INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
A theory of escalation: The use of coercive bargaining strategies in international conflict

Carlson, Lisa Jayne, Ph.D.
Rice University, 1994
RICE UNIVERSITY

A THEORY OF ESCALATION: THE USE OF COERCIVE BARGAINING STRATEGIES IN INTERNATIONAL CONFLICT

by

LISA J. CARLSON

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE

DOCTOR OF PHILOSOPHY

APPROVED, THESIS COMMITTEE

Dr. T. Clifton Morgan, Chair
Department of Political Science

Dr. Richard J. Stoll
Department of Political Science

Dr. Rick K. Wilson
Department of Political Science

Dr. Suchan Chae
Department of Economics

Houston, Texas

May, 1994
ABSTRACT

A THEORY OF ESCALATION: THE USE OF COERCIVE BARGAINING STRATEGIES IN INTERNATIONAL CONFLICT

by

LISA J. CARLSON

This thesis is a theoretical and empirical investigation of the escalation process that results when states are in dispute over some issue(s) in the international system. Escalation is viewed as a cost-imposing bargaining strategy enacted by states for the purpose of eliciting concessions from the adversary. Each state has a cost tolerance which identifies the maximum costs a state is willing to suffer to achieve their demand on the issue at stake. A player determines the likelihood that escalation will produce the adversary's concession by comparing the cost tolerances of both players. Once a player decides that the opponent has a greater willingness to bear the costs of escalation, that player concedes and quits the escalation game. A formal theory of the escalation process is developed which produces several hypotheses identifying the conditions under which states are expected to escalate in conflict and, once the decision to escalate is made, the level of escalation that that state is likely to achieve. One advantage in developing a general theory of escalation is that the interrelationships among the variables are expected to hold across a wide variety of escalation contexts. These hypotheses are then tested empirically in order to assess the utility of the general theory of escalation. In one of these contexts, the extended deterrence crisis, potential attackers are expected to refrain from pursuing higher levels of escalation when the attacker can perceive the value that the defender attaches to its ally and when the
defender has the capacity to impose serious escalation costs on its opponent. Another of the hypotheses tested produced the expectation that lower cost tolerant actors were more likely to achieve their highest level of escalation on their first move in the conflict given that they were unable to prevail in a long, drawn-out game of escalation with their stronger opponent. The results of the empirical tests indicate that the general theory of escalation developed in the thesis is useful in identifying the conditions that motivate states to escalate, the level of escalation that is likely to be achieved and the conditions under which states concede.
ACKNOWLEDGMENTS

This thesis is dedicated to my Mother and Father.

I would especially like to thank Cliff Morgan for his guidance, wisdom and infinite patience throughout the entire process. I am, without question, a better scholar as a result of his tireless efforts.

I would also like to express my indebtedness to the rest of the thesis committee: Ric Stoll, Rick Wilson and Suchan Chae.

Special recognition must be given to my dear friends, Eric Browne who provided a constant source of strength and inspiration from beginning to end, and to LaVonna Blair for her encouragement, support and friendship.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>What are We Studying</td>
<td>3</td>
</tr>
<tr>
<td>Method of Analysis</td>
<td>8</td>
</tr>
<tr>
<td>Organization of the Thesis</td>
<td>10</td>
</tr>
<tr>
<td>II. LITERATURE REVIEW</td>
<td>13</td>
</tr>
<tr>
<td>Outcomes of the Escalation Process</td>
<td>15</td>
</tr>
<tr>
<td>System Level and Psychological Escalation Dynamics</td>
<td>24</td>
</tr>
<tr>
<td>Escalation as a Bargaining Strategy</td>
<td>31</td>
</tr>
<tr>
<td>Conclusion</td>
<td>34</td>
</tr>
<tr>
<td>III. THEORY AND MODELS</td>
<td>36</td>
</tr>
<tr>
<td>The Escalation Context</td>
<td>39</td>
</tr>
<tr>
<td>Model I</td>
<td>52</td>
</tr>
<tr>
<td>Assumptions</td>
<td>52</td>
</tr>
<tr>
<td>Solution Concept and Equilibria Outcomes</td>
<td>58</td>
</tr>
<tr>
<td>Summary of Model I</td>
<td>64</td>
</tr>
<tr>
<td>Model II</td>
<td>67</td>
</tr>
<tr>
<td>Assumptions</td>
<td>67</td>
</tr>
<tr>
<td>Equilibria Outcomes</td>
<td>70</td>
</tr>
<tr>
<td>Hypotheses</td>
<td>72</td>
</tr>
<tr>
<td>Summary of Model II</td>
<td>74</td>
</tr>
<tr>
<td>Model III</td>
<td>78</td>
</tr>
<tr>
<td>Assumptions</td>
<td>78</td>
</tr>
<tr>
<td>Solution Concept and Equilibria Outcomes</td>
<td>81</td>
</tr>
<tr>
<td>Hypotheses</td>
<td>84</td>
</tr>
<tr>
<td>Summary of Model III</td>
<td>85</td>
</tr>
<tr>
<td>---------------------</td>
<td>----</td>
</tr>
</tbody>
</table>

**CHAPTER | PAGE**
--- | ---
Model IV | 92  
Assumptions | 92  
A Low Cost Tolerant i | 97  
Solution Concept and Equilibria Outcomes | 98  
Hypotheses | 99  
A High Cost Tolerant i | 101  
Hypotheses | 102  
Player j | 103  
Hypotheses | 104  
Summary of Model IV | 105  
Conclusion | 106  

**IV. ESCALATION IN EXTENDED DETERRENCE CRISSES | 110**

Theoretical Summation | 110  
The Extended Deterrence Context | 114  
Operational Hypothesis: Costs | 117  
The Value of the Issue at Stake | 118  
Research Design | 121  
Data Analysis | 124  
Conclusion | 129  

**V. ESCALATION IN MILITARIZED INTERSTATE DISPUTES | 132**

Theoretical Summation | 133  
Operational Hypothesis: The Dependent Variable | 138  
The Independent Variable | 141  
Research Design | 142
Data Analysis ................................................. 144
Conclusion .................................................... 149

CHAPTER PAGE

VI. CONCLUSION .......................................... 152

Empirical Evidence ......................................... 156

Limitations of the Current Study and Future Research .......... 157

REFERENCES ............................................... 159

APPENDIX 1 ................................................. 172
APPENDIX 2 .................................................. 173, 174
APPENDIX 3 .................................................. 175
APPENDIX 4 .................................................. 176
APPENDIX 5 .................................................. 177, 178
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1. Logit Estimates for Deterrence Success (Attacker Refrains from Escalating to Higher Levels)</td>
<td>125</td>
</tr>
<tr>
<td>5.1. Logit Estimates for Weaker State's Escalation Behavior</td>
<td>145</td>
</tr>
<tr>
<td>5.2. The Level of Escalation Achieved by the Weaker State in a Militarized Interstate Disputes</td>
<td>148</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Complete Information Game</td>
<td>56</td>
</tr>
<tr>
<td>2. A Low and High Cost Tolerant i</td>
<td>68</td>
</tr>
<tr>
<td>3. One-sided Incomplete Information Game</td>
<td>79</td>
</tr>
<tr>
<td>4. Two-sided Incomplete Information: A Low Cost Tolerant (LCT) i</td>
<td>93, 94</td>
</tr>
<tr>
<td>vs. a Low or High Cost Tolerant (HCT) j and A HCT i</td>
<td></td>
</tr>
<tr>
<td>vs. a LCT or HCT j</td>
<td></td>
</tr>
</tbody>
</table>
Chapter I
Introduction

Escalation is a concept intimately connected with the study of conflict resolution. Across a wide variety of scholarly disciplines, conflict studies are replete with references to escalation. For evolutionary biologists, an escalated fight between animals could mean the eventual extinction of the losing animal's species (Maynard-Smith, 1972; 1974; Parker, 1974; Hammerstein and Parker, 1982). For international relations theorists, escalation between nuclear states could result in the mutual self-destruction of nation states (Brodie, 1959; Schelling, 1960; Snyder, 1961; Kahn, 1965) and the eventual extinction of human life. However, the outcome of escalation need not be so dramatic nor does an outcome have to materialize before the disruptive and damaging effects of escalation can be evaluated. Presidents have stepped down (Johnson in 1968), cities have been destroyed (London and Dresden in World War II) and in 1993, the combatants in Yugoslavia's civil war have contributed to the systematic genocide of an entire race of people. Whatever the particular behavioral manifestation, escalation can be a frightening concept to both the actors directly involved and for the policy makers who are responsible for making these escalation decisions.

The use of the term escalation engenders almost immediate comprehension for everyone. The general impression that most people are left with when they hear the term escalation is that for some reason the conflict worsened. Perhaps punches were thrown, riots broke out or a nation's military was used to punish a rival.

However, the centrality of escalation as a focus for theoretical and empirical inquiry varies a great deal. In some cases, escalation is a convenient stepping stone to investigate some other dimension of a conflict situation. For others, escalation is associated with a particular context, e.g., arms races (Richardson, 1960; Singer, 1962;
Caspar, 1967; Intriligator and Brito, 1984; 1989). The theoretical eclecticism that characterizes the study of escalation suggests that few explanations are satisfactory to the majority of scholars working within this area. In spite of the number of efforts that address some aspect of the problem of escalation, our understanding of the dynamics of the escalation process is quite primitive. One of the weakest links in our understanding of escalation has been the failure to produce a set of falsifiable hypotheses that once tested, can be used as evidence to assess the validity of the theory that supported those derivations.

The result is a great deal of theoretical speculation and a proliferation of different models but few attempts to falsify some of these explanations. A thorough understanding of any event must be based on a systematic theory whose derivations are supported empirically. The purpose of this thesis is an attempt to explain why nations escalate and perhaps more importantly, why they stop escalating. Toward that end, I present several models of the escalation process, each model building on the previous one, which are used to derive several empirically testable hypotheses that address some aspect of escalation. Some of these hypotheses are tested empirically in different international contexts such as extended deterrence and crises. While most would agree that escalation is a general phenomenon that is relevant across most types of conflict situations, few efforts have been devoted to developing a theory of escalation at a sufficiently general level to viably transport the theoretical linkages specified between the variables to different contexts.

My aim in this study is more limited than developing a comprehensive theory that addresses all aspects of the escalation dynamic. Rather, the goal is to build a model that is used to answer some important questions but that can also serve as a solid theoretical foundation for investigating more complex problems in future research. Any explanation of why states escalate is only as good as the theory which supports it. Therefore, it is necessary to spend a great deal of time building the theoretical framework from the ground
up so that the models are developed in a logically consistent and coherent fashion.

A major problem for the student of escalation is in untangling what is known about escalation stems from the variety of definitions, conceptualizations and independent variables used to explain why states escalate. In this chapter, I present my conceptualization of the problem, identifying what is being studied and the method of analysis used to address various aspects of the escalation dynamic.

What are We Studying?

One of the more confusing aspects in sifting through studies of escalation is that the concept can have multiple references and scholars are not always careful to distinguish the manner in which escalation is used in the analysis. Escalation can be the dependent variable in the analysis, the variable which is to be explained. The primary aim of this thesis is to explain escalation by identifying the conditions under which states are expected to escalate in a conflict. And it is the presumption of this study that once we identify the factors that affect states' decisions to escalate, we will also be in a position to determine the conditions when states refuse to escalate altogether or when actors escalate for some period of time and then exit the conflict (Osgood, 1962; Bonoma, 1975). Before discussing the other ways in which escalation is used in different analyses, it is necessary to define the concept to know exactly what we are explaining. Whenever a term such as escalation is used frequently but is not the exclusive focus of study, providing a clear definition of the term seems almost unnecessary. We all know what escalation refers to generally but rigorous analysis demands more than theories based on an intuitive understanding of its major concepts.

Escalation is defined as the imposition of costs on both the sender and recipient of an action over time (Cross, 1969; Pillar, 1983; Lalman, 1990). This definition of escalation is consistent with more common treatments which view escalation as an "increase in the magnitude of hostility or the application of sanctions over time" (Bonoma,
1975; Gochman and Leng, 1983). Note that escalation is defined at the monadic level of analysis. Even though a conflict involves at least two participants, the behavior of a single actor within that dispute is described as escalation irrespective of the behavior of the other participant. One of the results of defining escalation this broadly is that many behaviors which would be excluded from other analyses are now fair game for investigation. It then becomes possible to derive hypotheses which identify how a single state will behave in a conflict.

Many definitions of escalation, particularly in empirical research, implicitly or explicitly require that escalation be defined at the dyadic level of analysis (Leng and Wheeler, 1979; Gochman and Maoz, 1984; Leng, 1993). The identification of an escalation action is determined by the previous level of escalation (hostility) set by the opponent. If the hostility level does not match or exceed the pre-established threshold set by the opposing player, the behavior is not considered escalatory. This dyadic definition of escalation is clearly inappropriate to use if we want to explain escalation, however, it is useful if the research question involves predicting to some event other than the escalation behavior itself, e.g., war, (Siverson and Miller, 1992). In addition, the dyadic definition fails to distinguish the act of escalation from a description of how states escalate. The definition of escalation used in this study does not contain any restrictions on the way that states escalate. The actors may impose fewer, greater or an equal amount of costs the next time they decide to act. The level to which states escalate, e.g., matching or exceeding the previous level set by the opponent, is in my framework, more appropriately viewed as one kind of escalation move and an expectation about state behavior that should be derived from the theory and not included as part of the definition of escalation. This is one of the reasons why the definition of escalation used here does not contain the stipulation that the level of costs imposed by an actor must be increasing on each subsequent move taken by that actor. How states escalate is a separate research question from why states escalate.
While it is fairly easy to make a clean analytical distinction between the two decisions, in practice it is more likely that these decisions are made simultaneously. It is important to analyze these decisions separately before any attempt is made to join theoretically these two dimensions of escalation.

This raises the question as to what kind of behavior qualifies as non-escalation. Escalation can be conceived as a continuum of cost imposition ranging from zero to some positive number. Presumably the upper limit is established by the resources that the participants can bring to bear in the conflict. Non-escalation is the imposition of zero costs or in effect, to cease participating in the conflict. The decision to stop imposing costs implies that escalation is within the control of the actors. However, this assumption is not accepted by those who view escalation as a process that results "when people do not stop to think" (Richardson, 1960; Boulding, 1962). While these assumptions may appear marginal on the surface, they are in fact the product of two very different conceptualizations of the escalation process. And it is these different conceptualizations that make up the other referents to escalation mentioned earlier.

A well articulated conceptualization of escalation is absolutely critical to any theory developed to explain some event. The conceptualization guides the selection of the appropriate assumptions used to build the model and assists in the identification of the key explanatory variables. The interaction of all of these factors combine to produce very different explanations, some of which are more or less useful in explaining escalation.

Escalation is often used as a description of some conflict process in which the actors engage in a set of interactions that move a conflict from some initial starting point to some other outcome (Richardson, 1960; Boulding, 1962; Wright, 1965; Pruitt, 1969; Grefffinus, 1990; Grefffinus and Gill, 1992). These set of interactions are captured by a reaction function in which one actor's behavior is to a large degree a function of the behavior of the other actor(s) in that system. One of the major research questions guiding
this work is identifying the conditions under which changes to the system's equilibrium point (escalation) becomes dysfunctional to the maintenance of that system. In effect, escalation spins out of control into a conflict spiral from which it becomes very difficult, if not impossible, for the actors to extricate themselves. This conceptualization has been a particularly powerful tool in providing insight on the dynamics of arms races (Isard and Anderton, 1985; Anderton, 1985) and the conditions under which arms acquisitions might lead to war (Intriligator and Brito, 1984; Morrow, 1987). However, the generalizability of this conceptual approach to other areas of escalation might be limited by the fact that escalation is defined at the dyadic level and another assumption which often accompanies these studies: an automatic action-reaction escalation dynamic.

A more useful conceptualization of escalation is to view escalation as one part of a larger context of state interaction rather than identifying that context in terms of escalation. Throughout the thesis, escalation is viewed as a sub-process of the larger process of interstate bargaining (Schelling, 1960; 1966; Young, 1968; Snyder, 1972; Pillar, 1983; Patchen, 1988). This perspective uses escalation in a third way, as a tactic that states manipulate to try and improve their negotiating position (Schelling, 1966). As such, escalation is one of the values that the independent variable called 'bargaining tactics' can assume. The aim of many bargaining models that analyze the coercive use of force attempt to explain how escalation affects the eventual outcome received by both parties (Ellsberg, 1961; Cross, 1969; Wittman, 1979; Pillar, 1983). While my aim is more limited than specifying a theory which incorporates the dynamics of bargaining directly, the bargaining conceptualization identifies the key motivation underlying the use of coercion: the attempt to convince the other side to submit to one's demands.

The importance of explicitly identifying the reason why states escalate should not be underestimated. Many treatments of escalation, even analyses that purport to explain escalation behavior, assume that escalation on behalf of one party automatically produces
counter-escalation on behalf of the other (Kahn, 1965; Ikle, 1971; O'Neill, 1986; Zagare, 1992). This explanation returns us to a reaction-function type conceptualization of escalation which spirals upward when players do not stop to think. While escalation can serve as both a tactic manipulated by the actors themselves and a variable to be explained, we must avoid falling into the trap of explaining escalation with escalation or a "bad begets bad" notion (Leng, 1980; Keohane, 1986). Without penetrating below the surface to specify the effects that escalation has on players' evaluation of their situation, the escalation begets escalation explanation is a tautology which sheds little insight on why states pursue escalation. The bargaining conceptualization anchors the analysis of state interaction. Without some framework guiding the study, the explanation for escalation is free to wander all over the map.

In sum, one of the major impediments to gaining a fuller understanding of the escalation process is the lack of well integrated, empirically supported theories that explain the process and not the outcome of an escalation process. The source of these problems has been inadequate conceptualizations of the event being studied and the failure to specify the logical linkages between the key explanatory variables affecting state behavior. One of the more serious problems is the lack of attention paid to the way in which escalation is used in the analysis. Escalation can be used to describe a process, it can be an independent variable manipulated by an individual actor and escalation can serve as an independent variable. Any or all of these uses are legitimate. However, it is incumbent upon the researcher to clarify its meaning and to avoid slipping back and forth between different usages of the term. A bargaining conceptualization has the advantage of placing escalation within a general framework and avoids wedding the term to any one particular context.
Method of Analysis

The aim of this thesis is twofold. First, I develop a general theory of the escalation process that is used to predict both the conditions under which states are expected to escalate in a conflict and the level to which states will escalate once the decision to escalate has been made. The models developed generate several hypotheses that specify the interactions among different kinds of variables and produce different conclusions regarding the conditions that are likely to cause escalation.

The value of any theory depends on how well the hypotheses are supported empirically. A great deal of the empirical evidence that exists on the process of escalation has been derived from the use of case studies (George et al., 1971; Ikle, 1971; Smoke, 1977; Snyder and Diesing, 1977). This method is very useful in uncovering some of the rich detail and insights often missed in more aggregate analyses of conflict. However, these studies many not be generalizable and so it becomes important to test the hypotheses against a larger number of cases. Thus, a second aim of the thesis is to test empirically some the hypotheses derived from the model(s) in different international contexts, e.g., extended deterrence crises (Huth, 1988) and militarized interstate disputes (Gochman and Maoz, 1984). The theory is intended to be general so that the variables posited to effect the escalation dynamic are relevant in any conflict situation. Thus, one of the aims of future research is to test a different set of hypotheses in other contexts such as arms races and war.

The theory developed is based on game theory. Game-theoretic models can be either static, e.g., 2x2 matrix games, or dynamic, e.g., sequential play with multiple opportunities to move. Because the research in this thesis involves the dynamics of escalation, the latter approach is used. Increasingly, the expected utility and game-theoretic approach to understanding escalation in a bargaining environment has been able to produce important insights on the factors that affect the escalation dynamic (Shubik,
1968; Bonoma, 1975; Bueno de Mesquita, 1984; Morrow, 1985; Brams and Kilgour, 1987; Maoz, 1990; Lalman, 1990; Bueno de Mesquita and Lalman, 1992; Zagare, 1992). There are at least two important issues raised in these analyses which are developed fully in the model(s) in this thesis. First, the costs for doing business in an escalation game are not only generated endogenously but are subject to some upper threshold defined by the players themselves (Brockner and Rubin, 1985; Maoz, 1985). However, what is often neglected in work that does not adopt an explicit bargaining view of escalation is that the players are in conflict over some issue. In other words, players are not escalating because the opponent has escalated; they escalate because this is one of the ways to achieve some satisfaction on the issue at stake. And it is the relative evaluation of that issue which determines the level of escalation or defines the upper limit on the costs a player is willing to suffer to play the escalation game. The relative cost-benefit tradeoff that a player is willing to make provides an important cap on the level of escalation to be observed. It is not uncommon to find escalation studies that assume costs, benefits and the like are all variable without also identifying when they become variable or why escalation should ever stop (Smoke, 1977; O'Neill, 1986).

The model of escalation developed in this thesis is built in four stages. Each model builds on the previous version and is a generalization of the models which precede it. Any mathematical model must be formulated, at least initially, at a fairly abstract level. Even though the model(s) represent a stylized and simplified escalation game, this doesn't necessarily mean that the model is oversimplified and produces only trivial results. The loss of detail is compensated for by the increase in clarity and rigor that simplification provides. Once the basic framework is established, complicating features can be incorporated into the model.

The primary aim in building the model in different stages is that each model produces a set of expectations regarding the escalation behavior of states. Following the
presentation of each of these models, I offer a theoretical discussion of the model's derivations and identify different factors found in the empirical setting which are used to represent the parameters of the hypotheses operationally. One important aspect of the discussion is demonstrating the ability of that particular model to address the research questions of interest. The conceptual weaknesses identified for that model are used as a basis for extending the model into another version. Because each model is a generalization of the former, these models explain everything that the previous models were able to address but also explains some other dimension of the escalation question not logically permissible in previous versions.

Escalation has been analyzed with a variety of different paradigms most of which are based on different assumptions and different explanatory variables depending on the particular scholarly discipline in question. While each of these respective fields has tapped some important aspect of escalation in general, we have yet to exploit this resource for purposes of building a theory of escalation that resolves some of theoretical and empirical inconsistencies that currently impede our progress in understanding the subject. The goal is to draw from several of these research traditions to produce a theoretical framework that can be used to build upon in future research and guide our current understanding of why nations escalate.

Organization of the Thesis

The following chapter delves more deeply into the literature on escalation. In particular, the strengths and weaknesses of the various approaches to escalation are discussed with a greater emphasis on the theoretical arguments advanced to explain escalation. Much of the empirical work on escalation has attempted to predict to an outcome associated with the escalation process and this research is accessed for purposes of identifying any empirical regularities associated with particular patterns of escalation.

The third chapter is the heart of the thesis. This chapter fully explicates the
conceptualization of escalation, defines the key terms and identifies the explanatory variables. As stated previously, the models are developed in four different stages and at each stage of the analysis, I will spend some time justifying the assumptions used. To facilitate the discussion of the model's conceptual strengths and weaknesses in addressing the major research questions of interest, I use the example of President Lyndon Johnson's decision to escalate American involvement in Vietnam in the early 1960s. In a sense, these models are used to post-dict Johnson's decisions which are then compared against the historical record. The Vietnam example is not meant to be a full-fledged case study of escalation but is used mainly in an effort to clarify the discussion and provide some intuition behind each of the models. This discussion also includes a theoretical justification for the modifications that are necessary to equip the next model with the ability to address certain questions not accessed by the previous model.

Chapter four presents the results of an empirical test of one of the derivations from the third model developed in chapter three. The hypothesis is tested against cases drawn from extended deterrence crises (Huth and Russett, 1984; Huth, 1988). Because deterrence and escalation are opposite sides of the same coin (Brodie, 1959; Lichbach, 1987), this set of cases provides an extremely rich context in which to test a hypothesis identifying when states ignore the opponent's threat and back down versus the conditions that lead a nation to press forward and escalate.

Chapter five explores the results of a different hypothesis tested in the context of militarized interstate disputes (Gochman and Maoz, 1984). This hypothesis is derived from the fourth model which is formulated to tap a different dimension of escalation. The hypotheses derived from model I-III all identify the conditions under which states are expected to escalate in a conflict. The fourth model takes these conclusions one step further and asks: given that the decision to escalate has been made, what level of escalation will that actor choose as his opening gambit? Like immediate extended deterrence
situations, militarized disputes are high conflict contexts in which the probability that one or both actors will use force has increased dramatically (Gochman and Maoz, 1984; Siverson and Miller, 1992).

The final chapter assesses the overall accomplishments of the thesis. Admittedly, the task of building a theory from the ground up necessitates that the scope of important questions that can be addressed thoroughly in any one research project be limited to a select few. However, the questions that I have identified are among the most important in that the answers to these questions must be addressed satisfactorily before we can press on with future research. The empirical evidence presented in this thesis will determine whether the theory of escalation developed is a useful way to understand the escalation problems of today and perhaps will be able to assist policy makers in understanding the dynamics of the conflicts of tomorrow.
Chapter II
Literature Review

The study of the social sciences is in large measure the study of human conflict and the ways in which humans go about resolving those conflicts. Whenever conflict occurs, there is some probability that the protagonists will adopt a strategy of escalation before the conflict is resolved. Escalation has been the subject of many theoretical inquiries in disciplines as diverse as psychology and evolutionary biology. The problem facing both the practitioner and the student of escalation is in untangling what is known about escalation both theoretically and empirically especially when inconsistencies and incompatibilities abound both across and within paradigms.

The common link between many theories that analyze escalation either directly or indirectly is the identification of those factors which cause actors to resort to the use higher forms of coercive power before the conflict is eventually resolved. Beyond that common denominator; however, theories of escalation differ widely in terms of the basic conceptualizations and definitions of escalation, the assumptions upon which the models are based and the independent variables used to explain the process or the outcome of escalation. Despite the theoretical eclecticism that characterizes the study of escalation, the existing research on escalation falls within two broad categories.

The bulk of the research on interstate escalation focuses on explaining and predicting the outcomes associated with an implicit escalation process. The outcome that absorbs most of the scholarly attention in international relations is war. Almost by definition, most wars are viewed as the result of an escalation process in which state conflict begins at lower levels of coercion and progresses to higher forms of violence. The explanatory and predictive power of these studies is derived from the association between the values that particular variables assume at the outset of a conflict and the affect these variables have on the likelihood that a conflict ends in war. The assumption driving
this approach to escalation is that the underlying structural features of the conflict have a
decisive impact on the whether the actors resort to war to resolve their conflict. The
causes of war are found either in the long term developments in the relations between
states (Hobson, 1904; Richardson, 1960; Choucri and North, 1975 ) or in the more
immediate, short term attributes of nations involved in a crisis situation (Azar, 1972;
Cusack and Eberwein, 1982; Gochman and Leng, 1983; Gochman and Maoz, 1984 ).

