitating to behavioral patterns containing the fewest number of component events, patterns defined at a higher event threshold would be expected to provide a truer reading of the system's underlying configuration. By this line of reasoning, the shift in our results at the five event threshold is fully compatible with the multiple systems construct.

Of course, any discussion of measurement error must go beyond
the behavioral side of these relationships to consider the equally vulnerable indicators of problem structure. Several potential sources of error are pertinent here. For instance, just as with the aggregate statistics contributing to capacity to act, there is always the possibility of inaccurate or incomparable values in the trade data and military manpower figures. An even more critical sort of error concerns the validity of these indicators as measures of substantively specialized problem-solving capabilities. First, as noted in Chapter III, a government's military capability may depend not only on the size of its armed forces but also on their training, organization, and access to material and weaponry. Second, the commodity-weighted share of world exports is weakened considerably by its failure to take into account such obviously important factors as the demand elasticity and strategic value of specific commodity groups. Finally, given our initial indecision regarding the relative merits of the two weighted-share indices and the failure of our empirical analysis to demonstrate any appreciable differences between them, it might be argued that we have not yet achieved a suitable operationalization of trade structure. At least this would explain why the trade system analyses tended to support the single system model more often than not.

We have seen that design errors involving inaccurate or misleading measurement procedures can account for our findings regardless of the system's underlying configuration. Next we will consider a potential source of difficulty that is applicable only to the multiple systems model. This concerns the possibility that our characterization of
problems dealing with international trade and the settlement of military
security disputes is inappropriate for defining substantively distinct
systems under the multiple systems approach. The most worrisome aspect
of problem specification has to do with the level of abstraction used
to delineate a problem's characteristic substantive content. If
problems are specified either too broadly or too narrowly we are likely
to get a distorted picture of the empirical relationships between
positional structure and behavior. In operational terms, this means
that the concepts on both sides of the relationship must be specified
at appropriate and, of course, equivalent levels of abstraction.

Finally, let us briefly mention one other potential source of error
that is especially relevant to the multiple systems model. Throughout
this study we have maintained that governmental foreign policy
behavior is the product of numerous other factors in addition to the
systemic level effects of the stratification process. Unless these
confounding influences are brought into the analysis it is virtually
impossible to obtain a true (within the limits of measurement error)
reading of the impact of positional structure. However, by assuming
that extraneous factors have approximately the same cumulative effect
within each putative system, then such factors can be ignored so long
as our results are used for purposes of comparison across systems.
The importance of this assumption for the present inquiry cannot be
overstated - if it is mistaken then we are unable to infer that differ­
ences in the magnitudes of coefficients are due only to differences in
the systemic constructs guiding our analysis. For example, let us
suppose that certain causal factors tend to be more or less potent depending on a problem area's substantive content. If this were the case, then the relationships observed in different problem systems would be contaminated by different confounding influences which might either dampen or reinforce the hypothesized structural effects. Obviously, under such conditions there would be little chance of detecting the underlying systemic configuration by the method of comparisons used in this study.

The foregoing discussion has left us with a sobering reminder of the many uncertainties and potential difficulties besieging our inquiry. The point in mentioning these problems here is to demonstrate that we cannot rule out either of the system constructs based on the evidence of this one study. Nevertheless, we should also note that the questions raised above represent potential difficulties, they are by no means proven ones. Although it is most unlikely that such problems are entirely absent from this study, neither can we be sure that the findings are wholly an artifact of the design. To the extent that these findings do signify something other than the vagaries of observational or design error, we must revise our initial premises regarding the incompatibility of the single system and multiple systems models. Some alternative and necessarily more speculative interpretations of these findings are considered in the following section.
Alternative Interpretations

From the very beginning of this study we have portrayed the single system and multiple systems models as competing representations of intergovernmental relations. This competition was predicated on the belief that the two models comprised alternative and fundamentally incompatible images of the same real world referent. Thus, in order for one model to be shown superior to the other (in the sense of providing the closer approximation to the empirical referent) it must be judged consistently superior over a broad range of observational data. This, of course, is just what our results do not show. In the previous section, attention was given to the many ways in which these findings could be merely an artifact of our measurement or analysis procedures in which case either model could prevail over the other irrespective of the mixed and seemingly anomalous outcome of the present study. Here we follow a different tack by relaxing the incompatibility assumption to discover what alternative models or explanations would account for our results' divided support of the two system constructs. Three plausible and partially overlapping interpretations will be presented, along with some initial suggestions for the further investigation of each.

System Transition

Perhaps the simplest way to account for our findings is by means of a system transition model. There can be little doubt that systemic level features of world politics do change over time, although it is not always clear how such changes are brought about or even where they
may lead. As noted earlier in Chapter III, many observers have pointed to the early 1970s as an important transitional period. For example, Seyom Brown (1973:286) perceived the main trends of this period as "...a disintegration of the cold-war coalitions, the rise of nonsecurity issues to the top of diplomatic agendas, and a diversification of friendships and adversary relations." Similar characteristics are put forward by Keohane and Nye (1977) as the key ingredients of what they term "complex interdependence". They contend, first, that societies are increasingly connected by channels other than formal interstate relations among governmental elites; that where multiple issues are considered they are no longer arranged in a strict hierarchy dominated by military security; and, finally, that the use of military force has ceased to be a viable instrument of foreign policy. We should add that Keohane and Nye do not envision complex interdependence as the current state of global affairs but only as an ideal type that is increasingly well approximated on some issues and in some bilateral and multilateral relationships.

What are the significant and long-run consequences of these changes? The authors mentioned above, among others, suggest that goals and the power resources to effectuate them will become tied to specific issues areas. Although the outcome is by no means certain, the trend would seem to be in the direction indicated by the multiple systems model. Furthermore, we would not expect such deep-rooted changes to take shape in a matter of only a few months or years but rather over the course of
several years or even decades. Thus, even though reports of these
trends began to surface in the early seventies, it is likely that the
trends themselves are expressions of deeper currents originating as
far back as the mid to early sixties.

It is possible, then that the ambivalence of our empirical findings -
far from being simply an artifact of the design - actually manifests
a transitional phase from a systemic configuration resembling a single,
comprehensive system to one more in accord with the multiple systems
construct. Although this interpretation fits the general outcome of
the study reasonably well, it offers few insights into some of the more
specific patterns of results. For example, the transition explanation
is unable to account for the apparent shift toward the multiple systems
model evidenced at the five event threshold; nor does it help us to
understand why individual hypotheses were found to support a particular
construct. This, of course, is not to say that these findings could
never be explained by a system transition model, only that they are
presently inexplicable given our scant knowledge of global systems
and their transformation.

In principle, it would be relatively easy to test the system tran-
sition hypothesis merely by replicating the present investigation at
different points in time. Ideally, such a study would encompass two
widely spaced temporal domains, such as the mid 1950s and the mid 1970s,
to allow for the maximum possible change between observations. Moreover,
it would be advisable to narrow the temporal domain of each observation
period from the decade used in this study to something between two to
five years in order to lessen the chances of significant system transformation within an observation period. In practice, however, a replication design is severely encumbered by the time and resources necessary to generate the requisite behavioral data.

On balance, then, while the system transition interpretation may be a satisfactory explanation for the mixed outcome of this study, it does not appear to account for the full range of observed findings. Furthermore, even granting that the decade under investigation was marked by some transformation of systemic level characteristics, it is by no means certain that the direction of this change was toward a multiple systems configuration. Let us, then, see what other possible models may be indicated by the results.

**Split System Model**

An alternative explanation for our findings may be conceived as a split system or branching model of international positional structure. This model embodies elements of both system constructs by positing a configuration that resembles a single, undifferentiated structural arrangement among the middle and lowest ranking nations but, at the highest levels, branches out into what appear to be multiple, issue-specific hierarchies. This branching occurs because the highest ranking actors are also the most salient; they are perceived as more important and their capabilities are examined more carefully and with finer distinctions than are other actors'. For the most part, the elite nations arrayed on any one branch structure are likely to be
the same as those located on other problem area branches, although their relative positions are apt to differ from branch to branch. Moreover, we would expect these top ranking actors to give sustained attention to a wide range of issues and, hence, to generate reasonably consistent patterns of behavior within particular issues but not necessarily across them. On the other hand, nations occupying lower positions probably attend to fewer issues but are more consistent in their overall behavior. In short, problem-specific capabilities count most for the most prominent actors whereas these distinctions become blurred for those in lesser positions.