Another set of escalation research concentrates on explaining the process of
interstate escalation. The dependent variable of these studies is escalation and the goal is
to determine theoretically how and why states respond to one another in some fashion.
War, the critical dependent variable in the previous approach, is viewed from a process
perspective as simply another level of escalation in a series of past and future interactions.
There are two main approaches used to study the escalation process. The systems
perspective characterizes the relationship between actors as an interaction system in which
the escalation of that relationship is sustained and fueled by a psychological dynamic. The
bargaining approach conceptualizes escalation as a subprocess of the larger context of
interstate bargaining. This approach is based on the assumption that the behavior of the
actors is of vital importance to the eventual outcome of the conflict (Schelling, 1960; 1966;
Kahn, 1965; Young, 1968; Bonoma, 1975; Powell, 1987; 1988; 1990; Lalman, 1990;
Zagare, 1992). Parties to a conflict use escalation tactics purposively to convince the
adversary to give in to the former's demands. The process is characterized by various
combinations of decision making throughout the duration of the conflict.

The use of these different approaches in understanding escalation carry with them
certain advantages and disadvantages that must be taken into consideration when
constructing a theory of escalation. The structural and systems approach have identified a
set of key explanatory variables that tend to be correlated consistently with higher forms of
violence. The major problem with these studies is that while there exists a broad
consensus concerning the key variables, the affects of these variables on war, and especially on escalation, is open to theoretical debate. In addition, the use of psychological variables to explain the escalation dynamic tends to produce hypotheses which are difficult to falsify empirically.

On the other hand, process theorists pay particular attention to the context of escalation and the variables that affect the actors' escalation decisions. Even though the majority of this work conceptualizes escalation as part of a broader framework of interstate bargaining, the specific definitions of escalation, the variables that affect the process, and the assumptions made about the actors often result in the use of models that capture the extremes of state escalation behavior.

Outcomes of the Escalation Process

A vast majority of the research on escalation in international relations can trace its theoretical origins to realism whose primary aim is to explain the onset of war or to predict to some other outcome of which war is one of the values that the dependent variable can assume (Morgenthau, 1950; Waltz, 1979; Leng and Singer, 1988). The interest in escalation coincided specifically with the recognition that intense conflicts that occur between states, such as crises or disputes, precede most if not all wars (Hermann, 1972; Snyder and Diesing, 1977; Lebow, 1981 Siverson and Miller, 1992). The distinguishing feature of a crisis is a decision maker's perception that the conflict represents a grave threat to national values and that there exists a relatively short time frame within which to respond to the crisis (Hermann, 1972; Wilkenfeld and Brecher, 1989). The most important feature of the crisis is the recognition on behalf of both parties that one or both sides may resort to the use of force, thereby increasing the probability that war will occur between the two sides (Snyder and Diesing, 1977). The crisis, however, may be resolved without either of the participants ever firing a shot. Most of the empirical work on the escalation to war relies on a subset of crisis cases, the militarized interstate dispute, in
which states at a minimum threaten to use force (Siverson and Miller, 1992).

Many, though not all of the studies addressing the outcomes of an escalation process, conceptualize escalation as an attribute of the conflict. The escalation behavior of an actor is pegged to and defined in terms of the adversary’s behavior. If the behavior of an actor does not match or exceed the level of violence previously set by the opponent, the behavior is not considered escalatory. The reason for linking individual behavior with the level of violence in the conflict is clear. These studies are formulated at the dyadic level of analysis and the research question concerns the impact that certain variables have on the likelihood that the conflict will end in war. Less emphasis is placed on identifying the conditions under which an individual state chooses to escalate.

Within the context of a dispute, the variable posited to have one of the strongest affects on the escalation to war is a nation’s relative power, and in particular, its military capabilities. A nation's relative military capabilities are an important component in the decision to wage war because capabilities are a weighty determinate of the chances of winning a war. Military capabilities are used to overrun the enemy’s military forces, to destroy property and extinguish lives. In short, capabilities are used to make the war effort so painful to the other side that they are forced to give in and/or go to the negotiating table (Pillar, 1983). All other things being equal, the stronger a state’s military, the greater the likelihood that that actor can wage a successful war (Morgan, 1990). The causal importance of the relative capability distribution in determining the onset of war; however, is an old debate in the realist literature and focuses on the war-proneness (escalation proneness) of certain types of capability distributions (Siverson and Sullivan, 1983). The link between relative capabilities and the escalation to war is drawn theoretically through the mediating affect of decision makers' uncertainty as to which of the two sides will win a war. There exists empirical evidence to support the proposition that an equal power distribution promotes peace by inducing leaders to behave cautiously.
when the likelihood of winning or losing is perceived to be roughly even (Morgenthau, 1950; Claude Jr., 1961; Champion and Stoll, 1978). There is also evidence to support the contrary proposition that it is precisely this ambiguity over who will win that makes war more likely (Garnham, 1976a; 1976b; Weede, 1976; Organski and Kugler, 1980).

Recent efforts that have analyzed the power thesis claim that power alone is insufficient to comprehend the conditions under which a state will choose to go to war (Bueno de Mesquita, 1981; Wagner, 1984). The missing component that determines the condition under which an actor will go to war is the value that each side gains from escalation and war. This variable, as well as the other alternatives available to the players in the conflict, must be evaluated before any determination on the likelihood of war can be assessed theoretically. What is usually missing from theoretical analyses of this kind is the evaluation of the issue at stake to the players.

While it may true that the relative capability distribution is one of the components that affects a state's willingness to go to war to resolve a conflict, the theoretical link between the capability distribution and the willingness to escalate in a conflict short of war is less well understood. It is reasonable to expect that a state can be certain at the outset of a dispute that it will lose a war and yet still choose to escalate the conflict on the presumption that the adversary might back down. Similarly, the militarily, stronger side in the conflict can be confident of victory and yet fail to escalate (Jervis, 1988). The theoretical gap between the capability to escalate and the willingness to escalate is found in the absence of any discussion as to why a state might be motivated to escalate in a conflict. Thus, to fill in the missing conceptual link between capabilities and escalation, we must first determine what motivates states to escalate and second, how do capabilities affect that motivation?

Deterrence theorists have paid particular attention to the interconnections between power, threats and the value that brings the sides into conflict. Deterrence is a form of
coercive power that attempts to "persuade a potential aggressor that the costs of an action will outweigh the benefits from pressing ahead" (Brodie, 1959; Synder, 1961; George and Smoke, 1974; Morgan, 1977). The issue in a deterrence situation is the price of victory because the costliness of escalation is not an absolute evaluation but one made relative to the value of the issue at stake. Even if a potential aggressor is confident that he is militarily stronger than the defender of the status quo (hence more likely, but not certain, to win an armed conflict), the aggressor must also consider whether the costs that the defender can impose during the attempt to achieve some value are greater than any benefit to be gained from conflict. If so, the aggressor will be less likely to press ahead.

States are motivated to escalate to achieve some value in the dispute or to prevent the other side from realizing those gains. The means to achieve this goal are through the threat of cost imposition which is represented by a player's military capabilities. The overall aim of a deterrent threat is to prevent an armed attack, however, this also includes deterring the opponent from taking escalatory steps that could eventually result in war (Brodie, 1959; Kissinger, 1969). As such, the concepts of deterrence and escalation are tightly interconnected. However, the focus of the vast majority of deterrence studies is on identifying the conditions under which deterrence will be successful. Consequently, even though deterrence provides a rich context within which to study the factors affecting the escalation process, many deterrence models are structured in such a way that they are of minimal use in explaining the process by which deterrence is established.

One of the major drawbacks of formal models of strategic deterrence is that they are often based on the assumption of complete information. This means that players know the outcome of the game with certainty (Brams, 1985; Brams and Kilgour, 1985a; 1985b; 1986c; 1988a; 1988c; Zagare, 1985; 1987; 1992). This assumption is useful in deriving a set of expectations about what should occur under an ideal set of circumstances. However, this assumption is so restrictive that when the winner and loser is determined
prior to the game, escalation becomes unnecessary since the loser cedes immediately
because the winner's threat to escalate is perfectly credible. Complete information
transforms the deterrence/escalation game into a mental exercise for the players in which
no observable escalation behavior is ever expected. Escalation only becomes possible if at
least one of the players does not have complete information (Wagner, 1982; Nalebuff,
player can be viewed as one means to gather information over which actor will ultimately
prevail in the contest.

Another limitation of many of these models for studying escalation is that the
deterrence game is often conceptualized as a game of chicken in which mutual escalation
leads to disaster, e.g., nuclear war (Snyder and Diesing, 1977; Jervis, 1979). A 2x2
game matrix conceptualization of deterrence obscures the fact that deterrence can be
established following an escalation move, e.g., a missile launch, because these models are
based on the implicit assumption that escalation is an automatic action-reaction dynamic
that always ends in disaster. However, deterrence can still prevail even if a state threatens
less than total costs (Langlois, 1989). Once we relax the automatic escalation assumption,
there is no reason to expect that the escalation game is structured like a game of chicken.
This shifts the emphasis to identifying the conditions under which escalation is too costly
to sustain and the conditions when players continue to escalate. In other words, even
though both players may fear the costs of escalation, some may fear them a little less and
can derive a bargaining advantage from being in that position (Wagner, 1982).

Deterrence is not only the capacity to make the costs of the adversary's action
outweigh the benefits. Deterrence is fundamentally a question of the intentions of the
other side (Schelling, 1960; Jervis, 1979; Singer, 1961). The aggressor must also
determine the probability that the defender will carry through with the punishment. Since
the aggressor does not know for certain whether the defender's threat to retaliate is
genuine, the aggressor has to pick up cues from both the environment and the defender's actual behavior to determine the likelihood being punished. And if the defender is serious about the threat of retaliation, it is in his interest to communicate this intention as clearly as possible (George and Smoke, 1974). This increases the likelihood that the aggressor will back down and the defender can avoid paying the costs of retaliation.

Recent attempts to model the deterrence problem formally begin with the assumption that deterrence does not involve a one shot calculation on the part of either of the players (Wagner, 1983; Langlois, 1989; Kilgour and Zagare, 1991; Nalebuff, 1991). The ingredients for successful deterrence may be present at the outset of a crisis; however, whether the ultimate outcome of deterrence is success or failure is established during the process of resolving the conflict (Nalebuff, 1991). In other words, the way in which states manage the crisis may be equally if not more important than the set of initial conditions in determining the eventual outcome of a crisis (George et al., 1971; Lauren, 1979; Lebow, 1981).

The game that the players are engaged in involves a demonstration that the issue at stake is important enough to bear the burden of imposing costs on himself to convince the adversary to back down. The importance or salience of the issue at stake has been tied to the concept of a nation's resolve. In general resolve refers to the relative unwillingness to give in to the adversary (stand firm in the face of demands). It has even been suggested that all crises are competitions of the players' resolve (Powell, 1987; Morrow, 1989).

Empirical evidence, particularly in crisis bargaining studies, suggests that the greater value states place on the issue, the more likely that states escalate and the crisis ends in war (Barringer, 1972; Leng, 1980; Weede, 1983; Huth and Russett, 1988). Unlike deterrence theory whose primary focus is on negative sanctions, crisis bargaining research examines the impact of various mixes of coercive and accommodative bargaining strategies on the likelihood the dispute ends in war. These studies provide important
empirical evidence that the behavior of the participants in a crisis may be equally if not more important than the underlying attributes that define the crisis context for determining the eventual outcome to the crisis (Azar, 1972; Hermann, 1972; Brecher and Wilkenfeld, 1989; Wilkenfeld, 1991). One of the theoretical difficulties of these studies, however, is that they "use a more inductive approach to ground the hypotheses" (Leng and Walker, 1982) making it difficult to determine the causal direction of a bargaining strategy and the probability of escalation. One of the most war-prone bargaining strategies is a bullying strategy in which players use escalating, negative inducements exclusively to try and get the opponent to back down. However, in the absence of a solid theory to interpret the empirical results, it is not clear whether a player's high evaluation for war causes them to choose a bullying strategy or whether a bullying strategy causes a war that may or may not be intended by either party to the dispute. Much of the theory guiding the affects of coercion on the opposing player and an increase in the probability of war is based either on psychological reciprocity theories in which "like is exchanged for like" (Jackson, 1973; Leng 1983; Greffenius, 1990; Greffenius and Gill, 1991) or on the idea that the costs of escalation create their own momentum (Holsti et al., 1968; Hoole and Zinnes, 1976; Kaplowitz, 1984; Leng, 1988). However, since some bargaining strategies are defined by different mixes of escalation and accommodation, it must be the case that there are conditions under which escalation is matched by escalation and conditions when escalation produces concession.

Many of these studies treat the selection of a bargaining strategy and the evaluation of an outcome as independent. This is a rather dubious assumption since a player's evaluation over all outcomes is inextricably linked with the means to achieve that outcome. But why states adopt certain strategies over others is not addressed theoretically in this research since the strategy selection of the players is determined empirically and then correlated with the outcome. The conditions under which various moves (a series of
moves determines the overall strategy) are selected remains unknown. So while we know that the process matters, we do not know how or why without an investigation into the conditions under which certain kinds of interaction patterns emerge.

One type of interaction pattern, the arms race, has been the subject of a great deal of theoretical and empirical work (Huntington, 1958; Richardson, 1960; Caspary; 1967; Smith, 1980; Intriligator and Brito, 1984; 1989; Morrow, 1989; Wallace, 1979; 1982; Diehl, 1983; 1985). Arms races presumably embody the classic case of an escalation situation because states are amassing huge, costly arsenals as a direct reaction to the opponent's behavior. Richardson (1960) is credited with beginning the development of the scientific, mathematical study of arms races and indeed has spawned a whole cottage industry of research devoted to this subject area. Richardson, however, was more concerned with describing one of several theoretical processes which could precede the outbreak of a war (Intriligator and Brito, 1984). An arms race is one of the behavioral manifestations of states who have an extremely antagonistic and hostile relationship with one another. The mutual feelings of insecurity and hostility that are brought on or exacerbated by the arms race can lead to the belief that war between the two sides is inevitable. The only question that remains is the timing of the conflict. Thus one of the possible linkages between an arms race and war is that one of the actors may have an incentive to launch a pre-emptive strike against its adversary to gain the advantage (Morrow, 1989).

There are two central research questions that guide the arms race literature. On the one hand, there are attempts to establish the empirical existence of an arms race. For those arguing that the determinates of a nation's military buildup are found in the domestic polity itself (Ostrom, 1977a; Cusack and Ward, 1981) the entire notion of an escalating arms race in which states are competing with one another for military advantages is untenable theoretically. The implication of course is that escalation too becomes meaningless
because the external-political linkage that supposedly drives arms acquisitions is severed theoretically and empirically. This does not necessarily imply that a rival nation's military capabilities are unimportant. In fact most states or leaders justify their military acquisitions in strategic terms (Kurth, 1981). The point is that arms races are not escalatory per se but simply a building process sustained by the internal, political needs of the state's leaders.

One of the most serious theoretical gaps of this literature is the absence of any discussion of the theoretical conditions under which an arms race should begin or wind down. The stopping point for an arms race is usually imposed by assuming that the arms race ends in war. But even that theoretical linkage has been subject to intense debate (Wallace, 1979; Altfeld, 1983; Morrow, 1989). The difficulty is that we have one interaction process, the arms race, and the outcome of war or no war. In effect, we have a theoretical black box that fails to provide the linkage between one process and another which entails resolving the conflict.

The literature that examines the affects of players' attributes and bargaining behavior has identified key variables associated with the escalation to higher forms of violence, e.g., war. In particular, military capabilities are viewed as vitally important given that they represent a states' most overt expression of influence in the international system (Morgenthau, 1950). There is general agreement that the military is used to punish the opponent by denying the adversary something of value (Snyder, 1961). Most of the literature on outcomes has failed to establish any strong theoretical links between the relative capability distribution and the likelihood of war. Fewer theoretical links are advanced to explain a player's willingness to escalate in a conflict short of war. In fact some argue that the connection between capabilities and the willingness to escalate is rather loose, even vague. The stronger component in a states' decision to escalate is the evaluation of the issue at stake (Jervis, 1988; Snyder and Diesing, 1977). A state may be
confident of winning a war and yet fail to escalate if the loser can escalate and raise the
price of victory to an unacceptable level. Moreover, the importance of the relative costs
and gains must be integrated into a theory that expressly focuses on the process of
escalation. Moreover, the models must be based on assumptions that permit states to
escalate in the conflict. The emphasis of the literature review shifts away from outcomes
to two approaches used to study the escalation process: the psychological dynamic
approach and escalation as a bargaining tactic.

System Level and Psychological Escalation Dynamics

One common approach is to conceptualize the escalation process as a description
of the relationship between a pair of actors. Each states' behavior is a response to both its
own internal needs and a reaction to the behavior of the other state. The joint reactions of
the players form a system of interaction which is sustained by these internal and external
factors. The system of interaction is analyzed in terms of its "stability properties" insofar
as a stable set of interactions leads to one outcome and an unstable dynamic leads to
another. The research emphasis of Richardson (1960), as well as other systems theorists
(Boulding, 1962; Pruitt, 1969), is on explaining the necessary conditions to create a stable
process which may precede an outbreak of warfare. Escalation implies more than just a
pair of states interacting, however. Escalation refers to an increase in the magnitude or
level of hostility that characterizes states' relations with one another (Grefenius, 1990;
Grefenius and Gill, 1992). Viewed in this light, arms races become one of the behavioral
manifestations of the underlying hostility of the actors; hostility which is generated and
exacerbated by the security dilemma (Jervis, 1979). The lack of an international sovereign
means that states are forced to devise ways of ensuring their own survival. One method to
deter a potential aggressors from attacking one's nation is to buildup its military capability.
This act produces the effect of increasing the security of the arming nation while
decreasing the security of other nations who now confront a stronger, more formidable
potential opponent. The insecure state responds by increasing his own military capabilities to alleviate its position of relative military inferiority. Thus the dilemma in the quest for security. A dilemma is one thing; the danger of not knowing the other's intentions is another. States can never be certain of the reasons why a nation is arming because they are operating in an uncertain environment. The effects of both sides arming is to reinforce the beliefs and suspicions that the other side is intending to use those military capabilities against them. To guard against this possibility (or to make it too costly for the other side to actually use its military for some political gain), states arm themselves. And so the cycle continues. This; however, is but one of the possible outcomes. Because states lack complete information, there is a strong possibility that states will misperceive their adversary's intentions. Rather then viewing the adversary's military buildup as a desire to increase its own security to maintain the peace, states misinterpret the adversary's behavior as a signal that it is going to use its arsenal for violent purposes. If one or both sides misperceive the adversary's intentions, this can lead to the belief that war between them is inevitable. The only question that remains is the timing of the conflict. Herein lies the link between an escalating hostility spiral and war. If states believe that war is inevitable, they may also reason that it is better to be the first to launch a pre-emptive strike against the adversary to gain at least a temporary advantage (Jervis, 1976; Levy, 1983).

Returning to the earlier discussion, all states face the security dilemma at all times. However, all states are not engaged in a mutually antagonistic relationship with one another. A critical question, which for the most part remains unanswered by either systems or spiral theorists, is the condition(s) necessary to instigate the escalation dynamic in the first place. Since the main concern of this research is "the stability properties" associated with the process leading to an outcome, the identification of certain variables or some kind of assumption is necessary to account for the behavior of states in a reaction
system. For the most part, escalation is described as an automatic action-reaction process (Kahn, 1965; Young, 1968) or a positive feedback loop in which escalation on one side produces counter-escalation on the other. The theoretical linkages drawn between action and reaction are different from the conceptual arguments advanced to explain horizontal escalation within a single conflict (Kahn, 1965; Epstein, 1983) or the diffusion of war across boundaries (Most and Starr, 1980; Most et al., 1990). Because the coercion begets coercion argument pertain to two distinct states, the argument is also different from addiction arguments that refer to the same nation using violence over time to achieve its political objectives (Richardson, 1960).

The theoretical linkages drawn between the stimulus (escalation) and the response (counter-escalation) are found in the psychological impact of the opponent's escalation behavior. It is assumed uniformly that escalation always engenders negative psychological reactions and that these reactions almost always produce counter-escalation on behalf of the other player. Most of the empirical evidence supporting these theories is based on lab experiments that represent escalation conceptually as a prisoner's dilemma. Escalation was defined as selecting the "defection" alternative and it was not uncommon for the experimenter to witness an unending series of defections following the first escalation move. When the participants were queried as to why they decided to defect, a majority responded that they were motivated by feelings of revenge, anger or betrayal (Tedeschi, 1970; Vincent and Schwerin, 1971 Bonoma, 1975; Pruitt and Rubin, 1986). Thus, the players become deadlocked in a coercion begets coercion or a "bad for bad" (Keohane, 1986) relationship from which it becomes extremely difficult, if not impossible, to extricate themselves.

Other experiments such as dollar auction games (Shubik, 1971; O'Neil, 1986; Leininger, 1989) are similar to the previous studies but differ in terms of the operationalization of escalation. Escalation in these studies refers to any bid level
exceeding the previous bid of the other player. Players begin with the same levels of resources that can be used to bid in increments on the value of a prize, a dollar. The winner, who is the highest bidder, and the loser pay the auctioneer the amount of their final bids. The results of these experiments show repeatedly that players escalate their bids well beyond the dollar value they are competing for in the game. In other words, the losses incurred well outweigh any benefits they receive for playing the game so it is "irrational" in retrospect for these players to play. The explanation for this empirical observation often refers to a player's "desire not to lose" (Ikle, 1971; Patchen, 1988) or the act of escalation increasing the values at stake." (Smoke, 1977; O'Neill, 1986). Presumably, "the desire not to lose" is equivalent to "the desire to win" is similar to an "increase in the values at stake" in both a technical and operational sense. Again, the theoretical linkage between the escalation behavior of the two actors is mediated through the nascent psychological reactions of the opponent escalating. But unlike the previous psychological studies mentioned, the explanatory power of the variables just discussed rely implicitly on the assumption that the "utility pie" is not fixed. There is no sense that there are cost-benefit tradeoffs made by the players or perhaps more generously, for every cost increase resulting from escalation, there is a counter-balancing value increase that justifies further escalation. By definition, the costs of escalation can never exceed the benefits of giving in and states should by implication escalate forever.

One explanation for the escalatory spiral is that the players become "entrapped" in the situation (Brockner and Rubin, 1985). Because each escalatory move (a bid) is only a small contribution to the overall cost of the game and because players have ill-conceived or vague notions of the limits to which they will play, actors tend to play well past the point when losses outweigh the gains of continuing to play (Brockner and Rubin, 1985). Brockner and Rubin formulate this argument along the lines of a waiting game where, much like the simpler versions of War of Attrition games (Maynard-Smith, 1972; 1974;
Parker, 1974) the basic consideration of the actors is how long (in real time) to hold out before quitting the game.

This theoretical argument raises two important dimensions to the overall problem of understanding escalation. First, the general claim is that players formulate thresholds (even if ill-conceived and not always adhered to) to orient their decision as to whether the costs of continuing the escalation game outweigh the benefits of stopping and cutting one's losses. Beyond this threshold value, players simply accumulate costs without receiving anything of value in return. If we accept this as a reasonable assumption, then this provides an endogenous cap to the level of escalation that should be observed. The empirical variation that is observed over the levels to which players are willing to escalate becomes a function of the different threshold values for the players. The second point raised by this discussion is that the rate at which costs accumulate can have a decisive impact on the players' decisions over whether to stay in the game or to exit. The reasoning process is that the cost differential between one second and two seconds is practically imperceptible and hence insignificant. So too is the difference between two seconds and three. But say for instance that time originally proceeded by the second but then suddenly each time the clock ticked, an hour elapsed rather than a second. It is no longer the case that the two time costs are indistinguishable. This discussion assumes that the actor is competing against time or nature but not against another actor. The logic underlying the impact of the rate of escalation on an actors' expectations and behavior has been extended to the two party case. Escalation or the rate of escalation is a behavioral choice executed by one or both of the actors. One the one hand, it is argued that small, incremental escalation steps allow the adversary time to adjust, absorb the costs of escalation fairly effortlessly and most importantly, gives the opponent the opportunity to counter-escalate (Ikle, 1971; Patchen, 1988; Haig, 1984). Escalation produces counter-escalation and so on until the process culminates into a costly outcome that may in the end provide very little
value. There is; however, some probability that an incremental escalatory move will be sufficient to cause the adversary to give in which provides the escalating actors with a good outcome on the cheap. On the other hand, a sudden, decisive escalation move is more likely to shift the opponent's expectations about the costs involved in the conflict which convinces him to back down. The downside associated with this strategy is that the higher one escalates, the closer one moves toward their own cost threshold (Rosen, 1972; Patchen, 1988). If the strong escalation move is insufficient to convince the opponent to quit, the opponent can counter-escalate and impose burdensome costs on the actor who made the first move.

From a prescriptive point of view, crisis management theorists often caution against the use of high levels of escalation initially (George et al., 1971; Snyder and Diesing, 1977; Craig and George, 1990). The logic is that by starting low, states preserve their "options" to move to higher levels of coercive force should the situation warrant such a move. The implicit assumption underlying this prescription is that it is extremely difficult for states to escalate at lower levels once they have escalated high, e.g., lower levels of escalation "options" are foreclosed.

Conclusions drawn from a number of different psychological experiments show consistently that a change in the rate of escalation has a significant impact on players' behavior subsequent to the action (Tedeschi, 1970; Vincent and Schwerin, 1971; Bonoma, 1975; Patchen, 1988). This variable, more than any other that falls into the "coercive" category, is more likely to elicit a cooperative gesture on the part of the victim. This is contrasted with cases where a constant rate of escalation was imposed on the opponent. In the majority of these cases, the opponent was non-responsive and comparatively less likely to switch from an escalation strategy to a cooperative one. Unfortunately, these data were not interpreted within any theoretic framework that could explain the results. Nevertheless, the data are consistent with a number of conjectures
emanating from outside the psychological literature. If these insights are to be of use, they must be integrated within a theory that produces hypotheses that are derived and then tested empirically.

There is probably a kernel of truth to the proposition (even in the social sciences) that things in motion tend to stay in motion. Yet the use of psychological variables to explain the dynamic carries with it a number of difficulties both theoretically and empirically that detract from their usefulness in theory construction. It is quite plausible that escalation causes negative emotions in the players and that these emotions can assume a range of different values. However, it is not clear what the underlying metric for a little or more angry would mean in this context. The problem is broader than this in that the conditions under which values of the variables will produce different reactions is not stated, indeed it is precluded theoretically. In practice, these variables themselves become constants because the same outcome occurs by definition. As a practical and theoretical matter, there is nothing that the players themselves can do to halt the escalation dynamic. The cessation of escalation must be a function of some arbitrary stopping point, e.g., war. The theoretical implication from the standpoint of escalation is that the process should continue unabated until one or both of the players is destroyed.

The only necessary information from the analyst's point of view is the level of resources that each side possesses. The cognitions and beliefs of the actors are exogenous to the game. However, empirical evidence is accumulating to suggest that states display different patterns of escalation behavior, all of which do not progress in an upward linear fashion. Moreover, some of these escalation patterns are more or less likely to lead to war. This suggests that the exclusion of other variables biases analyses of escalation in favor of finding escalation where different behavior may in fact exist. We in fact learn very little about the process of escalation when we approach it as an assumption for why some dynamics spiral upward.
There main problem associated with the research on escalation reviewed so far is that escalation has been conceptualized without regard to the an actor's underlying motivation to escalate and perhaps more importantly, to cease escalating. The assumptions used to model the escalation process lead to expectations regarding the extremes of states' escalation behavior. Either states escalate once and exit from the conflict or states are expected to escalate forever. These assumptions are of minimal use in understanding the escalation process since both of these expectations are falsified empirically. Therefore, theories addressing the dynamics of the escalation process must be based on a conceptualization of the escalation process that link the affect of important explanatory variables on actors' choices.

*Escalation as a Bargaining Strategy*

Many who study the international escalation process view escalation as a subset of the process of interstate bargaining (Schelling, 1960; 1966; Kahn, 1965; Young, 1968; Snyder, 1972; Lockhart, 1973; Smoke, 1977; Pillar, 1983; Maoz, 1985; Morrow, 1985; Powell, 1987; 1989; Patchen, 1988; Zagare, 1992). Bargaining is defined as a means by which two or more purposive actors reach a negotiated settlement to their dispute (Young, 1968). Schelling (1960) was among the first to recognize that the participants of most interstate conflicts have at least some interests in common even if this interest is as minimal as wanting to avoid mutual disaster. If in those rare instances the interests of the actors are diametrically opposed, bargaining becomes impossible. This is a case of a pure conflict or in game theoretic terms a zero sum game where the gains to one actor are equivalent to the losses suffered by the opposing player. The common interests of the parties serve as a foundation upon which some sort of compromise can be built.

The first condition necessary for bargaining to occur is the recognition that the parties have common, though not completely compatible interests. The second condition is that the players do not have complete information regarding the situation that both they
and their adversary face (Wagner, 1982; Powell, 1989). When players know in advance the agreement point, they will move immediately to that outcome and forego bargaining (Cross, 1969). This presumes of course that bargaining is not a cost-free strategy. If bargaining is costliness, players will bargain indefinitely. But if this criteria is satisfied, why would a bargainer ever accept anything less than his full demand? The answer, in the limit, is that he would not accept any agreement except the adversary’s full capitulation. We come full circle to the recognition that this is a zero-sum conflict. And because the players know this, and know the end game, a crisis or conflict will never occur between two completely informed bargainers. Thus, a third criterion for bargaining to occur is that the players have some incentive to reach an agreement. Players are motivated to find a solution to their conflict because bargaining is costly. In general, bargaining costs reduce the value of disagreement (Pillar, 1983; Patchen, 1988). It is the “power to hurt” that makes it so costly for the other side to continue playing which motivates that actor to accept some agreement over enduring the punishment. But as Schelling (1960) points out, successful coercion is coercion that is never used. Rather, successful coercion is the threat of future violence that so terrifies the opponent that he comes to believe that backing down immediately is in his best interest. This is not the diplomacy of handshakes and good will; it is the diplomacy of violence and the exploitation of the power to hurt the other side.

The ability to threaten, to frighten, to cajole and to manipulate the other side into backing down can only occur if the one or both of the players are uncertain about the their bargaining environment. If both players know the absolute minimum value the other is willing to accept, they will either agree to disagree or they will jump immediately to a bargaining solution and forego the costs of bargaining. Neither player can manipulate the beliefs of the opponent because the opponent knows everything with certainty about the game.
It is the incredulous nature of the threat to be carried out for noncompliance which stimulated theorizing on the interconnections between threats, credibility and nuclear war. As soon as the Soviet Union acquired a second strike capability (part of the Soviet's nuclear force could survive a U.S. first strike attack and counter-attack the U.S.) threatening the adversary with a nuclear strike would appear incredible given that the implication of a second strike also means the destruction of one's own society. Schelling began to theorize on the ways in which deterrence could be made credible. His conclusion was based on the idea that while it may not be credible to threaten the instigation of nuclear war directly, it may be credible to take actions which might produce this outcome. That is, states are tied together by a rope taking steps on a frictionless surface sliding uncontrollably toward the nuclear abyss. These steps taken may lead to war which lie outside the control of either party. For if one could control these actions, the adversary may reason that complying with the demands of the other is unnecessary given their ability to stop the dynamic at will. Thus, it is the autonomous risk that things will "get out of control" that makes the outcome of nuclear war credible. When both sides are taking steps and creating risks that things will get out of control, the crisis becomes a "competition in risk taking." The risk is generated by escalation. So escalation becomes a tactic that is designed to exploit the fear, or manipulate the adversary's beliefs that the costs of continuing the crisis outweigh the benefits of giving in now.

Schelling has conceived of the "competition in risk taking" as a type of brinkmanship game where players operate on the margins of disaster. The first player to 'blink' loses. Although it is probably more accurate to view this as a 'wink' since we are dealing with purposive state behavior (Brodbeck, 1959). One can also examine escalation as a form of limited punishment (Powell, 1988; 1990). The motivation behind this action is twofold. On the one hand, applying a limited cost on the adversary should decrease the value of disagreement in the short run. On the other hand, this strategy
communicates that more punishment is on its way. In effect, states are engaging in a strategy that has two components. First, actors are actually imposing costs on the adversary. Secondly, the act of escalation is backing up the credibility of the threat of future action by actually escalating in the conflict (Powell, 1990).

Conclusion

States can come into conflict with one another in a variety of international contexts. Whether that context is an arms race, a crisis, a deterrence situation or a war, states attempt to resolve the conflict in ways that advance their interests. Ideally, both sides to the dispute prefer to have their adversary satisfy their demands completely on the issue. However, since neither player’s goal can be satisfied simultaneously, the process of dispute resolution involves attempts to induce or persuade the adversary to agree to a solution more favorable to their interests. Within these conflicts, escalation becomes an important tool of persuasion. By actually applying sanctions on the opponent, escalation attempts to communicate to the other side that the costs of future punishment will outweigh the benefits of giving in to the coercive state’s wishes.

The bargaining process provides the conceptual bridge that links conflict initiation with the eventual outcome of the conflict. Since escalation can be an important dimension of the bargaining process, gaining a stronger appreciation for the dynamics of escalation can assist scholars in understanding the reasons for the variation in the outcomes observed empirically across different international relations contexts. Yet, there have been very few attempts to construct a well integrated theory of escalation that is capable of producing falsifiable hypotheses regarding the conditions under which states escalate or the reasons why states cease escalating. Part of the reason for this general lack of progress in the area of escalation has been the tendency to view escalation as context specific. The manner in which escalation is used varies widely depending on whether we are referring to nuclear deterrence (Brams and Kilgour, 1987; Zagare, 1987; Powell, 1990), arms races (Jervis,
1976; Intriligator and Brito, 1984) or militarized interstate dispute (Gochman and Maoz, 1984). However, once a general theory of escalation is constructed, the hypotheses regarding escalation can be tested against cases drawn from these different contexts. The advantage of this approach is that the insights derived from these more context specific studies can be pulled together into one coherent theoretical framework that explains escalation behavior.

Formal theory is a useful tool in constructing a general theory of the escalation process (Bueno de Mesquita, 1984; Maoz, 1985; Morrow, 1985; Lalman, 1990). One of the main advantages that formalization bring to theory development is analytical rigor. The simplification process is important in identifying key assumptions and clarifying concepts. Once the complexity of an escalation situation has been stripped down to its essential components, it then becomes possible to incorporate complicating features into this basic model structure in a logical and consistent fashion.

The next chapter develops the theory and models used to derive hypotheses regarding the dynamics of escalation. These models are designed to answer two general research questions. First, what are the conditions under which states escalate in conflict and what conditions are necessary before states decide to end the dispute? The second research question extends that analysis one step further and asks: given that states have decided to escalate, what level of escalation are we likely to observe? Chapter three develops the tools to address these questions.
Chapter III
Theory and Models

Introduction

The principal argument of the previous chapter is that our understanding of the escalation process has suffered from inadequate conceptualizations and incomplete theoretical frameworks used to explain the escalation process. While it is important to begin the modeling enterprise with a simplifying set of assumptions, the failure to extend analyses beyond these restrictive set of assumptions perpetuates an understanding of escalation which is limited to the margins of state behavior. States are either expected to escalate until they have exhausted all of their resources (e.g., war of attrition, Maynard-Smith, 1972; 1974; Fudenberg and Tirole, 1991) or states escalate once and drop out of the conflict (Powell, 1987; Zagare, 1987; 1992). These studies may provide an interesting set of conclusions for certain research questions but they are inconsistent empirically with the vast majority of escalation observed during the process of dispute resolution (Bremer, 1982; Cusack and Eberwein, 1982; Maoz, 1983).

This chapter develops the theory of escalation. The theory begins by conceptualizing escalation as part of interstate bargaining. Escalation is defined precisely along with additional concepts that support the overall context in which escalation occurs. A clear conceptualization of escalation is a critical part of the theoretical construction because it anchors the interpretation of the theory. The variables posited to affect the escalation dynamic are identified and the interrelationships among these variables are specified in terms of the affect they have on the escalatory behavior of the actors. The variables are introduced as they become relevant in different stages of the theory's development. The over-arching goal of the thesis is to derive several hypotheses, some of which are tested empirically in later chapters. The theory is specified for the most general case which permits testing these hypotheses across a variety of escalation contexts.
such as crises, extended deterrence, arms races and war. The hypotheses are derived in
this chapter and given substantive interpretation. In subsequent chapters, a subset of the
hypotheses are tested in some of the contexts mentioned above.

The theory of escalation developed is based on non-cooperative game theory. All
game theoretical models deal with situations involving interdependent decision making
(Young, 1968) in which the outcome of any contest is a function of the intersection of the
actors' choices. Game theory is a useful tool to analyze the problems inherent in all
escalation situations for several reasons. From a conceptual point of view, even though
the behavior of a single state can be defined as escalatory, the final outcome of any conflict
situation involves the actions taken by both participants in a conflict. Thus, states' decisions
over whether to escalate and choices between the level of escalation to employ must be made with a view toward the anticipated reaction of the opponent. Game theory is
designed specifically to understand the problems of this type of strategic interaction.

From the point of view of modeling the escalation process, game theory offers
several advantages over other types of theoretical construction. First, formalizing the key
components of the model requires that the assumptions used to generate the model are
forced to the surface and are made explicit. Second, the model must be simplified to
maintain the mathematical tractability of the problem. The simplification process is
important to the extent that the most basic elements necessary to any escalation situation
are more readily identified and the interrelationships among the variables are exposed. The
danger that simplification carries is that the contextual richness of the situation being
modeled is reduced to a trivial problem theoretically. Descriptive accuracy is an important
element of all models. However, the explanatory and predictive structure of the model
must first be established before the complexity of the empirical domain can be understood
within a broader context.
The model of escalation is developed in a series of stages. The first model version depicts a simple game of escalation in which the actors compete in a highly stylized environment. The simple model includes many of the basic components that capture the escalation process. This model serves as a foundation to guide the progression in incorporating complicating features into the model. This modeling procedure produces a new model version which is simply an extension of the preceding model(s).

There are several reasons for constructing the escalation model in stages. First, because each model is a generalization of its immediate predecessor, the hypotheses that each model version produces are compared and the conditions under which earlier hypotheses hold are refined. These hypotheses are compared theoretically to determine if, or exactly how, our expectations change as greater realism is injected into the model's structure. More complicated versions show how these hypotheses deviate from the baseline prediction derived from the simple model. Second, a theoretical discussion accompanies each model and exposes any conceptual weaknesses in that model's ability to address certain research questions of interest. The extensions necessary to produce such answers are incorporated into the model in a logical, consistent fashion. This effort produces a general theory of the escalation dynamic that applies across several different conflict situations.

This chapter is structured as follows. The first section is a discussion of the general conceptual context that leads to the genesis of the escalation game. The second section details the assumptions and components of the basic escalation model. Throughout the course of the discussion, the model's components and the hypotheses derived from each model are given substantive interpretation. The conclusions of each model are analyzed in terms of their ability to address the various research questions of interest. The conceptual weakness of the model is highlighted and this leads to a discussion of the modelling extensions necessary to lead us in the direction desired.
The Escalation Context

All states have various goals that they pursue in the international environment. Sometimes states pursue aims that come into conflict with the foreign policy preferences of other actors in that system. However, whether the conflict remains latent or not depends on the parties mutual recognition that their differences need to be resolved in some fashion. For approximately forty years, the Soviet Union and the United States harbored diametrically opposed policy preferences over the type of domestic political system that should govern the majority of states in the international system. The superpowers, however, only came into direct conflict with one another when it seemed as if ignoring the opponent's actions would lead to an unacceptable advantage for one side or the other.

The central method for resolving interstate conflict is bargaining. Bargaining is defined as a process by which two or more purposive actors reach an agreement (Young, 1975: 5). The bargaining process unfolds as the interplay of a series of actions that players take to induce the opponent to agree to a settlement. The bargaining context simply provides the protagonists with a forum within which to resolve their underlying differences over the issue at stake. However, sometimes the players' interests are so diametrically opposed that any attempt to find common ground as the basis for a negotiated agreement will fail. Under this condition, escalation is not used to persuade, influence or convince the other side to back down. The diplomatic dimension of escalation becomes ineffectual and is abandoned in favor of inflicting pure violence which does not persuade the opponent to back down but forces him to do so (Schelling, 1966). There are one of two ways in which this type of conflict can end. First, both players continue to escalate until the pain has become insufferable to one side thereby forcing that actor to drop out of the conflict or because both players are mutually destroyed in the process (Powell, 1987). Second, we can assume that some exogenous mechanism steps in and imposes an outcome on the players. This assumption is not uncommon to formal theories of
deterrence which assume that this mechanism acts as a trigger that unleashes an automatic escalation to nuclear war (Nalebuff, 1986; Powell, 1987; 1990; Brams and Kilgour, 1987). Imposing a solution on the players is a useful simplifying assumption because the theory of escalation developed in this chapter is less concerned with the outcome of a conflict than it is with identifying the conditions under which mutual escalation will occur in the empirical setting.

For those players who do have common interests and are attempting to discover the areas where their interests overlap, the immediate conflict is over the distribution of gains that will represent the final settlement. Bargaining can be characterized at one extreme by a player who gives in as the other does expecting some kind of intermediate outcome. At the other extreme, players expect the opponent to make all of the concessions (Cross, 1965). Most bargaining situations fall somewhere in between. The players articulate initial demands that are of higher value that what they will eventually settle on providing room on both sides to make concessions toward the final settlement.

Players are motivated to make concessions and to achieve an agreement because bargaining is a costly process that makes disagreement a costly state of affairs for both players. The costs of bargaining can be evaluated in terms of time, the costs involved in gathering information or a whole host of other costs that accumulate every time the players reach an impasse. The affect of these increasing costs is a reduction in the net benefit eventually acquired by the bargainers thereby making the current disagreement a relatively less attractive alternative.

Players can avoid the costliness of disagreement and induce some sort of solution through two general means. Accommodative actions motivate the opponent to achieve a settlement by rewarding concessionary behavior with a reciprocative reduction in a player's own demands. Concession making is often associated with the belief that it is an effective tool to hasten agreement by signaling one's flexibility over the level of benefit
sought (Pruitt, 1981). The hope is that by conveying the impression of 'playing fair' or bargaining in good faith, the opponent will be similarly motivated to offer concessions. The affect of a concession is to increase the value of the eventual settlement for the recipient of that offer and this increases the chances that this bargainer will respond in kind (Zeuthen, 1968). Yet bargainers are sensitive to the potential tradeoffs between facilitating an agreement and minimizing their exposure to losses. Concessions may be interpreted as a sign of weakness and not as a gesture of good faith bargaining. The potential danger is that the opponent may solidify his bargaining position by assuming that greater concessions from a weak adversary will be forthcoming.

The alter ego of accommodation is coercion. Coercion is a broad category of tactics whose elements include threats, intimidation, blackmail and most importantly, escalation. In the broadest sense, the purpose of coercion is to convince the adversary that the costs for failing to give in will outweigh the benefits of holding out for a better deal. States accomplish this by reiterating their initial demand and threatening the adversary with punishment if compliance with that demand is not forthcoming. Accommodative bargaining tactics increase the value of the eventual agreement by conceeding something of value that represents movement in the direction of the opponent's preferred outcome. Similarly, coercion also increases the value of agreement not because the solution has become more favorable but because the costs for non-compliance are mounting which depreciate the value of the final outcome.¹

A necessary condition for the success of coercive threats is that the victim must believe that compliance will prevent the execution of that threat. If it is believed that the

¹ The conceptualization of bargaining offered here differs from many economic models of the bargaining process (e.g., Rubinstein, 1982; Fudenberg and Tirole, 1983; Sutton, 1986). Rubenstein (1982) conceives the bargaining problem as how to divide a pie between two players. Players bargain until an offer is accepted by one side. In my conceptualization of the bargaining process, the players are not concerned with distributing the gains from a pie. First, players may or may not agree on the size of the pie. Second, the players bargaining behavior consists of efforts aimed at convincing the opponent to cede the entire value of the prize.
threat will be carried out irrespective of an actor's behavior, then there is no reason why an actor should concede given that yielding simply magnifies the victim's losses. However, the coercing state has no incentive to punish a submissive actor because the executioner also suffers costs for carrying out the punishment. Therefore, it is the exploitation of the potential use of violence to elicit a concession and not the actual application of violence that qualifies as the most successful manipulation of force (Schelling, 1966).

The recipient of a threat can of course either yield to intimidation or issue a counter-threat of his own. The failure to comply presents the coercive state with the dilemma of giving in or carrying out the punishment. Theoretical inquiries designed to explain outcomes of strategic situations such as the one described above often collapse the decision to retaliate into one grandiose cost imposing activity. However, this assumption ignores the motivation of the retaliator to minimize his costs and obscures the subtleties involved in unraveling the bargaining dynamic over time. It is certainly the case that the coercive state's decision to impose costs on the adversary can involve a continuum of different levels of punishment during different stages in the conflict. However, the choice of the level of costs to impose is actually secondary to the more fundamental decision that has been made: the decision to escalate.

Escalation is a particular kind of coercion that imposes costs on both the recipient and sender throughout the bargaining process. Many definitions of escalation frequently contain stipulations that the magnitude or severity of the punishment increases with each successive escalation action taken by the participants. These stipulations are tantamount to underestimating the behaviors that can qualify as escalation. This definition limits both the theoretical and empirical applicability of the concept to a few select cases. Moreover, this definition confuses escalation proper with a description of how states escalate.

---

2 No assumption is made that the costs the imposer suffers in escalating are equivalent to the costs suffered by the recipient of that escalation move.
Actor's choices over the magnitude or the severity of their cost imposition is an expectation about state behavior that should be derived from the theory and not used as a criterion for inclusion.

The definition of escalation offered here is sufficiently broad to include the imposition of all levels of cost over time. A state's repertoire of escalation actions range anywhere from issuing a diplomatic protest to dropping a nuclear weapon on the rival country. The evaluation of these costs also varies from recipient to recipient. The American naval blockade of Cuba in late October of 1962 was costly for the U.S. both in terms of the financial expenditures necessary to sail these ships from their home port to their destination but also in terms of the unrealized costs that sparking a nuclear war with the Soviet Union might entail. The cost to the Soviets was that by adhering to the blockade, they were unable to get supplies through to Cuba and perhaps the costs Khrushchev himself experienced as a result of appearing to waver against the adversary. The costs of escalation can also be evaluated with regard to the destruction of property and/or the loss of military and civilian lives. Clearly, the horror that nuclear weapons inspires is the rate of devastation wreaked on all forms of existence. But lesser campaigns can carry serious costs as well. The Chinese bombing campaign on the island of Quemoy was an attempt to make it so costly for Chiang Kai-Shek to remain on the island that he would evacuate his Nationalist forces.

Other specific examples of escalation could be given. The general point, however, is that escalation is employed for two reasons. First, escalation raises the current costs of bargaining for players thereby decreasing a player's willingness to endure the punishment. Second, escalation is also an attempt to exploit the adversary's fear that the current costs of escalation are marginal compared to the future devastation that will occur if the adversary does not concede immediately. These two aims of escalation are separated for analytical purposes, however, in practice the signals that escalation sends are interdependent and
reinforce one another. The believability of future punishment is strengthened or made more credible both by past and current escalatory behavior and due to the fact that escalation is costly for the imposer (Powell, 1987). A willingness to suffer a self-imposed reduction in the net benefit from bargaining conveys the escalating state’s seriousness in getting his demands satisfied.

It is unnecessary to assume, nor is it likely to be the case, that the players are affected similarly by the costs of bargaining. Thus, bargainers who occupy a relatively superior position in terms of the ability to withstand the costs of bargaining can exploit this advantage by wringing concessions from the other side. For example, the landlord who owns the only vacant rental unit in the village can demand an exorbitant monthly fee from potential occupants who have no other available housing alternatives. The renter’s cost in rejecting the asking price is that he has nowhere to live. The landlord suffers the costs of having to readvertise the vacancy and costs of repeating the process with some other potential occupant.

A necessary condition for escalation is the opponent’s failure to comply with a demand. Every non-compliance, however, is not matched by an escalation response. At some point in the bargaining process, one or both of the players conclude that further escalation is an ineffective strategy to secure anything of value. Escalation adds nothing but costs to an outcome which is already fairly grim. Consequently, the dejected player gives in to the adversary’s demands. The identification of a threshold distinguishing the value of one more escalation move versus giving in raises an important theoretical question as to the criteria players use to define conceptually the maximum amount of escalation costs they will suffer before they exit from the bargaining process.

One common conceptualization of the limits to which players will escalate is illustrated by a story drawn from folk-lore. The story is set in the late 1840s at the zenith of the California gold rush. Every Saturday night a dance was held at the town hall. One
evening, two gold miners began to argue over which of the two was going to have the last dance with the dance hall girl. It was an established custom that when miners came into conflict with one another that they would resolve the dispute by playing a game instead of using their fists to beat one another to a pulp. The winner of the game would get the hand of the dance hall girl for the final tango of the evening. The rules of the game are simple. The miners alternate throwing gold coins into the San Fransisco Bay which are taken from a pile of gold hidden from the opponent's view. It is assumed that the price a miner is willing to pay to play the game is equivalent to his level of resources. There are no limitations on the number of coins a player can throw on each turn and the player who throws the most gold into the Bay wins the dance hall girl's hand. Clearly, if the game is played to resource exhaustion, the loser in this game is the player who has the smaller pile of gold. The order in which the players toss gold is irrelevant in determining the winner in all cases except when the players' resources are equivalent. In this case, the player throwing second wins the game with one gold coin. However, neither player can determine a priori who will win and lose given that their piles of gold are hidden from sight. If we relax the assumption that the miner's pile of gold is concealed from the other side, then no miner would ever throw a single coin into the Bay. The player with the smaller pile of gold would concede defeat immediately to his opponent in order to prevent losing all his gold in addition to inevitably losing the last dance.

The San Fransisco story represents a war of attrition type conceptualization of escalation (Maynard-Smith, 1972; 1974; Parker, 1974; Hammerstein, 1982; Fudenberg and Tirole, 1991). The players continue to escalate (throw coins into the Bay) up to and including the point at which they have exhausted all of their resources. The loser of the conflict is the first player to expend all of his resources. The success of the war of attrition as an interstate war fighting strategy is also based conceptually on the ability to wear down the physical and moral resources of the enemy (Maoz, 1990). The war of attrition
conceptualization of escalation may be a partially accurate description of the way that some conflicts unfold over time. However, there are several implicit assumptions underlying this perspective which limit its utility as a general principle to guide the explanation and prediction of states' escalation behavior.

It might make sense theoretically for a state to consume all of its resources if it believed that it would never have to engage in another fight. Since states can never be guaranteed of this outcome in an anarchical international setting, states will behave in ways that minimize their vulnerability in the next conflict. The utter devastation characterizing Europe following WWII, left the Europeans exposed and unable to resist the intrusion of potential aggressors. The American response was a massive infiltration of economic and military assistance given not for altruistic motives but out of the perceived necessity to shore up Europe's ability to defend itself in the event of future hostilities. The interdependence of the state system requires that states reserve some of their resources for use in future conflicts as well as sustaining, with as little disruption as possible, their current way of life.

An interrelated problem of the war of attrition conceptualization for the study of interstate escalation is the implicit and untenable assumption that both players are ignorant and also assign the same value to the disputed prize. The practical consequence is an exclusive focus on the capacity to inflict costs (exhaust resources) and a dismissal of the bargaining dynamic that motivates actors' escalation behavior in the first place. For instance, what if one of the miners hated dancing? Would we still expect that miner to throw most of his gold into the Bay for a prize that would mean very little to him if he won? Probably not. The reason is that the next conflict may offer a prize so valuable to the miner that he would be willing to use the resources saved in the last dispute to fight for the prize in the current bout. This implies two different things about the way that states manage their resources during a conflict. On the one hand, states *save* some portion of
their resources for future conflicts and on the other, states *expend* resources in the current conflict based on the strength of their desire to win the object at stake. The greater the value of a prize to a player, the greater the level of resources they are expected to consume in the current conflict to obtain the prize. Conversely, the less salient an object is fewer resources are spent to secure the prize.

The maximum costs a player is willing to suffer to achieve some object of value is referred to as a player's cost tolerance. Cost tolerance represents the upper limit or threshold to which a player is willing to escalate (expend resources) to achieve some valued prize in the conflict. Note that the maximum costs a player will endure is not necessarily synonymous with the level of resources that can be accessed for the purpose of fighting. It is the value of the prize and not the amount of resources one can expend that determines the price a player is willing to pay for an object. When the total value of the costs of escalation have accumulated to equal the value of the prize, a player has achieved his cost tolerance. Experiencing any further escalation means that the costs of continuing to play the game now outweigh the benefits. At this point, actors receive more value from quitting the game and cutting their losses.

This notion of cost tolerance is based on the assumption that players treat the costs of escalation as cumulative. The accumulation of costs incurred on previous escalation moves can affect players' subsequent decisions regarding the benefits from continuing or quitting the escalations game. The cumulative cost assumption is common in many war of attrition models, e.g., Maynard-Smith, 1972; Nalebuff and Riley, 1984 and dollar auction models (Shubik, 1971; O'Neill, 1986). These costs continue to mount until one or both of the players quits the escalation game (reach their cost tolerance in that period).