This model is compatible with the results in one key respect: it accounts for the observed shift in favor of the multiple systems model when the analyses were repeated on problem dyads composed of five or more events. We arrived at this conclusion after a thorough inspection of our data that paid particularly close attention to the number of events in each dyadic case. As one might expect, we discovered that the cases dropping out of the analysis at the higher event threshold were generally those involving low ranking actors. This finding raised an intriguing question: was the shift in support primarily a function of greater accuracy in the behavioral patterns or was it due to the omission of most low ranking actors? We investigated this question by performing another round of regression analyses, this time admitting only those dyadic cases in which the acting government ranked in the top third of the hierarchy. The results, which are
not reported here, produced a pattern of support replicating that found at the five event threshold. Although a definitive answer must await further research, these findings suggest that the multiple systems construct is better approximated by higher ranking actors in the manner predicted by the split system model.

Two additional features of this model deserve brief attention before we consider what further studies are required for corroborating evidence. First, and most important, there is an alternative and reasonably compelling explanation for why governmental behavior might appear as though it conformed to the split system model. As noted in Chapter IV, some analysts have maintained that news selection and reporting are heavily biased toward more extensive and detailed coverage of elite countries at the expense of small and poor nations in the Third World. If this is indeed the case, then observed differences in the behavioral patterns of high and low ranking nations may actually reflect differences in the precision of our event-based measurement procedures. Second, it is worth pointing out that the split system model is not necessarily distinct from the system transition interpretation outlined above. It may be, for example, that the branching configuration posited by this model is only an intermediate and quite transitory phase in the transformation process. Although we can do no more than conjecture, this line of reasoning suggests that the transition from a single system to multiple systems begins at the highest levels and then proceeds downward through the hierarchy.
Like the transition interpretation, the split system model cannot be accepted without further investigation based on a more extensive body of data. More specifically, such a study must have behavioral data available for a sufficiently large number of acting governments to permit independent analyses of the single system and multiple systems regions of the hierarchy. Moreover, very special care must be taken in the generation of these data to avoid distortion from biased news reportage that might prejudice the results in favor of the split system model and, in the process, increase the likelihood of what statisticians refer to as Type II error. Perhaps the most one can do in this regard is to utilize multiple and regional or country-specific event sources, although for reasons cited in Chapter IV, this strategy cannot guarantee the data's validity. We might also note that it would be possible to combine this "split system" design with the "system transition" design discussed earlier to test the hypothesis that distinct problem systems evolve by branching off from the top of the prevailing overall structure.

Linkage Model

The final model that shall be considered here is distinguished by its relaxation of the "no linkage" assumption advanced in the first chapter. This assumption, it will be recalled, simplifies reality under the multiple systems model by circumscribing structural effects to particular problem areas. In the absence of this assumption, that is, when cross-issue linkages are taken into account, the substantive
boundaries between what are ostensibly distinct systems can erode or even disappear altogether. Let us explore this matter more fully.

We have maintained that, in general, an international system is delineated by a distinctive arrangement of its constituent units—its positional structure—as well as certain resultant patterns of unit interactions; the multiple systems model then goes a step further by imposing the additional constraint of substantive homogeniety. When linkages occur across substantive areas, whether through overt bargaining processes aimed at extracting side payments or by less explicit means such as an inhibition to act in one area for fear of reprisals in another, the affected system(s) will experience some loss of coherence owing to the dilution of its (their) characteristic substantive core. Furthermore, when linkage strategies are successful and widespread, political processes and outcomes will follow similar patterns across a variety of problem areas since individual problem structures will give way to an overall structure based on the fungibility of power and the ability of governments to capitalize on their strongest areas (Keohane and Nye, 1977).