---

3 Many economists interpret these costs as "sunk costs" which are assumed to have no affect on the actors' future decision making. I am assuming something quite different. Every time a player escalates, both actors suffer some level of cost. These costs can affect a players' future evaluation of the benefits from remaining in the escalation game.
In contrast, many psychological models designed specifically to study escalation assume that costs are evaluated "as if" players had no memory of paying out or suffering costs in the previous period(s) (Smoke, 1977; Brockner and Rubin, 1985; Pruitt and Rubin, 1986). With this conceptualization, costs are not an important variable in players' decision to remain in the escalation game. However, the sunk cost assumption has the attractive feature of placing an endogenous cap on the level of escalation which, as mentioned previously, is consistent with what is observed empirically. Second, this assumption does not necessarily imply that the value of a player's cost tolerance remains constant throughout the course of the conflict (Rosen, 1972; Cannizzo, 1980). For instance, players who have a relatively larger industrial base and have a better capacity to replenish the material necessary to execute their escalatory actions, may experience a marginal decrease in the costs of fighting (Rosen, 1972). This implies that a stronger player's cost tolerance has a positive slope over time. In other words, this actor is willing to suffer greater costs over time but is unwilling to continue the dispute, if in any one escalation period, his cost tolerance is surpassed. Varying a player's cost tolerance over time is an interesting extension that will eventually be incorporated into the model. For now, the integrity of the theoretical argument is captured with the simplifying assumption that a player's cost tolerance is constant throughout the conflict.

As implied by the discussion, a player's cost tolerance or willingness to escalate can vary from conflict to conflict depending on the value of the issue at stake. It is difficult to determine in advance which issues will assume greater or lesser salience for a nation. The question is an important one though because it allows us to predict, in the limit, the level to which a state is willing to escalate in the conflict. However, certain issues are easier to identify than others. Probably the most salient issue for any state is the defense of its territorial integrity. By definition, a state is identified by its territorial boundaries and so protecting them implies maintaining its current existence. States may be
willing to pay high costs for issues other than territory. Sometimes the prize is who is going to have influence over another state as is the case in extended deterrent crises. In 1885, Russia seized territorial control over parts of northern Afghanistan which was then a British protectorate. The British had little inherent interest in Afghanistan except insofar as Afghanistan was a buffer against direct invasion to the jewel of the colonial empire, India. England threatened to declare war against Russia if the Russians persisted in reneging on the compromises they had accepted in previous negotiations. Russia, convinced that the British would escalate the conflict to defend Afghanistan, and indirectly India, backed down under the British threat to play a costly game of escalation.

If we substitute cost tolerance for the exhaustion of resources in the San Fransisco story, the expectation is that players will escalate until one of the players' cost tolerance is reached. The players will not spend all of their resources but they will escalate to the level of the loser's lower cost tolerance. The result depends on the assumption that the players (miners) do not know the exact value of the opponent's cost tolerance. Moreover, the players are ignorant in terms of their inability to assimilate the information that is being transmitted to them through their opponent's escalation behavior. The fact that the opponent has escalated, or escalated at a particular level, only signals to an ignorant player that he must escalate again. The ignorant player in effect learns nothing about his opponent's situation during the play of the game. Since escalation represents nothing more or less than a means to an end, the complexity of escalation as a manipulative bargaining tactic or as a tactic to convey information to the adversary is eliminated.

The outcome of any bargaining situation depends on the strategic interaction of both players during the process of resolving the conflict. The players' decisions over whether to escalate and what level of escalation to employ are therefore conditional on the beliefs that a player has about the opponent's likely counter-response, what the opponent believes about his own situation, etc. The knowledge that players lack is the cost tolerance
of their opponent. This attribute is the key determinate of a players' willingness to punish the adversary and to absorb the costs of escalation that accrue throughout the duration of the conflict. Thus, precise knowledge of the opponent's cost tolerance informs a player as to which of the two disputants is likely prevail in the conflict. When a player knows a priori that a loss is inevitable, that actor would never play the game. However, states, like the San Fransisco gold miners, must make escalation decisions in an uncertain and risky environment in which they do not know with certainty, but have expectations over, the willingness of the adversary to escalate. Escalation; however, operates as more than a means to some outcome. Escalation is used to signal one's true of false intentions to suffer costs such that the opponent becomes motivated to concede. Escalation cannot function in this capacity unless we assume that players can somehow reduce their uncertainty in the game. If we assume that players can use the cues provided to them through their opponent's behavior, then players will be able to either confirm or disconfirm their beliefs regarding the opponent's relative cost tolerance. These evaluations can then be used to reevaluate the benefits from escalating in the future.

The condition under which the escalation dynamic ends in a game played by two ignorant players, is the point at which the lower cost tolerance of one of the players is achieved. An escalation game, played by two players who can learn about their opponent's situation through escalation, can end at any point prior to and including the point at which a player's cost tolerance is reached. The escalation game is played only so long as the players believe there is a sufficient chance to prevail (players believe that they have a higher cost tolerance than the opponent). The failure to achieve this sufficient condition can occur at any stage in the game and when it does, players exit from the game to minimize their potential losses.

This opens up the possibility that players can manipulate the opponent's beliefs about their willingness to tolerate the costs of escalation. The incentive for a low cost
tolerant player to escalate against a higher cost tolerant adversary is the value that he
receives by manipulating the latter into giving in to the former's demands. Both players
are aware of the fact that the failure to reach agreement is becoming increasingly costly to
both sides. However, escalation is used in part to convince the opponent that the costs of
future punishment will be so devastating that it is in the opponent's best interest to give in
immediately and avoid suffering the costs associated with an inevitable defeat since the
coercive state is in a superior position to withstand the costs of escalation relative to the
adversary. The state with the higher cost tolerance also has an incentive to communicate
his bargaining advantage to the adversary. The sooner the higher cost tolerant player
convinces the opponent that he cannot possibly win, the fewer the costs the former has
to pay to get his demands satisfied.

Escalation is a coercive bargaining tactic states use to inflict punishment on the
adversary and is used to back up the believability of future cost imposition (Powell,
1989). States can heed these threats or counter-escalate so long as their cost tolerance
threshold is not surpassed. States identify the maximum costs they are willing to suffer in
a conflict because resources are precious and states want to minimize costs in conflicts
they will lose. Simultaneously, players are attempting to maximize their chances of
winning those conflicts where the stakes are defined to be more valuable, relatively
speaking. In those cases where victory is quite probable, states may also have an
incentive to escalate at lower levels to minimize the costs necessary to win. Because these
players are more willing to endure the costs of escalation, they can always shift to higher
levels of coercion if lower levels of escalation are insufficient to convince the adversary to
back down. Players decisions to escalate, counter-escalate or give in to the demands of
the opponent under the weight of coercive pressure are intertwined with their beliefs about
the opponent's bargaining position and an overall evaluation of the situation at hand. As
such, the escalation-bargaining game is replete with opportunities to misrepresent or to
genuinely convey that willingness to endure the costs of escalation.

To capture these theoretical arguments, the next section develops several models that are used to derive hypotheses identifying the conditions under which players escalate or concede in a conflict. The last model produces hypotheses concerning both the conditions of escalation and predicts the levels to which players will escalate. The first model is a simplified representation of the escalation process. The basic model is extended by relaxing some of that model's more restrictive assumptions. In this way, I am able to address different types of research questions that are better analyzed with more complex and realistic escalation scenarios.

Model I

All game-theoretic models include assumptions about the number of players, the level of information each player has, their strategy choices, the outcomes of the game and the values that the players attach to each of these outcomes. Once the model's components are specified, the conditions under which each of the outcomes is an equilibrium are determined. The equilibria serve as the predictions of the game and the explanation for those predictions is contained within the model's structure.

Assumptions

All conflicts involve at least two actors. The principal players on the international stage and in this study are nation-states. These actors are denoted generically throughout the thesis as players i and j. I assume that states are both rational and unitary actors. Rationality refers to actors who have a transitive and connected set of preferences (Abrams, 1980). A connected set of preferences simply means that actors can rank the set of all possible outcomes. However, it is not necessary that a player strictly prefer one outcome to another for every pairwise comparison. Players can be indifferent between two outcomes, however they must be able to compare the outcomes and state their indifference. The transitivity condition restricts the logical ordering of an actor's
preferences. For example, if actor i prefers any negotiated settlement to war and war is preferred to doing nothing, then i must prefer any negotiated settlement to doing nothing. If the transitivity condition is not satisfied, it becomes impossible to produce expectations regarding the kinds of choices an actor will make in a given circumstance. Connectedness and transitivity are both necessary and sufficient conditions to ensure that an actor is rational. Note that the requirements for rationality are independent from assumptions pertaining to the level of information available to the actors. Rationality is often associated with the conditions of complete and perfect information in which an actor knows everything about the game situation that they are playing (Allison, 1971). The information that an actor has in the game and the rationality of that actor are separate concepts.

Assuming different levels of information can affect the expected behavior of the actors; however, this does not change the basic fact we continue to analyze the behavior of a rational actor.

States are also assumed to behave as if they are unitary actors. The concept of "the state" is of course a reification of an amalgam of individuals who reside collectively within territorial boundaries. Since the state itself is an abstraction and cannot make decisions, I assume that the state is represented by a dominate leader who behaves "as if" these choices are made on behalf of the state's national interest (Bueno de Mesquita, 1981). This assumption is useful because it minimizes the problems inherent in representing (and aggregating) the preferences of a complex array of domestic political actors within each state. It is important to begin with the simple assumption to establish the basic escalation dynamic before complicating features are incorporated into the model.

Throughout the thesis, I concentrate exclusively on the dyadic level of analysis. The main reason for using this simplifying assumption is to avoid obscuring any fundamental relationships that may remain hidden when analyzing complex circumstances of coalition formation and dissolution at the system level.
The play of any game is a product of the combination of choices that players make throughout the game. A player's strategy for the game is a specification of which alternative he will choose at every stage of the game. I assume that players i and j have a binary choice between escalating or not escalating (conceding). Escalation is defined by two elements which occur simultaneously. First, an escalating player reiterates his original demand, which if satisfied by the opposing player on the next move ends the game. The stipulation that coercion involves reiterating a demand is consistent with the assumption that players want to hasten agreement, not by making concessions to the adversary, but by making bargaining too painful for the opponent to continue playing the game. Second, escalation involves taking some action which imposes costs both on the adversary and the initiator of the action. Each time one of the players selects the escalation alternative, a cost is levied on both players. These costs are additive and are reflected in the payoff value for any outcome that is preceded by one or more escalation moves. Escalation implies that the game continues by shifting the burden back to the opposing side who also has the option of escalating or giving in. This is true for all but the final move, which by definition, ends the game. Escalation on a player's final move implies that the players could not achieve a settlement of their differences. When this occurs, I assume that an exogenous player, Nature, steps in following the last escalation move and imposes a solution on the players. This is the only condition under which Nature plays in the simple escalation game.

The other choice available to the actors is to not escalate (concede). A concession is defined as giving in to the full demand of the opponent. I assume that a concession is equivalent to capitulation to avoid the complexity of integrating the negotiating dimension of state behavior into the escalation game. Players are not assessed a cost for the concession strategy. As soon as one of the players concedes, the game terminates.
Figure 1 illustrates the escalation game in extensive form. The extensive form represents the players, the sequence of play, the players' choices, the payoffs and the information each player has about the game. The game begins when one of the players makes a choice. I assume, without loss of generality, that player i moves first followed by player j who has observed i's action. If i concedes on his turn, j does not need to move because the game ends by definition. If i escalates, j must decide whether to escalate or concede to player i. Player i's concession plus the intersection of i and j's strategy choices produce three different outcomes to the game. Player i's immediate concession to j's demand is described as a loss for i and a win for j. When i escalates and j concedes to i's demand, the outcome is a win and loss for i and j respectively. The final outcome occurs when both players escalate. Since the players themselves can only end the game when either of the players concedes, Nature has the final move and imposes an outcome on the players to end the game. I assume that Nature selects a compromise solution that lies midway between the demands of both players. Figure 1 does not show Nature's move but the payoffs that Nature imposes on the players are represented, \((a_i, a_j)\).

A payoff is assigned to each of the players' three outcomes. A payoff is the value that each player attaches to that outcome. The value for each of the outcomes is represented generically in Figure 1 by the letters a, b and c. Specific numerical values are not assigned to the payoffs to represent the possibility that the players can have a different ordinal rankings over these outcomes under certain conditions. Even though players i and j are assigned the same letter to each of the three outcomes, this does not necessarily mean that the players attach the exact same value to those outcomes. Any escalation costs that are incurred on the way to some outcome are absorbed in the payoffs and not denoted by a separate term. There is only one payoff restriction in this game. For player i: \(b_i > c_i\). Player i must prefer j's concession to his own otherwise i always concedes immediately and the game would never get played. The following is an interpretation of the payoffs to
FIGURE 1
Complete Information
Game

\begin{itemize}
\item \textbf{(ai, aj)}
\item \textbf{(bi, bj)}
\item \textbf{(ci, cj)}
\end{itemize}
i and j. The subscripts on the payoffs represent the value to that particular player:

\[ a_i : \text{The value to i of the compromise outcome minus two escalation costs} \]
\[ a_j : \text{The value to j of the compromise outcome minus two escalation costs} \]
\[ b_i : \text{The value to i of j's concession minus one escalation cost (a win for i)} \]
\[ b_j : \text{The value to j of conceding to i's demand minus one escalation cost (a loss for j)} \]
\[ c_i : \text{The value to i of conceding to j's demand (a loss for i)} \]
\[ c_j : \text{The value to j of i's concession (a win for j)} \]

If we assume momentarily that there are no costs associated with the decision to escalate, both players rank the outcomes of the escalation game as follows: win > compromise > lose. Since the players cannot achieve their most preferred outcome simultaneously, we always predict that the players will achieve a compromise settlement to their dispute. Once we reintroduce the costs of escalation, this expectation may not hold. The reason is that under certain conditions, the costs of escalation incurred by either of the players may be so burdensome that the value of conceding to the opponent's demands is greater than the combined costs of escalation to achieve the compromise outcome. The players' preference ordering over these three outcomes hinges on the willingness of the players to absorb the costs of escalation relative to the evaluation of the issue at stake. The same player may have a different preference ordering over the outcomes in different conflict situations depending on his assessment of the value of issue under contention. For example, in the late 1800s, Bismarck was willing to suffer tremendous costs during his battles against Austria-Hungary, Denmark and France in an effort to achieve statehood for a greater Germany. However, in 1911, when the issue at stake in Adagir crisis was the solidification of the German ego as a great power, German politicians were unwilling to pay the probable costs necessary to force Germany's rivals to recognize its great power status if that meant a war against France, and more seriously England.

The final assumption of model I is that players have complete information. Complete information refers to a situation in which the players know the rules of the game
and any move by Nature is observed by both players (Rasmussen, 1989). The implication of this assumption is that the outcome of the game is determined a priori. Both players know which alternative the opponent will choose and therefore, which outcome will prevail. The importance of the complete information assumption as a first cut at the problem of escalation is that we can derive a baseline prediction concerning what will occur in an ideal escalation situation.

Complete information is represented in Figure 1 by the circles encompassing i and j's choice nodes. These circles are referred to as information sets and are used to illustrate the information that a player has concerning his location on the game tree. The information sets in games of complete information contain one choice node because players know their exact location in the tree.

Solution Concept and Equilibria Outcomes

The purpose of all game-theoretic models is to determine the conditions under which each outcome is an equilibrium. An equilibrium is defined as a combination of strategies in which no player has an incentive to deviate from his current strategy given the strategy of the opposing player. In other words, when both actors are playing their equilibrium strategies, neither player can improve his payoff by unilaterallyswitching to a different strategy. The solution concept used to identify the equilibria in this game is subgame perfection. Subgame perfection is based on the ordering of moves and the distinction between an equilibrium path and an equilibrium. The path is the path through the game tree that is followed in equilibrium. The equilibrium itself is a strategy combination which includes the players' responses to other players' deviations from the equilibrium path (Rasmussen, 1989: 83). The following are the conditions under which each outcome is an equilibrium:
<table>
<thead>
<tr>
<th>Outcome</th>
<th>Preference Orderings of the Players</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (c_i, c_j)</td>
<td>i: b_j &gt; c_i &gt; a_i  \ AND \ j: c_j &gt; a_j &gt; b_j</td>
</tr>
<tr>
<td>2. (b_i, b_j)</td>
<td>i: b_i &gt; c_i &gt; a_i  \ AND \ j: c_j &gt; b_j &gt; a_j</td>
</tr>
<tr>
<td>3. (b_i, b_j)</td>
<td>i: b_i &gt; a_i &gt; c_i  \ AND \ j: c_j &gt; b_j &gt; a_j</td>
</tr>
<tr>
<td>4. (a_i, a_j)</td>
<td>i: b_i &gt; a_i &gt; c_i  \ AND \ j: c_j &gt; a_j &gt; b_j</td>
</tr>
</tbody>
</table>

The game is solved through backward induction. Backward induction involves solving the game for the optimal choice of the last mover for each possible situation he might face and then working backward to compute the optimal choice for the player before (Fudenberg and Tirole, 1991). For example, suppose that the players' preference ordering are those of equilibria 1. To predict the c_i, c_j outcome, we begin with player j who has the final move in the game. At j's choice node, j compares the value he receives from escalating, a_j (the compromise), versus the value he receives from conceding to i's demands, b_j (j loses). Since a_j > b_j, j will escalate if given the opportunity to move. We now move backward through the tree to player i's choice node. Player i knows that j will escalate on his turn. If i escalates, both players will receive the compromise outcome imposed by Nature, a_i, a_j. The value a_j is compared with the value that i receives if he concedes immediately, c_i. Since c_j > a_i, i gives in on the first move. Thus, the outcome is a win for player j and a loss for player i, (c_i, c_j). Even though player i prefers the value associated with j's concession to his own, a completely informed i knows that it is impossible to achieve this outcome given j's preferences.

The other equilibria outcomes are solved in a similar fashion. In the second equilibrium, j prefers to concede to i rather than escalate and receive the compromise outcome, b_j > a_j. On the other hand, player i has a dominant strategy to escalate. This means that irrespective of j's strategy choice, i receives a higher value from escalating than from conceding to j's demands, b_i > a_i > c_i. The outcome under these sets of preferences
is a win for i and a loss for j. The compromise outcome holds when both players prefer escalation over their concession strategy as illustrated in equilibrium four. The technical reason that different outcomes occur is due to the various combinations of the players' preference orderings. The reason that different payoff orderings can occur is that under different sets of conditions players may be more or less willing to suffer the costs of escalation.

The basic conclusion of model I is that the stronger side (the player who can best withstand the costs of escalation) wins the conflict. This conclusion directly follows from the assumption that the players have complete information. For instance, if a compromise is i's worst outcome and j's second best outcome in the game, i knows that j will escalate if given the opportunity to move. Because the costs of escalation are too onerous for i to bear, i concedes defeat immediately. Since the winner and loser of a complete information game is determined a priori, the loser always concedes defeat to the stronger side without any escalation action ever taking place.

The importance of the hypotheses derived from model I becomes clearer once the variables are operationalized. Depending on the values that these variables assume, players can be distinguished on the basis of their willingness to play and/or remain in the escalation game. Once the terms of the hypothesis are translated into "operational" language, it then becomes possible to test these theoretical expectations empirically.

Players i and j concede immediately whenever the costs that the adversary can impose on them outweigh the benefits they receive from the compromise outcome. The most important indicator of a player's ability to impose costs in a conflict is its relative power. The measurement of power can include a number of components such as a country's economic and industrial base, the size of its land mass, its population which can be used in warfare and most importantly, an actor's military capabilities (Small and Singer, 1982). The level of an actor's relative capabilities is the most direct and most
relevant indicator of its ability to impose costs on the adversary (Morgenthau, 1950; Brodie, 1959; Bueno de Mesquita, 1981; Powell, 1987; 1988; Morgan, 1990).

A number of empirical studies investigating the link between power and the likely victor in a dyadic war conclude that the militarily, stronger side almost always wins. When upsets do occur, they are quite rare (Cannizzo, 1980; Bueno de Mesquita, 1981; Maoz, 1983; Ray and Vural, 1986). In situations other than war, such as a crisis bargaining context, power is not equated with the ability to "win" the negotiations but the stronger side can use its power to squeeze concessions from the other side to obtain a better outcome (Wittman, 1979; Pillar, 1983; Morrow, 1989; Morgan, 1990; 1993).

The costs that a players' military can impose take on two general forms. On the one hand, capabilities can be used to destroy something of value that is either costly to repair or in the case where the military extinguishes a human life, the value is irreplaceable. On the other hand, particular uses of the military can prevent the opponent from securing the resources necessary to continue fighting efficiently. Here I am referring to actions such as quarantines, blockades, seizures of military equipment or cutting off supply lines that prevent the opponent from securing critical materials required to continue fighting (Gochman and Maoz, 1984; Huth, 1988).

The relative strength of a player's military represents an actor's capacity to impose costs and more importantly, the ability to sustain these costs over time. A stronger military does not necessarily guarantee that a player will prevail against a weaker adversary; however, all other things being equal, we should expect that the stronger side can hold out relatively longer (Organski and Kugler, 1980). One of the advantages that the more powerful side maintains is the ability to replenish its resources more quickly and efficiently than the weaker side (Rosen, 1972). A large military apparatus must be supported by a fairly advanced, industrialized economy. In times of crisis, a large economy can more efficiently divert resources into those industrial sectors of the economy
to sustain the military machine. An economy, whose industrial base is much smaller, may be less able to quickly replenish the hardware necessary to sustain the escalation pace in the long run.

A superior military power is also in a better position to immediately overwhelm the opponent's military forces. This is precisely what occurred during the US-Iraqi war in 1991. Even though the Iraqis had the benefit of fighting on their own territory, the military dominance of the U.S. lead coalition completely consumed the Iraqi fighting force. Even in situations where the disparity in capabilities is not as obvious, the militarily stronger state is still in a superior position to impose costs on the adversary.

The stronger side might also be able to impose greater costs on the adversary because of the possession of a particular type of technology not available to the opposing side. The most obvious example of this is the possession of nuclear weapons. Even though the Japanese demonstrated a profound willingness to suffer costs in the Pacific War against the US in the 1940s, the rapid devastation brought about in using nuclear weapons and more importantly, the inability of the Japanese to counter the move at such a rapid rate of cost imposition, led the Japanese leadership to sue for peace. The utility of owning nuclear weapons as a threat against aggressors is a question that has yet to be resolved in the scholarly literature. Nuclear weapons have not been used since August 9, 1945 and there are only a few instances in which a nuclear power threatened to use strategic weapons against a non-nuclear power, e.g., the U.S. threatened China during the Quemoy-Matsu crisis in 1958 (George et al., 1971). One of the more interesting arguments advanced to explain the restraint on the use of nuclear weapons from the point of view of escalation is that the costs involved in using a strategic weapons are more devastating than the benefit they might be fighting for. Clearly this was the logic underlying the strategic doctrine of mutually assured destruction that governed the relations between the US and Soviets during the 1950s and 1960s.
Beyond the military dimension, there are other conditions under which a state might give in to the adversary's demands. Earlier I alluded to the advantage that a superior economy can bring to nations who must resupply themselves during warfare. The vulnerability of a state's economy to external economic threats and forces might also cause a state to cease behaving in ways contrary to the wishes of a rival power with superior resources. A state whose economy is based on very few, or in some cases, a single commodity are likely to evaluate the costs of the imposition of a tariff or even more seriously, an embargo, as more devastating than a nation whose economy is stronger and more diversified. In the latter case, the costs can be spread out over several sectors of the economy such that its impact is muted. But for nations whose existence depends on the exportation of cash crops or raw materials, as is the case in many of the African nations, the reduction in demand can be quite costly particularly if the availability of alternative consumers is not present. During the latter half of the 20th century, the U.S. could threaten cutting its foreign aid program to poor third world nations who were attempting to pursue foreign policies that deviated from American desires. Since these nations' economies depended heavily on the supply of American dollars and given the difficulty of securing other sources of monies, the possibility of being cut off was quite serious.

Another attribute which may distinguish between states' ability to tolerate escalation costs is the type of domestic governing structure of the state (Rummel, 1975; Morgan and Campbell, 1992; Bueno de Mesquita and Lalman, 1992). All political leaders who make decisions on behalf of the national interest are accountable to some portion of the populace for those foreign policy choices. In democratic societies, leaders are held accountable for their actions through the voting process which either returns or removes them from power. Rulers of non-democratic nations, such as dictatorships, kingships or communist systems, must satisfy the preferences of a small group of powerful political elites or else face removal from office through a coup, assassination or some other
institutionalized procedure (Zirker, 1986). Leaders of different governing institutions suffer the same potential costs, e.g., removal from power, however, the differential ability of the societies of these nations to articulate their dissatisfaction and to impose these costs on leaders may affect the willingness of these state leaders to absorb costs. Dictatorships may be better able to sustain these costs in the long run because some military personnel and the civilians who are losing their lives or going hungry have no legal avenue to pursue in punishing decision makers for their choices. The people absorbing the costs of escalation in democratic societies can sanction their leaders for undesirable policies. This might cause democratic leaders to avoid certain types of confrontations or to extricate themselves from conflicts that are long and costly, e.g., the British in the Boer War; France in Algeria; the U.S. in Vietnam.

The compromise outcome, \((a_i, a_j)\), occurs when the costs to both players of escalating are less then receiving a partial satisfaction of their demands. In general, the greater the benefit from getting one's demand satisfied, the greater the costs that one will bear in order to achieve that demand. Conversely, the less salient the issue, the fewer the costs one will pay to achieve full satisfaction on a particular demand.

**Summary of Model I**

The basic conclusion of the simple model is that the player who can best withstand the costs of escalation relative to the value of the issue at stake wins the game. This conclusion is consistent with our intuition about what should occur under an ideal set of circumstances. If the model were to fail to produce such a hypothesis, the theory itself would become suspect. This does not imply that all model derivations should confirm the obvious. However, given the restrictive nature of the assumptions underlying the simple model, this is precisely the conclusion we want to draw when nothing is left to chance.

The assumption that both players have complete information transforms the escalation game into a mental exercise for the actors. The winner and the loser of the
conflict is known a priori and so the loser concedes defeat immediately knowing that the adversary's threat to escalate is perfectly credible. Hence, what we should observe empirically falls at one extreme of the escalation continuum: players escalate once or threaten to escalate once and quit the game.

We can illustrate the insights of the model as well as its weaknesses by considering President Johnson's decision to escalate U.S. involvement in the Vietnam War. In 1964, Johnson was informed by his advisors that the U.S. effort to prop up the Diem regime in South Vietnam was failing due to recent Viet Cong (V.C.) successes against the South Vietnamese army. American foreign policy advisors concluded that the South Vietnamese government would fall into communist hands within a few short months unless the Americans accelerated their involvement in the conflict. Johnson's choice set can be simplified into two basic courses of action. On the one hand, Johnson considered air strikes against Hanoi in North Vietnam. The potential workability of this option was based on the assumption that the costliness of the bombing campaign would force Hanoi's hand in putting pressure on the Viet Cong to back down. On the other hand, Johnson had the option of essentially doing nothing and simply letting events in the South run their course. If President Johnson had complete information about the outcome of the Vietnam War back in 1964, he would have seen that the United States would spend x dollars, lose approximately 58,000 American lives and fail to achieve its objective that South Vietnamese government remain in the hands of a non-communist leader. President Johnson, knowing the end game in 1973, would have never chosen to launch air strikes against Hanoi nor would he have escalated the conflict at any point subsequent to 1964.

The disparity between what actually occurred in the Vietnam War and the story as interpreted theoretically in model I is that the complete information assumption trivializes President Johnson's choice. A priori knowledge of the winner and loser in a game always results in the loser conceding defeat immediately to avoid suffering the costs of escalation.
The Vietnam War, the First Schleswig-Holstein War of 1849, the War of the Armada of 1585, nor any other war or even conflict for that matter can occur if both players know the outcome before the game is even played.

President Johnson ordered the air strikes because he was not certain in the mid-1960s that the U.S. would lose the War. Johnson reasoned that if the Americans escalated long enough and escalated hard enough, Hanoi and the Viet Cong would have no choice other than to succumb to the American pressure. What the Americans failed to realize (or were extremely slow to realize) was that the Vietnamese were willing to suffer salacious costs to achieve their objective of national independence. Johnson's uncertainty over who would win, therefore, was a really an uncertainty over the cost tolerance of the Viet Cong and the North Vietnamese. But it took ten years, not one theoretical escalation move, to learn that this was the case.

The Vietnam example exposes one of the conceptual problems involved in failing to model escalation beyond the conditions of complete information. The prediction that states escalate once and then disengage themselves has little, if any, empirical support. The causal connection between the theoretical conclusions and the lack of correspondence with empirical observation is that the complete information assumption trivializes an actor's decision into little more than a mental exercise. We can predict the outcome but we would never actually observe a player taking an action. Returning to the previous example, if President Johnson had known in 1964 that North and South Vietnam would be unified under communist rule approximately ten years later, Johnson would have withdrawn all U.S. support for the South. When a player knows with certainty that he will lose, there is no reason why he would play the game. But Johnson did play the game. The timing of Johnson's decisions to escalate the war in Vietnam have been linked to an optimistic belief that victory was lurking around the corner (Schlesinger, 1968) and a more pessimistic strategy which was designed to temporarily stave off a military defeat.
(Ellsberg, 1971). The more fundamental point, however, is that Johnson's escalation decisions occurred because he could not predict the outcome of the conflict with certainty. The U.S. might lose, but America could have no chance of winning without escalating and imposing costs on Hanoi and the Viet Cong.

In terms of the simple model, an uncertain Johnson (player i) does not know whether the Viet Cong (player j) prefer the compromise outcome to j's concession because i lacks the knowledge of how j evaluates the issues at stake relative to the costs necessary to achieve j's demand. The knowledge over j's preferences becomes a crucial piece of information to i if i's worst outcome is the compromise. Escalation becomes a risky strategy for i since he can obtain either his best or worst outcome depending on j's cost tolerance. In sum, players escalate because they do not know a priori which state is best able to withstand the costs of escalation. Escalation is the manifestation of the belief that one has some chance of winning the conflict. If there is no chance, there is no escalation. To capture this theoretical argument, the complete information assumption in model I is relaxed. This extension produces model II.

**Model II**

*Assumptions*

Model II is a game of one-sided incomplete information. Figure 2 illustrates the one-sided incomplete information game. Player j continues to play the game with complete information. However, player i does not know the cost tolerance of player j. To simplify the analysis, I assume that there are two types of cost tolerant j players: low cost tolerant (LCT) and high cost tolerant (HCT). These two types of actors are distinguished on the basis of their willingness to suffer costs and this affects their preference orderings over the three outcomes. A low cost tolerant j is willing to withstand the costs of one escalation move. A LCT j player prefers c_j > b_j > a_j. Since a low cost tolerant j places
FIGURE 2

A Low and High Cost Tolerant Player i

A LCT j: \( c_j > b_j > a_j \)
A HCT j: \( c_j > a_j > b_j \)
relatively less value on the issue at stake, \( j \) concedes to \( i \) to avoid paying the costs of escalation if given the opportunity to move. A high cost tolerant \( j \) can withstand two of his own escalation moves which gives, \( c_j > a_j > b_j \). A HCT \( j \) always escalates on his turn. Player \( i \) knows his own preference ordering and the ordering associated with each \( j \) type of player. Player \( i \) knows that a LCT \( j \) concedes and a HCT \( j \) escalates but does not know \( j \)'s exact type.

Player \( j \)'s type is an attribute which is determined prior to the game. Nature makes the first move in the game and randomly selects a \( j \) type from a probability distribution of cost tolerant players. Nature draws a low cost tolerant (LCT) type \( j \) with probability, \( p \) and a high cost tolerant (HCT) type with probability, \( 1-p \). Since there are only two types of \( j \) players, \( p + 1-p = 1.0 \). Following Nature's draw, \( j \) is informed of his type. Player \( i \); however, is only informed that Nature has drawn a LCT type with probability, \( p \) and a HCT type with probability with a probability of \( 1-p \). This means that \( i \) is making decisions under conditions of risk. In Figure 2, player \( i \)'s information set contains two choice nodes to reflect \( i \)'s uncertainty over whether he is playing the game against a LCT \( j \) on the upper part of the tree or whether he is operating on the lower branch against a HCT \( j \). Both of \( j \)'s information sets are singletons since \( j \) knows his own type with certainty.

Game theorists assume that a player's preference ordering remains fixed for the duration of the game which means that a \( j \) player cannot switch types during the middle of game play. Many escalation theorists have violated this assumption implicitly when they explain escalation as the result of an increase in the object's value for one or both players (Ikle, 1971; Smoke, 1977). Not only does this violation tend to produce indeterminate predictions but without a clear specification of the conditions under which the payoffs become variable, the hypotheses become non-falsifiable.

Model II is a generalization of model I whenever \( p \) assumes the values of zero (\( j \) is HCT with certainty) or one (\( j \) is LCT with certainty). A probability value of zero or one
transforms the one-sided uncertainty game into a two-sided complete information game. Player i simply selects that alternative which gives him his best outcome given the certain knowledge that j escalates (j is HCT) or concedes (j is LCT).

There are six outcomes in this game; however, two of these outcomes cannot be realized given that each j type has a specific payoff ordering. The first of these outcomes is the compromise outcome, (aj, aj), when j is LCT. By definition, a LCT j prefers bj to aj and if given the opportunity to move, always concedes. Second, a HCT j will not concede to i's demands which means that the (bj, bj) cannot occur. The different combinations of the players' preference orderings produce four outcomes.

Equilibria Outcomes

The game is solved in the same fashion as the simple model by using the subgame perfect equilibrium concept.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Payoff Orderings of the Players</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (bi, bj)</td>
<td>bj &gt; aj &gt; ci AND cj &gt; bj &gt; aj</td>
</tr>
<tr>
<td>2. (ai, aj)</td>
<td>bi &gt; aj &gt; ci AND cj &gt; aj &gt; bj</td>
</tr>
</tbody>
</table>

Whenever i prefers the compromise outcome to his own concession, i has a dominant strategy to escalate against both types of j players. When ci < ai, uncertainty does not affect i's behavior because i is better off with either outcome associated with escalating to his own capitulation. Player j, who observes that i has escalated, selects the strategy that provides j with best outcome given his type. A LCT j concedes and i wins the game. If j is HCT, the players end the dispute with a compromise outcome.

Player i's decision becomes more complex when he does not have a dominant strategy to escalate, ci > ai. In this case, the costs of escalation are too burdensome for i such that i prefers to give in to j's demands rather than escalate and possibly receive the
compromise outcome. The only way that i can get his best outcome, a concession by j, is to escalate. When j is LCT, i will indeed win the game. However, escalating also risks the possibility that i gets his worst outcome, a costly compromise, when j is HCT. The only outcome that i can get for certain is a loss when i concedes, c_i. Player i must decide whether he should escalate and risk getting a_j or b_j versus conceding and getting c_i.

In games of complete information, it is adequate to represent the players' preference orderings over the outcomes by the utility value or payoff associated with each outcome. In model II, player i must consider a gamble in which one of n outcomes can occur. Player i is now making choices between different gambles and not among alternatives that provide him with a known value. Player i's preferences over these lotteries; however, may not correspond with the expected monetary value of the gamble. I assume that player i's preferences among gambles is represented by von Neumann-Morgenstern utility functions and i chooses that act which maximizes expected utility. For simplicity, i assume that i's utility function is linear to reflect i's risk neutrality (the expected value of the game equals the expected utility).

Since the escalation option no longer provides i with a certain payoff, i calculates the expected utility for escalating and compares that with the certain value he obtains by conceding. Player i knows with probability p that j is a low cost tolerant player. If j is LCT, i will receive b_j because a LCT j concedes on his turn. There is a (1-p) probability that j is a HCT player in which case i will receive the compromise outcome, a_i.

\[(p)*(b_j) + (1-p)*(a_i) = \text{the expected value for escalating} = 3.0\]

When c_i > a_i, player i compares the expected utility of escalation with the certain value i obtains by conceding to j's demands.

\[(p)*(b_j) + (1-p)*(a_i) = c_i = 3.1\]

When the expected utility for escalating equals the value i receives by conceding, i is indifferent between choosing either of his two strategies. If the left side of the equation
exceeds the utility of conceding, i escalates. If \( c_j \) is greater than the expected utility of escalation, i concedes immediately. By isolating the term \( p \), the probability that \( j \) is LCT, we can derive a threshold value that makes i indifferent between escalating and not escalating.\(^4\)

\[
p = \frac{(c_i - a_i)}{(b_j - a_j)}
\]

where;
\( p \) = probability that \( j \) is LCT
\( (c_i - a_i) \) = The difference in value between a compromise and losing
\( (b_j - a_j) \) = The difference in value between winning and losing

I offer an example to illustrate how this equation sheds light on identifying the conditions under which a state will escalate a conflict. Assume that the compromise outcome has a value of 1; \( j \)'s concession has a value of 5 and i's concession has a value of 3. The threshold value for \( p = .50 \). When there is an even chance that \( j \) is LCT or HCT, i is indifferent between escalating and conceding. If the probability that \( j \) is LCT is less than 50%, i concedes. By manipulating the values for a, b and c, we can derive several hypotheses identifying the conditions under which i escalates in a conflict.

*Hypotheses*

(H1) i is more likely to escalate when \( a_j \) increases, all other things being equal.

The payoff \( a_j \) represents the value of the compromise outcome. The value of the compromise outcome increases because the escalation costs that \( j \) can impose on i are not as serious and/or the issue at stake has increased in value to i. The expectation is that i is more likely to escalate on the first move. The costliness of escalation declines when i competes against a j player with a different set of attributes. For instance, the military strength of j might decline relative to i which places j in an inferior position to impose the

\(^4\) See Appendix 1 for the mathematical derivation of this equation.
costs necessary to get a stronger i player to back down. Even if i places the same value on the compromise outcome when i was relatively weak versus strong, the inability of j to sustain the imposition of serious escalation costs increases i's net benefit for the compromise outcome.

When costs are held constant, \( a_{ij} \) increases when i places more value on a partial fulfillment of his demands relative to giving in. One of the state goals of American foreign policy since 1945 is the promotion of democratic rule in smaller, third world countries (more often than not the translation from the theoretical to the practical meant the installation of a non-communist government not a democratic one). Even so, the American government has been more willing to intervene directly with military force placing its own military personnel in jeopardy in regions like the Caribbean and Latin America but prefers to provide indirect military assistance for indigenous fighting groups in African countries. The basic reason is that the U.S. places a higher value on 'stability' for its neighbors and is willing to pay a higher price to ensure this outcome relative to American behavior toward African states given the proximity of Latin America to the U.S. homeland. A potentially aggressive Latin American state is more of a potential threat to American security than a state thousands of miles from American shores.

(H2) i is more likely to escalate when \( b_{ij} \) increases.

This hypothesis states that i is more willing to suffer escalation costs when i places a greater value on j backing down from his demands (which implies of course that i places relatively less value of his own concession).

(H3) i is less likely to escalate when \( c_{ij} \) increases.

The payoff \( c_{ij} \) represents the value i derives from conceding immediately. The parameter values in this equation change when i's assessment of j's capacity to inflict escalation costs on i vary. Player's j's capacity to inflict damage can again be viewed in
terms of its relative military power. The greater the relative capabilities of a j player versus some i, the greater the costs that j can impose on i. As a result, we should observe far fewer conflicts between two states of relatively disparate power since many of these conflict will end with an i player conceding immediately to j's demands. The problem, however, is that we are often not aware of a conflict until a participant has taken some overt action against another. This implies that when we do observe actors with disparate capabilities involved in a conflict, that a militarily weaker side who initiates the conflict, places a greater value on the issue at stake to the extent that they are willing to absorb the punishment that a stronger adversary can mete out.

**Summary of Model II**

The one-sided incomplete information model produced several hypotheses about the conditions under which states are expected to escalate a conflict. One of the most important conclusions of model II is that it qualifies the prediction of the simple model. Formerly, the player that could best withstand the costs of escalation won the conflict. However, once uncertainty is introduced into the model, that conclusion holds only under a more restrictive set of conditions. These conditions are dependent on the beliefs that the uninformed player has about the opponent's willingness to suffer costs. If j is truly a LCT player but i believes that the probability that j is HCT is sufficient to make the expected value for escalating worse than giving in, i concedes. In this instance, player j, who is least able to tolerate escalation, wins the game. Player i, who would be the predicted winner in the complete information game, winds up the loser because of his beliefs.

Model II also raises the possibility that a player can achieve his worst possible outcome. This was precluded in model I because all players could at least guarantee themselves their second best outcome. The uninformed player pays a price for underestimating the cost tolerance of his opponent. This will occur most often when a
stronger nation state confronts a weaker adversary. In either case, i must pick up cues from his environment to assess the type of j player that he confronts. Whenever these cues are tangible, i can gather better information about j's tolerance.

Incorporating the assumption of incomplete information is an important extension of the previous model because it allows for the theoretical possibility that a state can rationally choose to escalate a conflict and receive a bad outcome. The assumption of complete information guaranteed that a player would do no worse than his second best outcome. The inability of an incompletely informed player i to predict the outcome of a conflict with complete accuracy means that under certain specified conditions players will accept the risks that escalation will either lead to their best or worst outcome. The incomplete information assumption also presents a more realistic portrayal of the complexity involved in making decisions during the bargaining process. Player i's decision to escalate or not is trivial if the outcome simply depended on his evaluation of the game exclusively. However, the interdependence between the players requires that i pay particular attention to the type of adversary he confronts. A player's type provides information with respect to that player's willingness to escalate a conflict and this effects the behavior of an uninformed player who would prefer to avoid the costs of escalation.

Returning to the Vietnam War example, the prediction from the complete information model was that Johnson would have conceded defeat immediately and never escalated American involvement in Vietnam. This prediction, of course, is inaccurate because Johnson had incomplete information about the North Vietnamese/Viet Cong's type. Johnson did not know in 1964 whether America would win or lose to North Vietnam; however, the President believed that there was a fair chance that the U.S. could prevail. The probability of outlasting the opponent in a game of escalation is a subjective evaluation about the magnitude of the opponent's cost tolerance or his willingness and ability to suffer the costs of escalation. It seems reasonable with hindsight to conclude
that Johnson initially believed that the American cost tolerance far surpassed that of the Viet Cong and Hanoi. The relatively low probability Johnson assigned to a higher cost tolerant Viet Cong in the South and the communists in the North was most likely a function of the superiority of American military capabilities. Johnson may have also taken into consideration the fact that the Vietnamese people had been fighting and absorbing massive costs unrelentlessly since 1954 and may have been growing weary. All of Johnson's escalation decisions occurred because of his beliefs about the willingness of the Viet Cong and Hanoi to continue with the escalation game. So why then did the Americans eventually withdraw from Vietnam in 1973?

This is an important question but one that cannot be addressed by model II. The basic reason is that while the assumption of incomplete information is a necessary condition for an actor to escalate, the assumption is not supported by additional axioms that make it possible for states to cease escalating. The escalation dynamic that President Johnson set into motion in 1964 would not stop until either the Viet Cong/Hanoi or the U.S. destroyed all of its resources in the process and was thereby physically incapable of continuing to play the game.

The assumption that model II lacks pertains to the ability of actors to reevaluate the wisdom of continuing the conflict after every round of escalation. Both players enter a conflict with a set of beliefs about the opponent's costs tolerance which determines on an actor's chances of prevailing in the conflict. However, there is no reason to suspect that an actor's perceived probability of winning remains constant from beginning to end. The flux in this probability is due to the fact that the adversary's behavior can provide important information about his willingness to bear costs. Every time the opponent escalates, the quality of information about that player's cost tolerance improves and this effects the probability that an actor can prevail against this adversary in the long run. Therefore, the cessation of escalatory behavior occurs because players decide that they
cannot win based on the observed behavior of their opponent and their updated beliefs about the adversary's cost tolerance.

The capacity of actors to control their escalation behavior opens up the possibility that actors can use coercion in a strategically, manipulative way throughout the bargaining process. Few players have an incentive to reveal the true nature of their situation particularly when they are bargaining from a relatively disadvantageous position. In terms of the escalation models(s), low cost tolerant players are not in a position to sustain the costs of escalation relative to high cost tolerant adversaries. However, we should not expect low cost tolerant players to concede defeat immediately. Because players lack complete information, a low cost tolerant actor may be able to convince the adversary that they are operating from a stronger bargaining position than is actually the case. The incentive for a LCT actor to escalate is the chance that the opponent backs down due to a mistaken belief that the opponent can outlast him in a grueling game of escalation.

There exist several opportunities for players to bluff their willingness to suffer costs in an uncertain decision context; however, not all of these opportunities will bring rewards. The chance that a bluff is successful also depends on the willingness of the actor who is the target of the manipulation to escalate. For example, even if President Johnson knew from the outset that the North Vietnamese and the Viet Cong would endure, Johnson could escalate in an attempt to manipulate these two groups into backing down. However, the V.C. and Hanoi's behavior would probably have been unaffected by their beliefs about the veracity of American escalation. The high cost tolerant V.C. and Northern communists were going to escalate irrespective of American behavior and Johnson's bluff would inevitably fail.

To integrate these conceptual arguments into the model, model II is extended by mapping additional assumptions into the incomplete information model. Specifically, players are assumed to incorporate new information into their preexisting beliefs and then
reassess the willingness of the opponent to escalate. The result is a set of updated beliefs which informs a player of the optimality of subsequent escalation versus the value of giving in to the opponent's demands.

**Model III**

*Assumptions*

Escalation is in part an attempt to signal information about one's cost tolerance to the opponent. A high cost tolerant player has an incentive to convey his true type in the hope that the LCT will be convinced to give in. If the LCT actor concedes, the HCT player minimizes the costs necessary to achieve his demands. The low cost tolerant player, who does not have an incentive to reveal his true type, might escalate in an effort to bluff the opposing player into believing that a LCT is really a HCT player. The behavior of the opponent provides a tangible piece of information to a player trying to assess the opponent's type. Player's use this information to update their beliefs about whether they are dealing with a genuine HCT player or a LCT player who is bluffing.

Model III is also a one sided incomplete information game and is represented in Figure 3. A few modifications are made to model II to give players the opportunity to revise their beliefs about the opponent's type in model III. First, the informational roles of i and j are switched. Now player i has complete information and j does not know with certainty whether i is LCT or HCT. Player j knows that Nature assigns i a LCT type with probability, p and a HCT type with probability, 1-p. Player j knows that a LCT i is willing to tolerate two escalation costs; one of his own and one of j's escalation moves. If j escalates, a LCT i prefers to give in rather than endure the costs of escalating a second time. A LCT i prefers $b_i > c_i > a_i > d_i$. A HCT i is willing to tolerate three escalation
FIGURE 3
One-sided
Incomplete
Information Game

LCT i: $b_i > c_i > a_i > d_i$
HCT i: $b_i > d_i > c_i > a_i$
j: $c_j > a_j > b_j > d_j$
costs and will escalate on his final move in the game. A HCT i prefers $b_i > d_i > c_i > a_i$. Player j knows the preferences associated with each i type but j does not know which type he is competing against.

The probabilities that Nature reveals to j are referred to as j’s prior estimates of i’s type. The second change in model III is that player i begins the game but also has an opportunity to move after j. If j escalates, i selects the strategy choice consistent with the preferences of i’s type. Third, there is one additional outcome in this game, $(d_i, d_j)$. This outcome represents the value of a compromise outcome minus three escalation costs to both players. Player j’s preferences are: $c_j > a_j > b_j > a_j$. Assume that j observes an escalation by i. Player j knows that a HCT i has a dominant strategy to escalate. However, j cannot be certain that an escalating player is HCT because some LCT may have incentive to bluff j into believing that an i player is really HCT. Thus, j must determine the conditional probability that a LCT i escalates. If i escalates on the first move, j revises his prior probability estimates of i’s type according to Bayes’s rule. A Bayesian updater uses the prior estimate of i’s type and the conditional probability that a player of that type escalates to calculate a posterior probability estimate that i is LCT or HCT. Player j calculates this probability according to Bayes’s rule:

$$\frac{P(\text{ple}) \cdot (p)}{P(\text{ple}) \cdot (p) + P((1-p)\text{le}) \cdot (1-p)}$$

where:
- $p$ = the prior probability that i is LCT
- $1-p$ = the prior probability that i is HCT
- $P(\text{ple})$ = the probability that a LCT escalates
- $P((1-p)\text{le})$ = the probability that a HCT escalates
- Player j knows that a HCT i has a dominant strategy to escalate; thus $P((1-p)\text{le}) = 1.0$

Bayes’s formula is a mechanism j uses to calculate the posterior probability that the escalating player that j has observed is LCT. The posterior probability that the escalating
player is HCT is 1 minus that value. Player j's posterior probabilities of i's type function in the same way as i's probability estimates in model II. These probability estimates are used to determine the expected value to j of conceding or escalating given that j has observed that i has escalated.

Solution Concept and Equilibria

The solution concept for this game is perfect Bayesian equilibrium. A perfect Bayesian equilibrium is a strategy combination and a set of beliefs such that at each node in the game, the strategies for the remainder of the game are Nash given the beliefs and strategies of the other players. Second, the beliefs at each information set are rational given the evidence appearing that far in the game (Rasmusen, 1989: 110). We are required to drop the subgame perfect equilibrium concept because the equilibrium no longer depends strictly on the strategies of the players but the strategies themselves depend on the players' beliefs. Under asymmetric information, an equilibrium is a strategy combination and a set of beliefs such that the strategies are best responses (Rasmusen, 1989).

We begin with the two simplest cases. First, whenever p assumes a value of zero or one, the game, like the game in model II, is transformed into a game of two-sided complete information. When i is LCT, i concedes immediately given that it is impossible for i to bluff a j who knows his type. If i is HCT, j concedes immediately. The remaining cases are analyzed when $0 < p < 1.0$. The second case occurs when i concedes on the first move. Player j knows that i is a LCT player and both receive the $(c_i, c_j)$ outcome. The reason that i might concede on the first move is that his complete information also includes the knowledge of j's posterior estimates of i's type. Player i concedes because he knows that j's posterior expected value for escalating exceeds the value of giving in to i's demand. Player i knows that he will lose so i concedes right away.
Assume that \( j \) has observed an escalation by \( i \). The one piece of information that \( j \) knows is that \( i \) did not give in. Thus, \( j \) uses this data to calculate a revised estimate for \( i \)'s type. The decision rule that \( j \) uses to determine whether he should escalate or give in has the same structure as \( i \)'s decision problem in model II. The only difference between the two is that \( j \)'s expected value for escalating is based on the posterior estimates of \( i \)'s type and not on \( j \)'s prior beliefs. If \( i \) is LCT and \( j \) escalates, \( j \) receives \( a_j \). If \( i \) is HCT and \( j \) escalates, \( j \)'s payoff is \( d_j \). Player \( j \) calculates the expected value of escalating as follows:

\[
\frac{P(ple)p}{P(ple)p + P((1-p)le)(1-p)} \times (aj) + 1 - \frac{P(ple)p}{P(ple)p + P((1-p)le)(1-p)} \times (dj) = (bj)
\]

where;

\( P(ple) \) = Probability that an escalating player is a LCT type. Call this \( Q \).
\( P((1-p)le) \) = Probability that an escalating player is a HCT type.

All HCT have a dominant strategy to escalate. Thus, this value = 1.0.

\( a_j \) = The value to \( j \) of \( i \)'s concession minus two escalation moves (\( i \) is LCT)
\( d_j \) = The value to \( j \) of the compromise outcome minus three escalation moves (\( i \) is HCT)
\( b_j \) = The value to \( j \) of conceding to \( i \)'s demands minus one escalation cost

Solving the equation for \( Q \) yields \( j \)'s posterior probability that an escalating player is low cost tolerant.\(^5\)

\[
Q = \frac{(1-p)(b_j - d_j)}{(p)(a_j - b_j)}
\]

Player \( j \) substitutes his prior estimate that \( i \) was LCT, \( p \), with the value \( Q \) to calculate an updated expected value for escalating versus giving in after observing \( i \)'s escalation. Even

\(^5\) See Appendix 2 for the mathematical derivation of this equation.
after updating, j does not know i's type with certainty. Therefore, if j escalates, j is uncertain as to whether i will escalate (i is HCT) or whether i will concede (i is LCT) on the last move in the game.

Player i is advantaged by the fact that he has complete information in model III. This means that i knows the strategy j will select based on j's revised estimates of i's type. In other words, a LCT i knows with certainty whether his attempt to bluff j into backing down will succeed or fail, e.g., j escalates and the bluff fails. Therefore, i wants to alternate between a strategy of bluffing and not bluffing such that after j revises his estimates of i's type, j is indifferent between his two strategy choices. Player i accomplishes this by devising a mixed strategy combination which involves escalating (bluffing) with probability, Q and conceding (not bluffing) with probability, 1-Q. If a LCT i plays this mixed strategy, all other things being equal, j will be indifferent between his two strategy choices after revising his estimates of i's type.

Note that the probability that a LCT i bluffing his type depends on j's prior estimates of i's type and j's payoffs for the outcomes in the game. In model II, i's decision to escalate or not depended strictly on i's estimates of j's type and on i's evaluation of his own payoff values in the game. All that matters to a model II player i is j's preference ordering (j's type). The magnitude of j's payoffs did not affect i's calculations so long as the preference ordering over j's outcome remained the same.