Of course, much depends on the incidence and patterning of specific linkages. If they tend to be rare occurrences distributed randomly across issues and actors their cumulative effects may be no more than a slight blurring of system boundaries which can probably be ignored for most purposes. On the other hand, if linkages are fairly common or recur in definite patterns the system configuration may be altered
considerably. Actually, many different configurations can be envisaged under the general rubric of a linkage model. For example, a persistent linkage between seemingly unrelated problems might eventually result in one of the problems being fully assimilated into the other's system; alternatively, the two problems could become permanently linked in a nascent problem system of their own. In either case, the referent configuration would look like the multiple systems model in all but the substantive coherence of its individual systems. Another variant of the linkage model can be imagined in which linkage strategies are quite prevalent in dealing with all but a small handful of well insulated problems. These circumstances would lead us to anticipate a configuration comprised of a few discrete problem-specific systems orbiting a larger, more centralized system undifferentiated by issue area.

In the absence of further knowledge regarding the extent and patterning of linkage processes we are quite limited in our ability to draw any specific connections between the observed findings and some variant of the linkage model. In fact, about all that can be said with any confidence is that our results are not inconsistent with the issue linkage idea. Moreover, unlike the transition and split system interpretations offered above, a more conclusive assessment of the linkage hypothesis cannot proceed simply by replicating the present analyses across different periods of time or groups of actors. Each of the previous models is quite explicit about what additional evidence would be required before it could be deemed a satisfactory explanation for
the mixed results of this study. The same cannot be said of the linkage model. The reason is that we do not yet know enough about how and why linkage phenomena occur to be able to predict their cumulative effect at the systemic level. A research agenda purporting to fill this gap would have to address at least the following questions.

First, under what conditions do governments link their own behavior on some issues to other actors' policies on other issues? One might begin to investigate this question with Rosenau's (1966) hypothesis that issue linkages are least likely to occur between actors that hold divergent goals in a single area but are in agreement on most other matters and between actors that can agree in one area but otherwise pursue divergent goals. Alternatively, attention might be given to disparities in actor capabilities as a possible source of issue linkages. It has been suggested, for example, that whenever possible actors will seek to compensate for their weaknesses by drawing linkages to their strongest areas (Keohane and Nye, 1977). Second, are certain types of actors more likely than others to pursue linkage strategies? One characteristic that seems worthy of further investigation is the complexity of a nation's domestic interests—according to Keohane and Nye (1977), the greater the complexity the more difficult it will be to draw effective linkages between unrelated issues. This proposition is especially interesting because of its potential systemic level consequences. Assuming that the highest ranking nations are also those with the most complex domestic constituencies, the proposition implies that
reliance on linkage strategies will vary inversely with hierarchical position. Finally, are there different kinds of linkage strategies in common use by contemporary governments? For example, it may be useful to distinguish linkages occurring as part of a bargaining process from those employed as a retaliatory measure against the prior actions of others.

Directions for Future Research

Having introduced a variety of plausible explanations for the empirical findings of this study, it would appear that any definitive conclusions will have to await the outcome of still further research. One matter that must be given close consideration in any future studies of the systemic organization of world politics concerns the question of how one might choose among the alternatives discussed in the preceding pages. Of course we cannot rule out the possibility that some entirely different interpretation would be better able to account for the results; nevertheless, in closing it will be useful to examine some of the issues sure to be involved in any attempt to untangle the various strands of explanation outlined above.

Certainly any future research on this topic must seek to gain some control over the most problematical aspects of the design and measurement procedures used in the present study. One issue particularly deserving of attention is the level of aggregation or expansiveness of the problems or problem areas serving as substantive
nuclei for the discrete systems posited under the multiple systems approach. For the most part, our specification of international trade and military settlement problems was guided by an existing classification scheme (Hermann and Coate, 1979); in retrospect, however, it seems possible that these two categories have been defined too broadly for the purpose of delineating discrete problem systems. It may be, for example, that trade problems should be further subdivided according to particular commodities such as petroleum, or by classes of manufactured goods such as conventional arms. Moreover, we should not expect that problems underlying different systems will be specified at comparable levels of aggregation. For instance, if petroleum trade were found to require special treatment as an individual system, this should not preclude specification of a second, more inclusive but less homogeneous system dealing with trade in all other raw commodities and agricultural products.