Several hypotheses identifying the conditions under which a LCT i more or less likely to bluff his type are derived by manipulating j's payoff values in equation 3.3. Depending on the direction of change in these parameter values, different j players are more or less likely to escalate in response to an escalation move by player i. The probability that some j will escalate has a proportionate affect on the probability that LCT i players' bluffs will succeed or fail. This probability in turn affects the probability that a LCT i escalates. For example, assume that a LCT i is competing against some j for whom
the costs of escalation have decreased. Because escalation is less costly to j, the net benefit of the outcome associated with j’s escalation has increased thereby increasing the probability that a j player will escalate if given the opportunity to move, certus paribus. Now j observes that i has escalated. A relative decrease in the costs of escalation for some j, causes j to revise his estimates downward that the escalating player is LCT, e.g., Q decreases. In other words, out of a population of i players that escalate against a j for whom the costs of escalation have decreased, fewer of these players will be bluffing their type because the probability that j escalates in response to i has increased. However, recall that Q represents not only a j player’s revised estimate that his opponent is bluffing his type but Q also represents the probability that a LCT i escalates in some mixed strategy combination. As Q decreases, the probability that any one LCT i player escalates against j decreases as well.

Hypotheses

(H1) A LCT i is less likely to bluff his type as j’s prior probability estimate that i is LCT increases

Hypothesis one simply states that a LCT i will have a lesser chance of convincing j that i is a HCT player when j holds strong initial beliefs that i is a LCT type. Consequently, j is more likely to escalate which increases the chance that a LCT i’s bluff fails.

If we assume that military capabilities provide an indication of a players' ability and willingness to tolerate the costs of escalation, then the hypothesis leads to the expectation that weaker actors will be less likely to escalate and initiate confrontations against stronger opponents.

(H2) A LCT i is less likely to bluff when dj increases.

The value dj represents the payoff j receives from the compromise outcome minus three escalation costs. An increase in dj indicates that some j actors are better able to both
infect and absorb the punishment of escalation. Again, as the disparity in the military capabilities between the two sides increases, the weaker actor occupies a position that makes it relatively more difficult for him to engage the stronger opponent in a costly escalation battle. The same level of coercion will be less effective against a player for whom the costs of escalation are relatively more tolerable. In other words, a stronger opponent raises the price that the weaker side must pay in order to achieve the same value in the conflict.

The tools or types of weapons that can be used to inflict damage may be equally as important as a state's overall strength in terms of lowering the costs necessary to achieve an outcome. By many accounts, the advent of the Dreadnought, a class of warship developed in Germany and England in the early 1900s, meant that the world's pre-Dreadnought fleets became practically valueless overnight (Hough, 1964; Lambelet, 1974; 1975; 1976). The Dreadnought carried higher caliber guns, had a heavier and more effective broadside and could maintain a high speed for relatively longer periods of time (Lambelet, 1974). This meant that any state who attempted to challenge a German and especially a British Dreadnought on the high seas had to be prepared for a costly and most likely, losing engagement.

(H3) A LCT i is less likely to bluff when bj decreases.

The payoff bj is the value j places on his own concession to i's demands. The more unreasonable i’s demand is from j’s perspective, the less value j gains from giving in. Thus, j will be more likely to escalate and any attempt to bluff j into backing down is more likely to fail, all other things being equal.

Summary of Model III

As model I demonstrates, players who have complete information can avoid conflict without any escalation actions ever taking place. The reason that we observe escalation empirically is due to the fact that players have incomplete information regarding
the opponent's relative willingness to suffer costs in the escalation game. The escalation process involves understanding both the conditions under which states escalate as well as the conditions under which states stop escalating. States escalate because they believe that there is a sufficient probability that they have a higher cost tolerance than the adversary to justify extending the escalation game. By definition, states cease escalating once they decide that they are unwilling to suffer the costs necessary to win the game. The ability of players to change their behavior is made possible by the assumption that players are Bayesian updaters who can learn about the opponent's relative cost tolerance by observing the opponent's willingness to bear escalation costs throughout the course of the conflict. The players' learning process involves refining and improving their estimates of the opponent's type. Any qualitative improvement in the information a player receives about the opponent's cost tolerance is used as a basis for reassessing the potential benefits to be had from remaining in the escalation game. This aspect of the game is an important theoretical improvement over many of the current models of escalation that rely on implicit or ad hoc assumptions to explain the conditions under which actors will exit from the escalation game.

As noted previously, one of the difficulties in explaining American escalatory behavior in Vietnam with model II was that the incomplete information assumption by itself could not explain why American decision makers began to de-escalate involvement in Vietnam around 1969. By assuming that players can learn about the opponent's cost tolerance, model III helps to explain why the Americans began to gradually reduce their involvement in Southeast Asia.

In 1964, the American leadership held the belief that the U.S. could win the war in Vietnam with a limited amount of force applied through the South Vietnamese Army. American optimism regarding the success of this military strategy was intertwined with the belief that neither Hanoi nor the Viet Cong possessed a greater willingness to suffer costs
to achieve their aims than the American backed South Vietnamese forces. By the end of 1964, intelligence reports began to circulate in Washington which indicated that the South Vietnamese Army was not performing as well as expected against the Viet Cong. However, this did not mean that President Johnson abandoned his belief that military victory was still a viable outcome for the American side. What these reports suggested to Johnson and his advisors was that victory would elude the South unless the American intervened in the conflict directly. In 1964, Johnson ordered the American Air Force to begin a bombing campaign against Hanoi in North Vietnam. This campaign took effect in 1965. Simultaneously, American Air personnel increased their activity in South Vietnam in an effort to flush out the Viet Cong. By 1967, America had dropped more bombs on Vietnam than on the entire European theatre in World War II (Ambrose, 1985). Despite the American escalation efforts from 1965 to 1967, Hanoi managed to increase their level of men and supplies devoted to the war effort and the Viet Cong continued to fight in the South. Over the course of those two years; however, the Viet Cong did not engage in any military activity that signaled to Johnson that his belief concerning the V.C.'s cost tolerance was in need of serious revision. It was generally agreed within American foreign policy circles that everything that was required for military victory was being done.

Then in February of 1968, the Viet Cong launched the Tet offensive. Even though the Americans could claim that it won the battle, the real import of the Viet Cong offensive was the effect on the heretofore unshakable American belief that the V.C. were unwilling to pay greater costs to achieve victory than the Americans. In the aftermath of Tet, the American leadership learned two things: about their involvement in Vietnam. First, the Americans abandoned their belief that the strategy of limited application of

---

6 The total number of U.S. and V.C. military personnel killed in action in February of 1968 was 41,991. Of that total, the Viet Cong lost 39,867 men or 95% of the total killed in action. American losses in February of 1968; however, almost doubled from the previous month (Gartner, 1992).
military force was working against a resilient Viet Cong (Zagare, 1977). Second, and more importantly, Johnson decided that the American leadership and citizenry were unwilling to pay the escalation costs necessary to implement a winnable military campaign against a higher cost tolerant Viet Cong and North Vietnam. By 1969, President Nixon issued the first declaration of American intent to gradually withdraw American troops from Southeast Asia.\(^7\)

Model III also captures some of the important theoretical uses of escalation beyond its function as a means to reach an outcome. Escalation is used to signal one's truthful or fallacious intent to withstand the costs of punishment throughout the conflict. Players are ill-advised, however, to bluff at every opportunity which presents itself. Players must be careful in discriminating between those occasions when manipulating the adversary will succeed or fail given that certain types of adversaries have a greater likelihood of escalating regardless of what the opponent does. A player with a royal flush is unlikely to fold no matter how the opponent antes; the same player with a pair of tens might be persuaded to give in. Even though the U.S. had dropped more bombs on Vietnam than all of history combined (Ambrose, 1985), the U.S. eventually learned that they could not win the war at an acceptable cost against a higher cost tolerant Viet Cong.

One of the motivations of a low cost tolerant i players is to maximize the chance that a bluff will succeed. In all of the models discussed so far, the ability of i to accomplish this task is limited due to the nature of i's two strategy choices, escalate or concede. Player i can escalate which increases j's posterior probability that i is HCT but i has no ability to affect the magnitude of that increase. Providing the players with the blunt choice of escalating or not is an appropriate assumption to use when the research question pertains to the conditions under which states escalate in the conflict. However, once the

\(^7\) It should be noted that the Americans did not cease imposing costs on Hanoi or the Viet Cong after 1969. President Nixon's policy of "Vietnamization" was a strategy designed to continue the application of military pressure against Hanoi and the V.C. but with a reduction in American casualties.
decision to escalate has been made, states face a complex array of choices in terms of what level of escalation to employ and the optimal sequencing of these escalation moves. In other words, the level of escalation may be an important variable that states themselves can manipulate to affect the opponent's subsequent escalation behavior (Cross, 1969; Bonoma, 1975; Brockner and Rubin, 1985; Patchen, 1988). Different types of players may have varied incentives to utilize different escalation levels depending on their beliefs about the adversary. However, all players are assumed to be motivated by the same concern: given my opponent's type, what is the optimal level of escalation or the minimum price necessary to get the adversary to back down?

The willingness of a high cost tolerant player to withstand the costs of escalation also extends to a HCT player's capacity to endure high levels of escalation for longer periods of time. A low cost tolerant player can absorb high escalation but only for a limited duration. Therefore, a LCT may be able to increase the chances his bluff will succeed if he escalates at a higher level relative to lower levels of escalation which might be expected from a LCT player. The risk a LCT player accepts with this strategy is the possibility that the costs associated with a failed bluff are relatively greater thereby making a bad outcome even worse.

A high cost tolerant player may also have an incentive to play different mixes of escalation levels. One feature of both model II and model III is that a high cost tolerant player always has a dominant strategy to escalate the conflict. This holds because the high cost tolerant player can sustain greater costs over the long run and escalation is, by definition, the only way to win a conflict. Yet some wins are more valuable than others when the same outcome can be achieved at a lower cost. For example, if a player has the option of paying $5 or $10 to secure some object of value, the actor prefers to pay the lower amount because this raises the net benefit of the object by $5. In this example, the player's choice over what amount to pay is simplified by the fact that he is certain to obtain
the object with only $5. The analog in the escalation game is that a player would impose just enough costs on the opponent to make him indifferent between escalating one more time and giving in to the opponent's demands. However, the problem players face is that they are uncertain as to the exact value of the opponent's threshold which makes him indifferent. Thus, the actors can never be certain of the lowest possible cost that will secure a win. A state can attempt to maximize the probability that the adversary gives in by escalating high but this risks paying more costs than is necessary to achieve the desired outcome. Lower levels of escalation minimize the risk of overpaying; however, the opponent is also in a relatively better position to respond with his own counter-escalation (Snyder, 1972; Milburn, 1977; Lauren, 1979; Patchen, 1988). Thus one of the tradeoffs a player must consider is how to strike a balance between the twin objectives of ensuring victory but at a price that does not exceed what is necessary to win.

One important feature of model III is the assumption that the uncertain player can learn something about the opponent's willingness to endure the costs of escalation by observing the opponent's behavior in the game. The fact that the opponent has escalated provides some information about the opponent's threshold but the level of costs imposed by that actor from one escalation move to the next may dramatically alter an updating players' previous expectations concerning the escalating player's willingness to suffer costs.

One of the assumptions of the San Fransisco gold miners story was that the players were permitted to expend any level of gold on each turn. For instance, miner 1 could throw one coin in the Bay on his first turn and throw in 38 coins in on the next round. Even though different escalation combinations were possible, players had no incentive to throw anything less than the minimum of one coin in on each turn. The reason was that the miners were unable to learn anything about their opponent's level of resources based on the number of coins tossed into the Bay. In other words, throwing
four coins into the Bay on the first move is no different than tossing 1,000,000 on move two. The only thing that matters from the miner's perspective is that the game continues because at least one gold coin was tossed into the Bay.

The most successful use of coercion is coercion that is merely threatened but still manages to convince the adversary to back down from his demands out of fear of future punishment (Schelling, 1966). However, sometimes it becomes necessary for states to back up the credibility of the threat that future pain will indeed materialize by escalating today (Powell, 1989; 1990). This behavior may or not be sufficient to convince the opponent to give in. If a player is ultimately going to succeed, then that actor must also convince the opponent that he cannot possibly stay in the escalation game as long as his rival. Thus, the value to $i$ of escalating lower or higher depends of course on the type of $j$ player $i$ competes with. Again, if the Americans knew the cost tolerance of the Viet Cong, it would not make much difference whether the American government increased the number of advisors in South Vietnam by 5000 or unrelentlessly bombed Hanoi during Christmas in 1972.

The logic of this argument presumes that neither actor in the escalation game knows their opponent's situation with certainty. In model III, player $i$'s decision to bluff or not is simplified when he knows whether that bluff will succeed or not. Similarly, the choice between escalating high or low is of little consequence to a completely informed player. Thus, it is important to incorporate the assumption that neither of the actors has complete information in order to examine the impact that players' beliefs and values have on their decisions to utilize different levels of escalation.
Model IV

Assumptions

Figure 4 illustrates this escalation game in which both i and j have incomplete information. Prior to the game, Nature draws randomly from a distribution of types and selects a type for both players. Nature draws a LCT and a HCT i with probability p and 1-p respectively. Nature also draws a LCT and a HCT j with probability s and 1-s respectively. Both players are informed of their own type and know the preference orderings associated with each type. Nature reveals to i the probability with which a j type was selected and Nature informs j of the probability of each i type. For example, Nature informs j that he is a LCT player but Nature only reveals to i that there is a .4 probability that j is LCT and a .6 probability that j is HCT. Nature repeats this process with player i. Recall that the players are not concerned with the absolute value of the opponent's type. A LCT player gains little from knowing that his opponent is also LCT. The information that knows the preferences associated with each i type but j does not know which type he is competing against.

The probabilities that Nature reveals to j are referred to as j's prior estimates of i's type. The second change in model III is that player i begins the game but also has an opportunity to move after j. If j escalates, i selects the strategy choice consistent with the preferences of i's type. Third, there is one additional outcome in this game, (di, dj). This outcome represents the value of a compromise outcome minus three escalation costs to both players.

Previously, the players could choose between a strategy of escalation or concession. In order to capture the argument that the level of escalation is an important component in the escalation game, model IV players are provided with the alternatives of low and high escalation. The assumption of a low-high escalation choice loses some of the more interesting and complex features that can emerge with an assumption that the
FIGURE 4

Two-sided Incomplete Information: A LCT i vs. a LCT or HCT j
FIGURE 4
Two-sided Incomplete Information: A HCT i vs. a LCT or HCT j

(1 - P)

S

(1 - S)

(1 - P)
player's strategy space is continuous. However, we can derive some interesting conclusions with the simplified assumption while retaining the tractability of the model. Future research will extend the model by incorporating a continuous strategy space.

The difference between low and high escalation is in the costliness of the action. For example, we can compare the costliness of i's declaration that j's ambassador is a persona non grata to i's action in seizing j's entire diplomatic corps and holding them hostage. In both cases, j suffers the costs that accompany the official cessation of relations with country i. However, assuming that j places some premium on the lives of its citizens, the latter action is more costly in the sense that the lives of j's citizens may be in jeopardy. Not only does j lose any benefits from official relations with i but j also suffers the additional cost that its citizens might be harmed. This suggests that the distinction between low and high levels of escalation partly involves finding some cost that different actions have in common and then determining whether one of the action differs in degree of cost or even kind. Killing three military personnel in a border raid imposes fewer costs than launching a major assault on the border and killing 3000 soldiers.

It should be noted that "low" escalation could be rescaled to "zero" escalation in which case the strategy choices between the previous and the current model are indistinguishable. While low escalation can equal a zero level of cost imposition, conceding and low escalation are not necessarily synonymous in all cases. To put closure on the model, I do assume that if the player who has the last move in the game escalates low, this is equivalent to giving in. However, prior to the last move, low escalation simply means imposing relatively fewer costs on the opponent and oneself. Viewed from this perspective, model IV is a generalization of the models I-III.

The various combinations of players' strategy choices produce eight outcomes for every pair of cost tolerant players. This provides for a total of 32 outcomes in the game.
Even though there are eight outcomes for each pair of i and j, only four of these outcomes will ever materialize because each i type strictly prefers one outcome over another on his last move. The outcomes are still described as win, lose and compromise. However, the conditions that produce these outcomes are defined differently in model IV. First, the compromise outcome occurs whenever players select the same strategy choice for all three moves: either all high escalation or all low escalation. Any other combination of choices produces a winner and a loser in the game. If the players select the same strategy on all three moves, Nature has the last move in the game and imposes a solution on the players. Second, a win is assigned to the player who is the last player to escalate high. The following hypothetical sequence of moves illustrates how wins and losses are determined. If i escalates low, j escalates high and i escalates low, then this outcome is defined as a win for j. If i had instead opted to escalate high on the last move, i would win and j would lose.

The payoffs for each outcome for a LCT i and a LCT j and a HCT i and a LCT j are denoted by the letters a through h. The payoffs for a LCT i and a HCT j and a HCT i and a HCT j are represented by k, l, m, n, r, s, t, v. The preference orderings for each i and j type are structured in such a way that neither type has a dominant strategy to escalate high or low. All players in this game have a mixed strategy against each opponent type. Previously, HCT actors had a dominant strategy to escalate. While it remains the case that HCT actors continue to have a dominant strategy to escalate, they do not have a dominant strategy to choose either level of escalation. The incentive of a HCT player to escalate low is based on the possibility of achieving a good outcome at a cheaper price. Because HCT players are more willing to suffer the costs of escalation, they can always shift to higher escalation if low escalation proves to be insufficient to convince the adversary to back down. The success of either type of escalation strategy depends on the type of opponent a HCT player competes against: the HCT actor might escalate low if he believes his rival is
LCT but might escalate high against other HCT types. A similar logic applies to a LCT player when deciding between his optimal mix of strategies to play against different opponent types.

I assume that the payoffs in the game are common knowledge so that the players can calculate the appropriate strategy mix from the opponent’s payoffs. I further assume that a LCT and a HCT player can attach different values to the outcomes even though these outcomes may be the same conceptually, e.g., a win for a LCT i has a different value that a win for a HCT i. The different evaluations of the outcomes derives from the way in which different player types assess the costs of escalation.

Since this model is fairly complex, a brief conceptual overview of the game is provided. The discussion is separated into three sections to correspond with a LCT i’s first move, a HCT i’s first move and ends with a discussion of player j’s decision. Player i’s last move in the game is not discussed in detail since i’s final choice is determined by the preferences associated with his type.

A Low Cost Tolerant i

From player j’s perspective in model III, Q represented the revised probability that the escalation j observed was executed by a LCT i. In other words, Q was the probability that i was bluffing his type. A completely informed player i in model III had the advantage of knowing whether his attempt to bluff j would succeed or fail. From i’s perspective, Q was one part of his mixed strategy of bluffing against j. To make j indifferent between escalating and conceding following j’s revision of i’s type, a LCT i bluffeded with probability Q and gave in with probability 1-Q. The concept of bluffing had a clear meaning for both players in model III when it was common knowledge that i knew j’s preference ordering with certainty. A LCT i, knowing that he could not remain in the escalation game if j escalated, could still escalate against j on the first move. In model IV, a bluffing, LCT i ceases to have any meaning. In order to bluff, a player must know the
opponent's situation. This is not the case for an uncertain model IV i player. Instead, a LCT i's choice over the level of escalation is best viewed as an attempt to maximize the probability that j gives in by selecting different levels of escalation in the game.

Even though a LCT i's behavior is not considered conceptually to be a bluff attempt, the mechanics of player i's decision to escalate high or low is based on the same logic a model III i player uses to determine the optimal mixed strategy against j. Instead of calculating a single mixed strategy, i must figure out two mixed strategies, one for each type of j player. However, i does not know which of the two mixed strategy to employ since j's type is not known with certainty. Once i determines the mixed strategies to be used against a LCT j and a HCT j, i calculates an overall mixed strategy which randomizes between the two mixed strategies to be used against both j types. Player i escalates high with probability, x and escalates low with 1-x. This overall mixed strategy makes both a LCT and a HCT j indifferent between escalating low or high on his move in the game.

**Solution Concept and Equilibria Outcomes**

The perfect Bayesian solution concept is used to identify the equilibria outcomes. A low cost tolerant i randomizes over his mixed strategies according to the following:

\[
p^s((x)^*(b) + (1-x)^*(f)) + p(1-s)^*((x)^*(l) + (1-x)^*(s)) = 3.4
\]

\[
p^s((x)^*(e) + (1-x)^*(g)) + p(1-s)^*((x)^*(n) + (1-x)^*(l))
\]

Isolating x, the probability that a LCT i escalates high, produces the following derivation.\(^8\)

\[
X = \frac{p((s^*(g - t - f + s) + p^*(t - s))}{p^*((s^*(b - f - l + s - e + g + n - t) + (1 - s - n + t))}
\]

\(^8\) See Appendix 3 for the mathematical derivation of this equation.
where:
\[ P = j's \text{ prior belief that } i \text{ is LCT} \]
\[ S = i's \text{ prior belief that } j \text{ is LCT} \]
\[ b = \text{ the value to a LCT } j \text{ for winning minus two escalation costs} \]
\[ f = \text{ the value to a LCT } j \text{ for winning minus one escalation cost} \]
\[ l = \text{ the value to a HCT } j \text{ for winning minus two escalation costs} \]
\[ s = \text{ the value to a HCT } j \text{ for winning minus one escalation cost} \]
\[ c = \text{ the value to a LCT } j \text{ for losing minus two escalation costs} \]
\[ g = \text{ the value to a LCT } j \text{ for losing minus one escalation cost} \]
\[ n = \text{ the value to a HCT } j \text{ for losing minus two escalation costs} \]
\[ t = \text{ the value to a HCT } j \text{ for losing minus one escalation cost} \]

The dependent variable in this equation, the probability that a LCT i escalates high, changes by manipulating the value of the independent variables on the right-hand side of the equation. The independent variables are the players' prior estimates of their opponent's type and j's payoffs for different outcomes in the game. There are ten different independent variables which produce a plethora of hypotheses identifying the conditions under which a LCT i escalates high or low. Since some of these hypotheses are fairly intuitive, I focus the discussion on those hypotheses that are of greater interest and/or are tested empirically in later chapters.

**Hypotheses**

(H1) **A LCT i is more likely to escalate high when S, the probability that j is LCT, decreases.**

The hypothesis refers to cases in which an i player is more likely to escalate high as the disparity between i and j's cost tolerance increases and i is the lower cost tolerant player. At first blush this hypothesis may seem counter-intuitive given that a LCT player is relatively less likely to ultimately prevail in the conflict and by escalating high, may simply make that loss all the more painful. However, the logic of the model exposes the reasons why this is a risk that the LCT player is likely to take.

There are two reasons why escalating low might minimize the chances that LCT i players can get their opponent to concede. First, if both players escalate and counter-escalate at lower levels, the lower cost tolerant player is bound to achieve his cost
threshold before the opponent. In the end, the LCT player has a bundle of costs but little of value to show for the effort. The second reason, which is interdependent with the first, is that a higher cost tolerant opponent is unlikely to modify his expectations that he is competing against a lower cost tolerant adversary and will therefore remain in the game. In essence, a Bayesian player j gains little additional information from small changes in i's escalation behavior. The marginal difference in i's escalation behavior may be insufficient to convince j to back down. Therefore, one of the only chances that a lesser cost tolerant player has of getting his demands satisfied is to start escalating hard right out of the gate. This maneuver may shift j's expectations to the point where j believes that the opponent who has just self-imposed some serious escalation costs may have a higher cost tolerance than j. Clearly, the risk that the LCT i assumes is that if j is unconvinced by the action, the lower cost tolerance actor i will give in shortly thereafter because he is unwilling to maintain that escalation pace.

One indicator that distinguishes between different types of cost tolerant players is their relative level of military capabilities. As noted previously, military capabilities are the primary means to inflict damage on the opponent. As the disparity between the two sides increases, the weaker side becomes increasingly less able to match the cost imposition carried out by the stronger side. Thus, if the weaker actor is to have any chance of convincing a stronger player to back down, the weaker actor must escalate to the maximum of his capacity on the first move.

(H2) A LCT i is more likely to escalate high as the expected utility of escalating low for either LCT or HCT j actors decreases.

The expected utility for escalating low could decline for either j type for several different reasons. One potential source might be the domestic political benefits j's leaders derive from "flexing the nation's strength" in the international system. This was a common argument used to explain President Reagan's various foreign policy adventures
in Grenada, Nicaragua and Libya (Ambrose, 1985). Reagan helped restore the American citizen's confidence in American military might and as a result, rewarded him with high popularity ratings and a landslide reelection victory in 1984.

A nation's relative military capabilities might also contribute to a relative decline in the expected utility of escalating low. This is particularly true if j is the weaker actor and is LCT. As mentioned previously, one of the ways that a weaker actor can improve the chance that his rival concedes, is to escalate high. This strategy increases the probability that the opponent strengthens his belief that his rival is a HCT actor. This argument suggests that two weak, LCT actors will escalate to their highest level from the outset of the conflict.

A High Cost Tolerant i

We now turn to a discussion of the behavior of a HCT i. Recall that in model III, a HCT players had a dominant strategy. Even assuming that a HCT i in model III did not know j's type with certainty, this would have no affect on his decision to escalate. In model IV, a HCT i can win by either escalating high or low on the first move. Because the HCT actor is relatively more willing to suffer the costs of escalation, the incentive for a HCT i to escalate low is the chance that he can achieve a victory at a lesser cost than would otherwise be the case. Under certain conditions, a HCT i will not be able to win by escalating low. Thus, unlike a HCT i in model III, a model IV HCT i does not have a dominant strategy and in this case, a dominant strategy to escalate either high or low. Consequently, a HCT i goes through the exact same process to find the optimal mixed strategy against a LCT and HCT j as did i's low cost tolerant counter-part. The mixed strategy that a HCT i uses against a low or high cost tolerant j is:

\[
(1-p)^s*(x^*(a) + (1-x)^*(e)) + (1-p)^s*(1-s)^*(x^*(k) + (1-x)^*(r)) = \frac{3.6}{(1-p)^s*(x^*(d) + (1-x)^*(h)) + (1-p)^s*(1-s)^*(x^*(n) + (1-x)^*(v))}
\]
Isolating $x'$, $^9$

$$p^*((s^*(v - h - r + e) + (r - v)) + s(h - v - e + r) + (v - r))$$

$$x' = \frac{3.7}{ps(e + k - r - a + d - h - n + v) + s(a - e - k + r - d - h + n - v) + p(r - k - v + n)} + (k - r - n - v)$$

The dependent variable, $x'$, is the probability that a HCT i escalates high. Again, this derivation contains ten independent variables and several hypotheses can be derived identifying the level to which a HCT i escalates.

**Hypotheses**

(H1) A HCT i is less likely to escalate high when S, the probability that j is LCT decreases.

Contrast this hypothesis with the expectation of the behavior of a LCT i. As the probability that j is HCT increases, e.g., 1-S gets larger, a HCT i is less likely to escalate high. Thus, two different i players will escalate differently against the same type of j opponent. When we couple a HCT i's behavior against a HCT j with j's updated belief that he is playing against a HCT i after observing a low level of escalation, the conclusion is that HCT players will begin a conflict with low, gradual escalation and continue to escalate in this fashion for some period of time. Both HCT players know that they can sustain the costs of escalation for longer period of time. This knowledge gives the players an incentive to escalate low to try and get a better outcome at a reduced price. The irony is that players may in the end amass more costs than if one or both players compressed those costs into a decisive escalation move. But it is also the case that the player that levies the most damage against a HCT player is another HCT player. This provides two high cost players with a different incentive to escalate low at least initially.

---

$^9$ See Appendix 4 for the mathematical derivation of this equation.
(H2) A HCT i is less likely to escalate high as the difference between the value of winning and losing to j increases.

The basic gist of this hypothesis is that the issue over which the actors are fighting has increased in value for j. Consequently, j is more likely to remain in the escalation game to ensure that he wins a more valuable prize, irrespective of the level of escalation that is occurring in the conflict.

Player j

Player j is a Bayesian updater who reassesses his beliefs about i's type after observing that i has escalated either high or low. Player j knows that both types of i actors play a mixed strategy of low and high escalation. Thus, observing a low (high)\(^{10}\) level of escalation does not automatically signal to j that i is either a LCT or HCT type. From the information provided to j through i's behavior, j revises his prior estimates of i's type and derives a set of posterior probability estimates. Both types use these revised estimates to calculate an expected value for escalating high or low. Previously, j knew that he could receive the value associated with his own concession with certainty, e.g., the game ended and j lost. In model IV, neither high nor low escalation provides j with a certain payoff.

As a result, j calculates an expected value for both strategies.

\[
\begin{align*}
\text{P(ple)*p} & \quad \text{P(ple)*p} \\
\text{P(ple)*p + P((1-p)le))*(1-p)} & \quad \text{P(ple)*p + P((1-p)le))*(1-p)} & \text{*(kj)} = \\
\text{P(ple)*p} & \quad \text{P(ple)*p} \\
\text{P(ple)*p + P((1-p)le))*(1-p)} & \quad \text{P(ple)*p + P((1-p)le))*(1-p)} & \text{*(mj)}
\end{align*}
\]

\(^{10}\) I assume that j is able to discern the difference between high and low levels of escalation. These estimates will clearly be subjective, however, j can make rough comparisons between various escalation actions.
where;
\( p \) = prior probability that \( i \) is LCT
\( (1-p) \) = prior probability that \( i \) is HCT
\( P(p|x) \) = probability that a LCT \( i \) escalates high. Call this \( w \).
\( P((1-p)|y) \) = probability that a HCT escalates high. Call this \( z \).
\( b_j \) = the value to \( j \) of winning minus two high escalation costs (\( i \) is LCT)
\( k_j \) = the value to \( j \) of the compromise outcome minus three (high) escalation costs (\( i \) is HCT)
\( d_j \) = the value to \( j \) of losing to \( i \) minus one high escalation cost
\( m_j \) = the value to \( j \) of losing minus two high escalation costs

Isolating \( w \),
\[
\frac{(m - k)}{p^*(b - k - d + m)} = 3.8
\]

The term \( w \) refers to \( j \)'s revised estimate that the player observed to have escalated high \( i \) is a LCT actor. Unlike \( Q \) in model III, \( w \) does not represent the probability that a LCT is bluffing his type. The posterior probability \( w \), informs \( j \) of the likelihood that the high escalating player has a lower or higher cost tolerance threshold than \( j \). The higher the value for \( w \), the stronger \( j \)'s incentive to escalate low to try and secure a good outcome at a cheaper price.

*Hypotheses*

(H1) A LCT \( i \) is more likely to escalate high the higher \( j \)'s prior estimate that \( i \) is LCT (\( p \) increases)

Returning to the example of military capabilities, the higher \( j \)'s initial estimate that his opponent is low cost tolerant (the disparity between the two sides is increasing), the more likely \( i \) is to escalate to the maximum of his capacity from the start of the conflict. Even though \( j \) can never be certain that \( i \) has escalated to his limit, higher escalation moves have a better chance at convincing \( j \) that the opponent is of a higher cost tolerance. The power disparity between the actors serves as a guide for the players in estimating the cost

\[11\] See Appendix 5 for the mathematical derivation of this equation.
tolerance of the opponent. However, if capabilities were the only variable that defined a player's cost tolerance, we would never observe a weaker actor escalating against a stronger opponent. The message that a weaker i is trying to communicate to a stronger j opponent is that i, despite his inferior capabilities is actually a higher cost tolerant player. Even though the weaker side lacks the ability to reciprocate the harsh punishment than can be handed down by the stronger player, this is more than compensated for by the fact that i is more willing to endure the pain to achieve his demands than is j.

Summary of Model IV

The assumptions of model IV come closest of all the models to reflecting the environment in which actual interstate escalation decisions are made. Neither player was assumed to have complete information and the players were given a choice over the levels of escalation they deemed most appropriate given their beliefs about the opponent's situation. Actors are driven to escalate because they want to achieve something of value that is worth fighting for up to a certain point. However, states also have an incentive to obtain an outcome at the lowest possible cost to themselves. This is possibly one of the reasons why different patterns of escalation are observed empirically (Pruitt, 1969; Greffinious, 1990). The chances that any one level of escalation is successful in convincing the adversary to back down depends on who the competitor might be. Since this information is never known with certainty, states alternate between various mixes of escalation as more information becomes available to them concerning the opponent's willingness to suffer the costs of escalation. Using lower levels of escalation may increase the chances that the net benefit of all potential outcomes. However, this cost saving technique may be offset by the fact that low escalation makes it possible for all players to stay in the escalation game. On the other hand, LCT players may not have the luxury of paying a lower price to have their demands satisfied. The ability of LCT players to win is a function of the opponent's belief that the former has a higher cost threshold.
One method to solidify this belief is to impose the maximum level of costs in one, decisive escalation move right from the beginning of the conflict.

**Conclusion**

Interstate conflict is resolved through a bargaining process. Representatives of states can come together in a spirit of cooperation and accommodation to facilitate an agreement. Sometimes relations are less amiable and states use the ugly side of power or the diplomacy of violence to create a bargaining atmosphere so painful that the opponent would rather extricate himself from the process than endure the punishment (Schelling, 1966). Escalation, however, is not about savagely imposing the maximal penalty on the adversary from the outset. With time and enough accumulation of costs, the outcome may resemble a punishment delivered in one massive blow but states generally operate in such a way that hostilities cease long before the conflict progresses to such a brutal stage. Imposing salacious costs on the adversary probably increases the chances of ultimately prevailing in the conflict but the incentive to use this tactic is tempered by the desire to minimize the costs necessary to achieve that outcome. This tradeoff occurs because escalation consumes resources and is costly to employ. The restraint exercised by players against their adversaries is really motivated by the desire to minimize the punishment inflicted on oneself during the process of harming the enemy. Escalation consumes resources but players do not escalate to the point of resource exhaustion. Instead, actors generate self-imposed limitations which define the maximum level of costs they will suffer in a conflict. This threshold represents the costs an actor will absorb relative to value of the issue at stake. When these costs equal the benefit that actor is fighting for, that actor is forced to exit from the escalation game.

One purpose escalation serves is to inflict pain on the opponent. Perhaps even more importantly, escalation is used to affect the opponent's future expectations concerning the costly consequences of failing to back down from his demands. Players
who have the capacity to assimilate new information and to revise their pre-existing beliefs about the opponent's willingness to escalate do not need to achieve their cost tolerance threshold before they concede to the opposing side. As soon as a player believes that he cannot outlast the opponent, the potential loser concedes immediately to avoid accumulating unnecessary escalation costs. Under certain circumstances, low cost tolerant players have an incentive to escalate to manipulate the opponent into drawing the faulty inference that he is in conflict with a high cost player who can outlast the latter in the long run. In sum, escalation is a complex bargaining tactic that inflicts costs, signals information and manipulates future expectations all with a view to making disagreement too costly for the opponent to continue.

The first, simple model of escalation was based on the assumption that the players had complete information. The model predicted that the player who could best withstand the costs of escalation would win the conflict. This prediction provided a baseline expectation that was subsequently refined once less restrictive assumptions were incorporated into the simple model. Complete information models, while important in providing a basic foundation for subsequent models, are limited in their capacity to say anything empirically meaningful about the escalation process. The basic problem is that escalation should never occur in these models. If both players are willing to absorb the costs of escalation, complete information obviates the need to escalate to find this out since states will jump immediately to the compromise outcome. Some conceptualizations of the escalation process, such as the war of attrition, predict the opposite extreme of escalation behavior—actors do not stop escalating until they have exhausted all of their resources. This conclusion hinges on the assumption that players are unable to learn anything of consequence about the opponent while playing the escalation game. This prediction is consistent with many efforts to explain escalation using psychological, explanatory variables. The hypotheses derived from psychological models of escalation are in
principle falsifiable; however, it is often extremely difficult to distinguish between the conditions under which similar emotional responses to escalation produce different behaviors and vice versa. The analysis of escalation becomes even more confusing and ad hoc when nothing remains constant in the model, e.g., payoffs change, the willingness to bear costs varies and so on.

The theory of escalation offered in this chapter represents a significant conceptual and theoretical advancement over the majority of escalation models that currently exist. The concepts, assumptions and key variables of the theory are integrated within a well specified game theoretical structure which permits the logical derivation of several empirically, testable hypotheses. The models used to analyze the process of escalation are developed over a series of stages which helps to expose the insights and the weaknesses of various modeling assumptions in addressing important research questions pertaining to the escalation dynamic. The predictions from the model began with a condition where escalation does not occur because players know the outcome to the second model where states can escalate but not quit and finally the third and fourth models where players control both the instigation and cessation of their escalatory behavior. The latter is especially important given that so many models can address one or the other escalatory action but rarely both.

The theory identifies several key variables that affect actors' decisions to escalate in a conflict. Obviously actors must pay particular attention to the costs that the adversary can impose on them thereby making disagreement a less attractive alternative. This assessment, however, is not made in isolation of the probable value that the opponent attaches to achieving his demands in the conflict. The interrelationships among these variables produced a number of the hypotheses that lead to expectations concerning the conditions under which states were expected to escalate a conflict. Other hypotheses
referred to the level to which states were expected to escalate given their beliefs about their adversaries' types.

The conditions under which each hypothesis holds is stated at a level of generality such that the hypotheses are applicable across a wide variety of escalation contexts. The usefulness of any theory must must ultimately be measured by its ability to explain and predict the empirical world. Too often, models of the escalation process are developed in such a way that the relationships are non-falsifiable or they are developed at such a high level of abstraction that operationalizing the hypotheses for purposes of testing is extremely difficult. The following two chapters are empirical tests of some of the hypotheses derived from different models produced in this chapter. These hypotheses are tested in the contexts of extended immediate deterrence and militarized interstate disputes. Once these relationships have been tested, the evidence is used to determine the theory's usefulness in explaining and predicting what has been so perplexing in the past.
Chapter IV
Escalation in Extended Deterrence Crises

Introduction

In the previous chapter, a theory of escalation was presented and several models were developed each of which addressed some aspect of interstate escalation. From these models, several empirically, falsifiable hypotheses were derived that identified the conditions under which states were expected to escalate in conflict. The theory used to generate the hypotheses was developed in such a way that the variable posited to affect the escalation dynamic were applicable across a wide variety of escalation contexts. Because the hypotheses are amenable to empirical falsification, they can be tested and used as evidence to assess the validity of the theory of escalation.

In this chapter, I present the results of an empirical test of one of the hypotheses derived from the third model. The hypothesis is tested with cases identified as extended deterrence crises (Huth and Russett, 1984; Huth, 1988). An extended deterrence crisis is defined when a potential attacker issues a threat against an ally of a defender nation who issues a counter-threat against the attacker (Morgan, 1977; Huth, 1988). The first section of this chapter restates the theory underlying the third model and identifies the hypotheses that are tested. This is followed by a presentation of the research design and an analysis of the empirical test. The last section is a discussion of the importance of the test results as it reflects on the theory of escalation in general and on the interconnections between escalation and deterrence specifically.

Theoretical Summation

Model III is a one-sided incomplete information game. Player j, who moves after i's initial choice, is the uncertain player in this game. Player j knows his own preference ordering over the outcomes but does not know whether his opponent is a low cost tolerant
(LCT) or a high cost tolerant (HCT) player. However, the absolute value of i's threshold, e.g., high or low, is of less concern to j than knowing i's cost tolerance relative to j's willingness to suffer costs. This factor is important because i has the last move in the game. Player j knows that on the last move an i with a relatively higher cost tolerance will escalate whereas an i with a lower threshold will give in on the last move. If j has complete information, his decision between escalating and conceding is trivial. Player j concedes immediately if i's threshold is higher than j's and j escalates against a lower cost tolerant i knowing that this type of i player will give in to j's demands on the last move. However, j does not know i's type in model III and must determine which strategy to select based on his estimates of i's cost tolerance relative to his own.

One of the assumptions of model III is that j is a Bayesian updater. After observing the strategy choice of the opponent, j revises his prior estimate of i's type in light of the new information conveyed to j through i's behavior. This learning mechanism is an important feature of the model because it makes the conditions that determine the cessation or continuation of the escalation dynamic endogenous. These conditions are j's beliefs about i's type which in turn determine the best strategy for j given the behavior of the opponent.

Player j's decision problem is simple if i concedes on the first move. Player j knows that only low cost tolerant i players give in immediately. Since i's move terminates the game, j receives the value associated with i's concession without incurring any escalation costs. Unlike an immediate concession, however, if i escalates on the first move, i's behavior does not automatically reveal i's true type to player j.

Player j knows that high cost tolerant i actors always escalate. All other things being equal, the fact that i escalates increases the probability that i is HCT. However, this only increases the chances that i is a HCT type; it does not guarantee that this is the case because some low cost tolerant i players may take advantage of j's uncertainty and escalate
in an effort to bluff \( j \) into backing down. The payoff to \( i \) of a successful bluff is the difference in the value between the payoff when \( i \) concedes and the value \( i \) gets when \( j \) concedes. In order to guard against the possibility that the escalating player is a low cost tolerant bluffer, \( j \) does not automatically concede after observing an escalation move. Instead, a Bayesian \( j \) player revises his prior estimate of \( i \)'s type and then calculates an expected value for escalating versus giving in using the posterior probability estimates of \( i \)'s type. Player \( j \) has an expected value for escalating since \( j \) does not know with certainty the outcome that will occur if he escalates. Because \( i \) has the final move in the game, \( j \) could receive his second best outcome if he correctly called a low cost tolerant player's bluff but if \( i \) is truly a high cost tolerant player, \( j \) receives his worst outcome by escalating.

Player \( j \)'s choice between escalating and giving in depends on his payoffs and prior and updated beliefs about \( i \)'s type. From \( i \)'s perspective, the choice between escalating and conceding immediately is simple if \( i \) is HCT: always escalate. However, a LCT \( i \)'s choice is more complex. LCT \( i \) players do not always want to escalate because \( j \) players will be unable to distinguish HCT \( i \) players from LCT \( i \) players and as a result, \( j \) player will on average, call his opponent's bluff more often by escalating. Thus, LCT \( i \) players want to escalate some of the time but concede at others to prevent \( j \) from figuring out a LCT \( i \)'s strategy. This raises the question as to when a LCT \( i \) is more or less likely to bluff his type against an uncertain adversary. The following equation derived from model III, specifies the probability with which a LCT \( i \) bluffs as a function of \( j \)'s prior estimates of \( i \)'s type and \( j \)'s payoffs for different outcomes in the game.

\[
Q = \frac{(1-p) \times (bj - dj)}{(p) \times (aj - dj)}
\]
where;
\[ p = j's\ prior\ probability\ estimate\ that\ i\ is\ LCT \]
\[ (1-p) = j's\ prior\ estimate\ that\ i\ is\ HCT \]
\[ b_j = \text{the\ value\ to\ } j\ \text{of\ giving\ in\ to\ } i's\ demands\ \text{minus\ one\ escalation\ cost} \]
\[ d_j = \text{the\ value\ to\ } j\ \text{of\ the\ compromise\ outcome\ \text{minus\ three\ escalation\ costs}} \]
\[ a_j = \text{the\ value\ to\ } j\ \text{of\ } i's\ concession\ \text{minus\ two\ escalation\ costs} \]

From player j's perspective, Q represents the revised probability that the player who escalated is also a low cost tolerant type. In other words, Q represents the probability that a LCT player is bluffing his type by escalating against j. Because i has complete information, a LCT i can determine whether his attempt to bluff j into giving in will succeed or fail. Thus, a LCT i plays a mixed strategy that randomizes between escalating and not escalating such that j is indifferent between his two strategy choices after observing an escalation by i, e.g., the expected value to j of escalating and not escalating are equal. To make j indifferent, a low cost tolerant i escalates with probability Q and gives in with probability 1-Q.

Note that the probability that a low cost tolerant i bluffs his type depends on i's estimate of j's priors and payoffs in the game. The values of j's parameters determine the likelihood that j escalates after having observed an escalation by i. Player j's probability of escalation directly determines whether a LCT i's bluff succeeds or fails. Since the success of i's bluffing strategy is a function of j's values in the game, the probability that a LCT i escalates either increases or decreases as a function of manipulating j's values. This procedure produces several hypotheses that identify the conditions under which a LCT i is more or less likely to escalate, i.e., bluff his type.

Two different hypotheses are derived from equation 3.3 both of which lead to the same expectation regarding a LCT i' behavior:
"A LCT i is less likely to escalate as the costs to j for escalating decrease"

AND

"A LCT i is less likely to escalate as the value j receives from i's concession increase"

These hypotheses are the result of manipulating two different variables in equation 3.3: (1) decreasing j's cost for escalating (dj) while holding all other variables constant and (2) increasing j's benefit from i's concession (aj). The direction of change in the value for each of these variables produces similar expectations regarding a LCT i's behavior: a LCT i is less likely to escalate. For purposes of empirical testing, these hypotheses are tested jointly rather than separately. Both the costs of escalation and the value of the issue at stake are important theoretically in a deterrence context (Schelling, 1960; Snyder, 1961; Kahn, 1965; Huth and Russett, 1984). Moreover, both of these elements are important in defining a player's cost tolerance. The maximum costs that a player is willing to suffer is a direct function of the value that an actor derives from having his demands satisfied. LCT attackers, who are in a conflict with defenders who are better able and more willing to pay the costs of escalation, are less likely to escalate further because they know that the bluff has a greater chance of failing.

*The Extended Deterrence Context*

A great deal of the theoretical work on deterrence concentrates on identifying the conditions under which deterrence is successful (Jervis, 1979; Brams, 1985; Brams and Kilgour, 1987; Zagare, 1987; Kilgour and Zagare, 1991). Similarly, most empirical tests of deterrence and extended deterrence theory are designed to explain the outcome of deterrence (Russett, 1963; Huth and Russett, 1984; 1988; 1993; Huth, 1988; Karsten et al., 1988; Wu, 1991). Several of the theoretical models of deterrence are based on the assumption that the players have complete information (e.g., Brams, 1985). This
assumption is often accompanied by an automatic action-reaction conceptualization of escalation in which the first escalation move leads inevitably to counter-escalation until disaster strikes both of the players. However, once we relax the assumption that both players have complete information and incorporate some mechanism that gives a player the opportunity to learn about the opponent's situation through their escalation behavior, the theoretical link between escalation and the success or failure of deterrence becomes a conditional proposition which is established during the process of conflict resolution rather than a foregone conclusion at the outset of the crisis (Lichbach, 1987; Nalebuff, 1992).

Deterrence is conceptualized as a bargaining process (Wagner, 1982). A deterrent threat is viewed as one type of negative sanction that promises grave damage if the attacker refuses to alter his behavior. Escalation becomes intertwined with the deterrent threat once a state deems it necessary to back up the credibility of that threat with the application of force when confronted with a defiant attacker (Powell, 1987; 1988). When the emphasis shifts to the process of establishing deterrence within a crisis, the outcomes of deterrence success and failure simply reflect a summation of the escalatory behavior of the participants throughout the conflict. Deterrence success implies that the attacker backed down from his demands rather than opting to escalate to higher levels against the defender. Deterrence failure implies the converse of that behavior (Langlois, 1989; Powell, 1990; Nalebuff, 1991). Discussions of deterrence theory often emphasize the dilemma confronting the defender in terms of whether or not to follow through on a threat to punish the attacker. However, what is often neglected in analyses of deterrence is that following the issuance of a threat and counter-threat, the momentum in the conflict shifts back to the attacker (for an exception, see Wagner, 1982). Therefore, it is the attacker's decision to heed the defender's warning or to ignore the counter-threat and continue to escalate which partially determines the course of the conflict and the ultimate outcome of
deterrence (George and Smoke, 1974; Jervis, 1979). Thus, the theoretical inquiry centers around the conditions that prevent the attacker from proceeding to higher levels of escalation (Brodie, 1959).

The hypothesis stated above is directly relevant to the context of extended deterrence. Prior to every game, both i and j make demands on the opposing player similar to the threat and counter-threat issued by the attacker and the defender, respectively. In cases of extended deterrence, the issue in the dispute is which of the two actors is going to have influence over the protege. The protege is not a distinct actor who makes choices in this context; the protege is the very issue that brings the primary actors into conflict. As soon as both sides have issued their threats, it is the attacker, player i, who must decide whether to disbelieve the defender's threat of retaliation and press on or whether to take the defender seriously and back down from his demands.

The decision problem from the defender's perspective is whether or not he should issue a counter-threat and escalate given that the defender is uncertain as to whether the attacker is bluffing or whether the attacker has a higher cost tolerance and is prepared to play a longer, more costly game of escalation than the defender. In cases where the defender believes that the attacker's threat to escalate is genuine, e.g., a HCT attacker, the defender concedes immediately. Whenever the defender believes at some level that the attacker is bluffing his type, the defender further escalates in the conflict.

From the point of view of the attacker, a HCT i always escalates irrespective of the defender's behavior. A LCT attacker, however, escalates some of the time and gives in at others depending on his estimate of the defender's willingness and ability to suffer costs. This willingness to suffer directly affects the probability that a LCT attacker's bluff will succeed or fail. The less likely a bluff is to succeed, the less likely it is that a LCT attacker escalates to higher levels. The probability that a LCT attacker escalates is a complex mixture of the LCT attacker's estimates of the defender's values in the game, the
defender's beliefs about the type of attacker who issues the threat and both sides beliefs regarding how the opponent perceives the situation. In particular, the hypothesis states that an attacker's bluff is a function of the costs suffered by a defender in absorbing and imposing escalation costs and the value that the defender attaches to the issue at stake, e.g., maintaining his relationship with the protege.

*Operational Hypothesis: Costs*

One of the variables that affects a LCT i's willingness to bluff is j's ability to both impose and absorb the costs of escalation, dj. An indicator that i players use to gauge j's ability to suffer costs is the relative military capabilities between the two sides (Snyder, 1961; Schelling, 1966; Huth, 1988). Player i is concerned with the relative capability ratio since the critical calculation for a player is whether a state can inflict more damage on the opponent than they would suffer in return (Russett, 1963).

There are two elements of j's military power that affect his ability to tolerate the costs of escalation. The first factor is the overall size of j's military forces relative to i's. First, stronger states are in a better position to overwhelm the enemy's force simply in terms of sheer numbers. The ability of the weaker side to score a decisive victory often hinges on the element of surprise which becomes very difficult in a crisis atmosphere (Huth, 1988). The stronger side also has the advantage of being able to replenish supplies in the long run given the high correspondence between industrial capabilities and military strength (Organski and Kugler, 1980). All other things being equal, as the costs suffered by some j actor (the defender) decline relative to i, the net benefit of the compromise outcome, dj, increases for that actor. Because j is better able to absorb the costs of escalation, fewer i players are likely to escalate to higher levels against a j for whom the costs of escalation have decreased. The reason is that this j player is more likely to escalate thereby increasing the chance that a LCT i's bluff will fail.
A LCT i also gains information about the value for dj by assessing the quality of j's weapons in addition to the quantity. In particular, j actors who possess nuclear weapons may threaten their use either directly or indirectly to prevent an i player from pressing forward. It may be the case that a direct threat to use nuclear weapons is incredible and unlikely to suppress a LCT i's escalation behavior (Organski and Kugler, 1980). However, even if the attacker perceives that the probability of a nuclear sanction is extremely low, the ability of the defender to exploit the fear that "things will spin out of control" (Schelling, 1966) may be sufficient to inhibit the attacker from increasing the risk that disaster could strike (Powell, 1989). The effect on the value of dj to j players who own a nuclear arsenal is similar to cases in where j has a superiority in conventional forces. When these j actors observe that the opponent has escalated, j's revised probability that that escalating player is also LCT, Q, decreases. In other words, of the attackers that do escalate against a defender for whom the costs of escalation are relatively lower, fewer of these attackers are LCT relative to the case where the costs to the defender for escalating are higher. The result is that fewer LCT actors will escalate to higher levels given that the defender is more likely to call their bluff.

*The Value of the Issue at Stake*

The likelihood that a LCT i bluffs his type is also a function of the value that j attaches to the issue at stake, aj. The greater the value, the greater the probability that j will remain in the escalation game and force his opponent to give in to his demands on that issue. In the case of an extended deterrence crisis, the issue is which of the two actors is going to gain or maintain influence over the protege.

A LCT attacker's ability to assess the costliness of escalation is enhanced by the fact that the attacker can observe the defender's capabilities and can make a rough comparison of the ratio of military power used to impose costs. The ability to determine the value that some defender attaches to his protege is more complex given that this
estimation requires the attacker to access the defender's mind to some extent. If the argument is correct that observable indices provide the attacker with higher quality information in terms of gauging the likelihood of a successful bluff, then the same logic should apply for a LCT attacker who is assessing the value of the relationship between the defender and his ally. As Schelling (1960) points out, the credibility of a defender's threat to retaliate is enhanced by actions that the defender takes which increase the loss the defender suffers for failing to act. In other words, LCT attackers who can observe the loss the defender suffers for failing to escalate are more likely to be deterred from escalating to higher levels than LCT attackers who cannot gauge this value independent of the defender's declaration that the protege is of great value.

One indicator of the value of the protege to the defender is the level of investment, financial or otherwise, that the defender has contributed to maintaining its relationship to the protege. The rewards from any kind of investment are evaluated primarily in terms of the flow of current and future benefits. When this investment is terminated abruptly, not only does the defender lose sunk costs but also suffers the deprivation of future gains.