Interestingly enough, the methods employed in the present study to comparatively evaluate the single system and multiple systems approaches would also be appropriate for investigating the system specification problems mentioned above. In both instances the research question is one of determining at what level of substantive aggregation stratification processes exert the most clearly discernable influence on patterns of interaction. In the present case, however, interest would be focused on successive levels of aggregation within a generalized problem area. Thus, the trade system examined in this study might be compared to more narrowly specified systems covering
trade in commodities on the one hand, and trade in manufactures on the other. Similarly, a commodity system could be compared with further subdivisions dealing with, for example, trade in agricultural products, trade in mineral ores, and trade in mineral fuels. In the end this process may involve experimentation with several different levels before a set of reasonably satisfactory systems is finally identified.

Note that as systems are defined at successively narrower levels of aggregation it will become increasingly necessary to address another of the potential sources of error discussed earlier in the chapter. We refer to the imprecision and lack of stability caused by the small number of events aggregated to form the monadic and dyadic patterns of behavior. There are essentially two ways to increase the number of events comprising a behavioral pattern: one is to lengthen the time span for which events are identified and the other is to expand the number of event sources. Although the former method is likely to be more effective in the short run, both methods will have to be exploited to their fullest extent if very low level systems are to be examined in the analysis.

Another measure that might be taken to reduce the potential for error would involve the identification and explicit control of any confounding factors that may be influencing the bivariate relationships between positional structure and behavior. This means that suspected confounding factors must be brought into the analysis. The simplest way to accomplish this is by incorporating these factors
as additional independent variables in a multiple regression analysis. For instance, if a high degree of geographical proximity were thought to promote collaborative intergovernmental behavior on certain types of issues, than geographical proximity ought to be included alongside structural proximity to control for this effect. Depending upon the number of suspected confounding factors and the complexity of their interrelationships, one might also turn to more formal causal modeling methods. The application of path analysis, for example, would enable the researcher to estimate the structure's direct causal impact on behavior as well as whatever indirect impact it might have through the influence of intervening variables. With sufficient theoretical development, path analysis could be used in much the same way that simple regression analysis was used in the present study. By this more sophisticated method, support for one or the other system model would be determined by comparisons of direct and indirect causal effects derived from multivariate causal models rather than by comparisons of individual regression coefficients.

Even if these and other efforts to address the most likely sources of error in our present design should meet with some success, it is possible that future studies still will fail to provide any clear distinction between the single system and multiple systems models of world politics. If empirical findings should continue to be ambivalent with respect to the two system models this would strengthen the case for further exploration of the alternative
interpretations outlined earlier in this chapter. We have already pointed out that investigation of the system transition and split system models can proceed as a straightforward extension of the methods used in the present study. The key feature of any research design for investigating these interpretations is the imposition of controls to facilitate detection of predicted variation in the results.

Owing to its greater complexity and more primitive development, the issue linkage model will require a somewhat different treatment. At least initially, issue linkage research ought to follow a traditional case-oriented mode of inquiry focusing on the handling of particular international issues or on a set of issues within particular bilateral relationships. Knowledge accumulated from the analysis of individual cases will serve an important heuristic function by contributing to the further refinement of the linkage concept and by providing an initial indication of the conditions likely to promote or hinder the occurrence of issue linkages. In addition, such studies might also be suggestive of some specific hypotheses concerning the systemic level implications of foreign policy linkage strategies.
LIST OF REFERENCES


BATH, C. RICHARD and DILMUS D. JAMES (1976) "Dependency analysis of Latin America: some criticisms, some suggestions," Latin American Research Review. Vol. 11, No. 3 (Fall).


_________ (1973) "Code manual for analytic deck of comparative foreign policy events--revised." Columbus: Ohio State University, mimeo.


_________ (1966a) "International relations and international conflicts: a sociological approach," Transactions of the Sixth World Congress of Sociology (Evian).


GANTZEL, KLAUS JURGEN (1973) "Dependency structures as the dominant pattern in world society," Journal of Peace Research, No. 3.


MASTERS, ROGER D. (1964) "World politics as a primitive political system." World Politics. Vol. 16, No. 4 (July 595-619.

MAYER, L.S. (1971) "Comment on 'The assignment of numbers to rank order categories.'" American Sociological Review Vol. 35 (August 916-917.)


SALMORO, BARBARA G. and LINDA P. BRADY (1972) "CREON supplemental and descriptive coding manual: revised." Columbus: Ohio State University, mimeo.