One of the strongest indications of the protege's value to the defender is the degree of military interdependence between the two sides (Schelling, 1960; Tillema and Van Wingen, 1982; Huth and Russett, 1984; 1988). The value of one state to another, either because of strategic location or some other factor, is best communicated not through a pledge to defend but by taking actions to create military interdependence that potential aggressors can see as well (Schelling, 1960). In particular, the maintenance of an effective fighting force requires expending vast amounts of financial resources to sustain these combat troops. This cost is compounded when material and troops must be transported to a protege located at some distance from the defender. The presence of troops also signals the defender's willingness to risk the costs associated with losing some of its own military personnel should hostilities ensue.
This argument suggests that potential signals of value which are not supported by observable indicators of the defender's loss should have no affect on the attacker's decision to press on (Morrow, 1989). Non-tangible indicators, especially formal alliance commitments, are ineffective tools in deterring escalation because they provide little useful information to the attacker who cannot discern what costs the opponent suffers or the value the opponent attaches to such an agreement. Therefore, we should expect no relationship between the attacker's willingness to escalate and an alliance between the defender and protege independent of other tangible signals of the value of the issue at stake for the defender, aj.

As Russett (1963) notes, "simply making an explicit promise to defend an ally, whether by treaty or declaration is not sufficient to deter an attacker." Part of the claim rests on the argument that oaths are often forgotten in times of crisis and there is some empirical evidence to support this proposition (Sabrosky, 1980). Of 190 commitments made between states from 1816-1965, the militarily stronger alliance partner abstained from entering the conflict on behalf of its ally approximately 66% of the time. This figure is even higher for major-minor disputes which comprise the bulk of the cases in extended deterrence crises (Huth, 1988). The argument here is that from the point of view of an attacker, an alliance between a protege and its defender is an unreliable signal of the defender's value given the ambiguous meaning of that agreement. Because the LCT attacker cannot determine the value of aj from an alliance alone, this should have no bearing on an attacker's decision to escalate to higher levels against the defender and protege.

The hypothesis leads to the expectation that attackers who confront defenders who are better able to withstand the costs of escalation will be less likely to bluff their type knowing that the chance of the bluff actually succeeding will decrease. In particular, attackers will look toward the relative capability balance between the two sides to
determine the defender’s ability to absorb the costs of escalation and the value for \( d_j \). Moreover, attackers are also expected to refrain from escalating when they are able to discern that they are competing against a defender who places a greater value on the issue at stake, the protege, \( a_j \). Both the costs of escalation and the relative value of the issue at stake to the defender are important elements in a LCT attacker’s decision to bluff his type and escalate to higher levels in the conflict. These two variables are combined into a single hypothesis for purposes of testing this expectation empirically.

**Research Design**

The hypothesis is tested using Paul K. Huth’s 1988 data on extended deterrence crises.\(^1\) The temporal domain of the data set extends from 1885-1983 and there are 58 cases of extended deterrence crises; 34 of the cases happened prior to World War II. World War I was divided into four separate cases of extended deterrence. The defender and protege were combined to form a single side in the crisis and the attacker represents the opposing side. The majority of the cases involved major power attackers and defenders while the protege was usually a minor power.

The dependent variable of the hypothesis is the probability that a LCT attacker escalates, \( Q \). There is no variable in the Huth (1988) data set that refers to the escalation behavior of the attacker specifically. However, there is a variable which provides a reasonable approximation of the escalation concept called the outcome of an extended deterrence crisis. The outcome of a deterrence crisis is classified as either a success or failure. Both of these terms summarize the attacker’s level of escalation prior to the resolution of the crisis. Deterrence success implies that the attacker disengaged from the

---

\(^1\) Four criteria are used to define cases of extended deterrence:

(1) There was an overt threat of force by the attacker against a protege of the defender. A threat was defined as an threat, display or use of force according to the criteria established by Gochman and Maoz, in the Militarized Interstate Dispute Data Set (1984).

(2) The attacker’s threat was perceived by the defender.

(3) The defender made a threat of retaliation.

(4) The attacker perceived the defender’s counter-threat.
conflict without escalating to higher levels. Deterrence failure means that the attacker continued to press onward to higher levels of escalation during the crisis. It is important to underscore that the empirical test in this chapter does not constitute an empirical test of deterrence theory. The theoretical dependent variable of interest is the probability that a LCT attacker escalates to higher levels once the crisis process is underway.

Successful deterrence is defined if a crisis resulted in fewer than 200 combined fatalities for the attacker and the defender/protege team or when the attacker is unable to force the defender to capitulate to its demands. In terms of the theory, successful deterrence means that a LCT attacker refrained from higher levels of escalation. Deterrence failure is defined if the fatalities of both sides exceed 200 or if the defender acquiesces to the attacker's demands. Deterrence failure represents cases in which the attacker did escalate the conflict to higher levels. The outcome of deterrence is measured as a dichotomous variable. Success is coded as 1; failure is assigned a value of 0.

The theory suggests that there are at least two different indicators of the costs of escalation to the defender, $d_j$, that should have an affect on the likelihood that a LCT escalates to higher levels. One indicator is the overall relative size of the attacker and defender's military capabilities. The variable used to represent this concept in the data set is the short term balance of military forces. The short term balance of forces reflects the ability of combatants to impose costs on one another within the first few months of the conflict (Huth, 1988). As the disparity in the short term balance of forces increasingly favors a defender, LCT attackers will be more likely to refrain from escalating to higher levels against these defenders. This variable is measured as the mobilization potential of all ground, air and reserve forces. The forces of the defender and protege were combined to form a single value. The variable is expressed as the ratio of the defender/protege forces over the forces of the attacker. The ratio was adjusted for distance to reflect the potential disadvantage the defender might experience if these forces have to be exported to
the conflict region. When the defender and the protege shared a border, no adjustment was made.

The other variable that affects a defender’s dj payoff is whether or not the defender possesses nuclear weapons. This variable is scored a 1 or 0 if the defender had or did not have nuclear weapons, respectively. Clearly, this variables assumes a value of 0 for all cases prior to WWII. All of the great power defenders had nuclear weapons after 1971.

There are two variables in that data set that are used to reflect the value of the protege to the defender, aj. The first is the presence of absence of a formal alliance commitment between the defender and the protege. Recall that the theoretical expectation is that this variable should have no impact on the behavior of the attacker. If the actors had either a defensive alliance or an entente, the variable received a value of 1; otherwise the variable was coded as 0. Huth assumed that a colony (protege) and its protectorate (defender) could be treated as formal alliance partners even if no official document existed to establish that fact.

Indices of the value of the protege to the defender that are overt and tangible are hypothesized to affect a LCT attacker’s willingness to escalate to higher levels. The variable that represents the value of the relationship between the defender and the protege is the immediate military balance of forces. The variable is measured as the number of troops stationed in or near the region of the protege. Land forces are measured strictly in terms of military personnel. The variable is expressed as the ratio of the combined land forces of the defender and protege over the land forces of the attacker.

Before proceeding with the data analysis, it should be noted that there is a selection bias associated with the set of cases used to test the hypothesis. The selection bias is against all of those cases in which the defender failed to issue a counter-threat against the

---

2 The measure used to adjust for distance was adopted from Bruce Bueno de Mesquita, The War Trap, 1981.
attacker and proceeded to give in immediately to the attacker's demands. All the cases in
the sample represent situations in which the defender believed at some level that the
attacker was bluffing on his threat to escalate the conflict to higher levels. The
consequence of the selection bias is that the empirical test is a "tough" test of theory
because both the defender and the attacker have demonstrated their willingness to escalate
in the conflict.

In sum, a LCT attacker is less likely to escalate to higher levels against defenders
for whom the costs of escalation are relatively low (reflected in an advantage in short term
military capabilities and the defender's possession of nuclear weapons) and are less likely
to escalate when the LCT attacker can observe that the issue at stake is important to the
defender (reflected by a military interdependence between the defender and its protegee).

Data Analysis

The hypothesis is tested using logit. Logit is an appropriate statistical technique to
use whenever the dependent variable is dichotomous. Common techniques to estimate the
model's parameters, such as ordinary least squares, produce inefficient coefficient
estimates because a categorical dependent variable violates the OLS assumption that the
relationship between the dependent and independent variables is linear. Logit is used
instead of other techniques designed to test relationships with a dichotomous dependent
variable, e.g., probit, because two out of the four independent variables are measured at
the categorical level.

Table 4.1 presents the results of the logit test. The overall fit of the model is quite
good. The logit model correctly predicts 83% of the cases whereas the naive model
predicts 58% of the cases accurately. This represents a 43% improvement over chance in
predicting the level to which a LCT attacker will escalate in the crisis. $X^2$ is significant at
the .01 level which means that there is only a 1% chance that the population proportions
are equal.
TABLE 4.1

Logit Estimates for Deterrence Success
(Attacker Refrains from Escalating to Higher Levels)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.941**</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(8.69)</td>
<td></td>
</tr>
<tr>
<td>Milbali</td>
<td>1.140**</td>
<td>.51</td>
</tr>
<tr>
<td>(immediate military balance)</td>
<td>(.564)</td>
<td></td>
</tr>
<tr>
<td>Milbals</td>
<td>.978*</td>
<td>.39</td>
</tr>
<tr>
<td>(short term military balance)</td>
<td>(.539)</td>
<td></td>
</tr>
<tr>
<td>Dnuke</td>
<td>1.537*</td>
<td>.67</td>
</tr>
<tr>
<td>(defender has nuclear weapons)</td>
<td>(.812)</td>
<td></td>
</tr>
<tr>
<td>Ally</td>
<td>-.406</td>
<td>-</td>
</tr>
<tr>
<td>(defender and protege allied)</td>
<td>(.634)</td>
<td></td>
</tr>
</tbody>
</table>

Numbers in parentheses are standard errors

***p < .01  **p < .025  *p < .05

N = 58
Chi-Square = 13.56***

% Predicted correctly
Null = 58%
Logit = 82.8%
The direction of the hypothesized relationships are all specified correctly and each explanatory variable is significant at the .05 level or below.

The theoretical importance of the variable is determined by examining the individual impact each variable has on the probability that the attacker refrains from escalating to higher levels. In regression analysis, OLS produces coefficients that translate a one unit change in an independent variable into a one unit change in the value of the dependent variable. However, given the dichotomous measurement of the dependent variable, there is no common metric underlying the independent and dependent variables. Logit coefficient must be translated using the equation, \( P(Y) = \frac{1}{1+e^{(-a + BX)}} \) which produces a value indicating the probability of observing an event. By manipulating the values of a single independent variable, holding all other variables constant at their means, we can determine the magnitude of the change in the attacker's behavior. Once translated, the coefficients represent the log-odds change in the probability of observing a successful trial of the dependent variable given some specified value of an independent variable, all other variables held constant. In other words, for some specified value of an independent variable, the probability that the attacker refrains from escalating to higher levels (a successful prediction) changes by some probability.

The variable that represents the value of the protege to the defender is the immediate balance of military forces. Different values of the variable were used to represent cases in which a defender's relative military posture changes from a position of relative weakness to one of relative strength. As the value of the immediate military balance of forces increasingly favors the defender/protege team, the attacker is 50% less likely to engage in higher levels of escalation. LCT attacker's who are able to gauge the value of \( aj \) to the defender (the value of the protege) are 50% less likely to press onward to higher levels of escalation. A relative increase in the value that the defender places on the opponent's concession means that the defender is more likely to escalate and any attempt
made by a LCT attacker to bluff the defender into backing down will have a higher chance of failing. Any non-serious attacker's attempt to overrun the protege on the cheap has a greater probability of failing given a defender's relative unwillingness to give in on a valuable issue without first engaging in a costly escalation fight. Therefore, fewer LCT players are willing to consume the oppressive escalation costs necessary to wage a battle that that player is likely to lose.

A slight advantage in the immediate military balances of forces for the defender decreases the likelihood that a LCT attacker presses on by 25%. In this instance, the value of the protege to the defender could be interpreted as weaker relative to the previous cases, and yet, the probability that the attacker refrains from escalating to higher levels remains fairly large. The appropriate comparison, however, is the difference between some indication that the issue at stake is important to an actor versus no signal of value at all. If a LCT actor is contemplating a bluff, the inability to determine the value that the opponent places on the LCT actor giving in (a) cannot have any affect on the bluffer's decision to escalate the conflict to higher levels.

The results of the empirical test show that the coefficient on the alliance variable is not significantly different form zero which is consistent with the theoretical expectation. Essentially, a formal alliance commitment does not provide a LCT player with any solid information regarding the value that the opponent attaches to the former's concession. In effect, the value for a is held constant and the decision of a LCT player to bluff his type and escalate to higher levels is made strictly on the basis of the opponent's other payoffs in the game.

The net value of a player's payoff for different outcomes in the escalation game is also a function of that player's ability to tolerate the costs of escalation. As the capacity to absorb costs increase, the net value of d increases and all other things being equal, a LCT attacker is less likely to escalate given that a bluff is more likely to fail. Taking the short
term balance of forces first, as the defender moves from a position of relative weakness to
one of relative strength, a LCT attacker is 39% less likely to escalate to higher levels. The
relative increase in a defender's capabilities informs that attacker that this defender's ability
to absorb the costs of escalation are relatively greater. Even though the defender may be
no more or less likely to escalate, the net benefit that a LCT attacker derives from
escalating, decreases against a defender who is better able to endure the punishment. The
superior ability of the defender to absorb costs is combined with the defender's greater
capacity to inflict more damage per escalation move on a relatively weaker opponent.
Therefore, fewer attackers are willing to escalate to the high levels necessary to force a
stronger defender to concede to the bluffer's demands.

The value of dj is also a function of the defender's capacity to inflict grave damage
on the attacker. Specifically, an attacker is 67% less likely to escalate to higher levels
when the defender has nuclear weapons. The strong impact of the nuclear weapons
variable on the attacker's behavior runs counter to many of the research findings in
previous work (Organski and Kugler, 1980; Huth and Russett, 1984; Huth, 1988; Geller,
1990). I assumed that if the relationship is spurious, it is probably driven by the impact of
major powers. I ran a series of tests to check the plausibility of this relationship. The first
thing to note is that the bivariate correlation between major power status and the
possession of nuclear weapons is .81 prior to 1971 and 1.0 in 1971 when China exploded
a nuclear devise.

The test was conducted to determine whether major powers were any more or less
successful in repelling attackers from proceeding to higher levels of escalation before and
after 1945. If the relationship between nuclear weapons and the attacker's behavior is
indeed spurious, we should expect no significant difference in the major powers' rate of
success. The first test involved dividing the temporal domain of the data set into a pre and
post nuclear era. The division resulted in 34 and 24 cases, respectively. The defender's
major or minor power status was cross-tabulated against the attacker's level of escalation. The X2 statistic is not statistically significant and indicates that there is no relationship between the defender's ability to prevent the attacker from escalating to higher levels with or without nuclear weapons.

However, another test was performed that casts doubt on the probability that major power status is actually driving the observed effect of the nuclear weapons variable. The variable major power status was substituted for the nuclear weapons variable in the logit analysis and then the test was repeated. Prior to 1945, the probability that the attacker refrains from higher levels of escalation is reduced by approximately 33% when the defender was a major power. However, after 1945, the coefficient on the major power variable is not significantly different from zero. Other checks on the validity of the relationship all resulted in too few cases to be confident in drawing any inferences about these relationships. However, the first series of empirical results are consistent with the theoretical expectation concerning the effect that a decrease in the defender's costs for escalation have on the attacker's behavior.

Conclusion

Deterrence and extended deterrence are policies designed to convince a potential aggressor to back down from his demands by exploiting the fear that the punishment for failing to do so will be so costly as to outweigh any benefits from ignoring the threat and pressing forward (George and Smoke, 1974). However, states who issue deterrent threats are less concerned with destroying the opponent physically than they are with eroding the opponent's will or motivation to continue with the conflict. The ultimate aim of a deterrent threat is to prevent another nation from attacking (Snyder, 1961). To prevent an attack, states try to deter behavior which moves the conflict in that direction. The behavior that actors are attempting to deter during the process of dispute resolution is escalation. The hypothesis tested in this chapter specifies two general conditions that
interact with one another to depress the likelihood that a bluffing attacker presses on to higher levels of escalation. These general conditions were interpreted within a context of extended deterrence to determine whether these variables affect an attacker's motivation to escalate the conflict to higher levels.

The two key independent variables hypothesized to affect an actor's escalation behavior are the costs to the defender for escalating and the value that the defender attaches to the issue at stake, the protege. At the outset, if the defender believes that the attacker has a higher cost tolerance that the defender, the defender fails to make a counter-threat and no escalation is ever observed. Therefore, all of the cases used to test the hypothesis represent cases in which the defender believed at some level that the attacker was bluffing his type. Consequently, the empirical test was a "tough" test of the theory given that both sides in the conflict have communicated their willingness to escalate.

The empirical test conducted in this chapter provides solid support for the hypothesis. The general conclusion is that both the costs that the defender can inflict and the costs that the defender suffers as a result of conceding to the attacker's demand exert a powerful effect on a LCT attacker's decision to escalate the conflict to higher levels. However, the value that the defender attaches to the attacker's concession can only exert an impact on the attacker's behavior if the attacker can observe this value and incorporate it into his calculation (Huth and Russett, 1984; 1988; 1993). Explicit actions taken by the defender concretize and clarify the defender's cost tolerance prior to the attacker taking any escalatory action. It is in those cases where the value of the relationship between the protege and the defender is murky, hence, where the cost tolerance of the defender is ill-defined, which seduce LCT attackers into escalating in the first place.

One of the reasons that indicators of value such as formal alliance commitments have no impact on the attacker's behavior is that alliances provide a LCT player with no useful information regarding the value of $a_j$. The claim that there is something inherent
about alliances which make attackers more likely to escalate appears to be misguided (Huth and Russett, 1984; 1988). The attacker is no more or less likely to escalate because an alliance, in the absence of any tangible and independent linkages that the attacker can observe between the defender and the protege, is a vacuous indicator of the defender's willingness to suffer costs to retain that value.

Theories of deterrence and extended deterrence invariably emphasize that the threat to retaliate must be credible and communicated clearly (George et al., 1971; Jervis, 1979; Lebow, 1981). However, one of the theoretical arguments presented in this chapter is that the credibility of retaliation is contained within the very structure of the extended deterrent situation itself because the attacker's perception of the defender's willingness to escalate is already established prior to either player taking any action. Nevertheless, the defender's willingness to suffer the costs of escalation is never known with certainty and some LCT attackers may challenge the defender to a game of escalation. The hypothesis stated in this chapter identifies the condition when bluffing players may escalate, however, they will exit from the escalation game long before the conflict reaches the higher stages of escalation.
Chapter V
Escalation in Militarized Interstate Disputes

Introduction

The decision to use or threaten to use military force is an important instrument that states access to influence the behavior of other nations. When the opponent uses coercion, the practical affect a player experiences is that the final value he receives for playing the game erodes with every escalation move. This game is actually fought on a psychological battlefield in which escalation becomes the weapon to frighten, convince or manipulate the adversary into backing down.

The previous chapter provided an empirical analysis of two hypotheses that identified how different variables interact to affect the likelihood that an actor escalates and tries to bluff the opponent into giving in. Identifying these conditions and empirically testing the under which a state escalates is an important step in understanding the dynamics of the escalation process. There are other dimensions of an actor's decision to escalate that are also of theoretical and practical interest, however, they have received little conceptual or empirical treatment in the literature.

Clearly, states have little or no chance of changing the opponent's expectations without threatening to escalate and/or actually imposing costs. Once an actor decides to escalate, the choice process does not stop there. The next decision an actor confronts involves determining the optimal level of escalation which will increase the chance that the opponent gives in but at the lowest possible cost. The hypothesis tested in this chapter identifies the relative level of escalation states are expected to choose given their beliefs about the type of opponent they are competing against.

Many theoretical discussions of the level of escalation assume that states escalate to some level and then proceed to analyze the impact that this particular level of escalation has
on the opponent's willingness to counter-escalate (Cross, 1969; Ikle, 1971; Epstein, 1983; Haig, 1984; Patchen, 1988; Freedman, 1991). While there exists some agreement that different levels of escalation have dissimilar effects on the opponent's willingness to remain in the escalation game (Tedeschi, 1970; Bonoma, 1975), few of these conclusions are derived from a theory that specifies these interrelationships in any systematic fashion (Vincent and Schwerein, 1971; Bonoma, 1975; Smoke, 1977). It is even more uncommon to find theoretical treatments addressing the conditions under which states choose particular levels of escalation in the first place. The hypothesis tested in this chapter is derived from the fourth model and identifies the level of escalation that different states employ given their beliefs about the adversary's type.

**Theoretical Summation**

Model IV is a two-sided incomplete information game in which neither i nor j knows whether their opponent is a low cost tolerant (LCT) or a high cost tolerant (HCT) type. The players are less concerned with the absolute value of the opponent's type than they are with the value of the opponent's cost tolerance relative to their own. If i and j have the same type, the information they want to know is whether their threshold is higher or lower than a player of similar type. The overall conceptual problem actors face is that while they want to prevail in the conflict, they want to do so at the lowest possible cost to themselves. These twin desires open up a wide range of strategic possibilities regarding the level of escalation to employ and raises questions concerning the timing of these moves within a conflict. The decision over what level of escalation to employ is captured in model IV by assuming that states have a choice between escalating high or low. The difference between these two strategies is the amount of costs suffered by the imposer and the recipient, e.g., high escalation carries higher costs for both parties relative to lower levels of escalation.
By assuming that the players’ escalation strategies can be represented by a low-high escalation dichotomy, I am able to incorporate the theoretical argument developed above while retaining the mathematical tractability of the model’s solution. To put closure on the model, I assume that a low escalation move by i on the last move is tantamount to giving in to j’s demands. However, low escalation by either player prior to the final move means imposing fewer (not necessarily zero) costs on the adversary and oneself.

Model IV begins with a HCT or LCT i’s decision to escalate high or low. The hypothesis tested in this chapter refers to the escalation behavior of LCT i actors, therefore, the discussion concentrates exclusively on the choice problem from a LCT i’s perspective. Unlike model III, an i player in model IV no longer has the luxury of knowing with certainty j’s preference ordering over the outcomes. However, this assumption does not alter conceptually the calculations i performs to determine whether escalating high or low is optimal. Since the payoffs to both a LCT and HCT j are common knowledge, a LCT i can determine a mixed strategy of low and high escalation such that the expected value to a j player of either type for escalating high and low are the same. Player i determines two mixed strategies: one to be played against a LCT j and another mixed strategy used against a HCT j. For example, a LCT i escalates high 25% of the time against a LCT j but escalates high 40% of the time when j is HCT. The problem, however, is that i does not know which probability of low escalation to play without also knowing j’s type with certainty. Thus, i calculates an overall mixed strategy that randomizes between low and high escalation such that both types of j players are indifferent between low and high escalation strategies.

A low cost tolerant i’s mixed strategy is derived using sequential equilibrium concept:

\[ p_s((x)^{(b)} + (1-x)^{(f)}) + p_{(1-s)^{(x)^{(l)}} + (1-x)^{(s)}} = 3.4 \]
\[ p_s((x)^{(e)} + (1-x)^{(g)}) + p_{(1-s)^{(x)^{(n)}} + (1-x)^{(t)}} \]
Isolating $x$,

\[
x = \frac{p((s^*(g - t + f + s) + (t - s)))}{p((s^*(b - f + 1 + s - e + g + n + t) + (1 - s - n + t)))}
\]

The value, $x$, is the probability that a LCT i escalates high against either j type player and $1 - x$ is the probability that a LCT i escalates low. The independent variables that affect i's decision to escalate high are: (1) S, i's prior belief that j is LCT and (2) P, j's prior belief that i is LCT and (3) the letters in brackets which represent a HCT and LCT j's payoffs for various outcomes in the game. By manipulating j's payoff values and/or the probability values i associates with j being a LCT player, we can derive several hypotheses that identify the conditions under which a LCT i is more or less likely to escalate high.

The hypothesis tested in this chapter is:

"As the disparity between the players' cost tolerances increase, the lower cost tolerant actor is more likely to escalate high"

The hypothesis is the result of manipulating the value for S, i's belief that j is a low cost tolerant player, while holding all other variables constant in the equation. Recall that players' decisions are based on a comparison between their cost tolerance and the opponent's cost tolerance. For i, knowing that there is a greater probability that j is LCT does not speak to i's need to know whether j can outlast i in a game of escalation. From model I, we know that a player concedes immediately when he knows that the opponent has a relatively higher cost tolerance threshold. The same logic applies here except that now i and j must actually play a game of escalation to determine which of the two possesses a greater willingness to suffer costs. To clarify the discussion, I briefly reexamine the intuition behind this hypothesis which was elaborated in more detail in the third chapter.
In general, high cost tolerant players know that they are in a better position to withstand the costs of escalation especially against low cost tolerant players. Therefore, HCT players may have a strong incentive to escalate the conflict at lower levels to minimize the costs to be paid in achieving their demands. Because of their relatively greater willingness to suffer costs, HCT players can shift to higher levels of escalation if the adversary refuses to back down. Low cost tolerant players, particularly if they are competing against a HCT opponent, may not have the same flexibility in choosing between various levels of escalation. LCT players are unlikely to prevail in a long, drawn out war of attrition since they will achieve their threshold before the opponent and will have little to show for it except a bundle of costs. This of course presumes that a HCT player believes that the low escalating player has a lower cost tolerance threshold than his own and will not drop out of the dispute. The willingness of the LCT actor to escalate is a function of how he values the issue at stake. If the prize is valuable enough, even though there is little chance of prevailing in a long escalation game, one of the only ways that a LCT player can convince the adversary that he can withstand the costs of escalation is to demonstrate to the high cost tolerant player that he is willing to suffer serious escalation costs. A low cost tolerant player can best accomplish this by escalating high from the outset of the conflict. The hope of the LCT player is that this move (from zero to high escalation) will be sufficient to convince a higher cost tolerant adversary that he cannot outlast the escalating player in the conflict and will concede as a result of that belief. The risk that the LCT player assumes is that if the HCT is not convinced, the inability of the LCT player to sustain that escalation pace compels him to exit the escalation game shortly thereafter.

The theory makes use of a conceptual distinction between "high" and "low" levels of escalation in order to derive expectations regarding the conditions under which different actors elect different levels of escalating in a conflict. The theory, however, does not
require an assumption that high or low escalation means the same thing to all actors and this includes actors of the same cost tolerance. The definition that accompanies high or low escalation is assumed to be subjectively determined by the actor based on the particular set of constraints decision makers believe operate on their choices within a dispute. The hypothesis tested in this chapter, therefore, does not predict to the absolute level of escalation a lower cost tolerant player will achieve on the first move in the conflict. The hypothesis that identifies this actor's behavior is based on the entire play of the game including that actor's first and final move. Thus, the dependent variable is the level of escalation pursued by the lower cost tolerant actor relative to his final action in the game. The logic of the theory suggests that lower cost tolerant actors, who decide to escalate, will escalate to the maximum of their ability from the outset so as to increase the chance that a higher cost tolerant actor will back down. If this demonstration fails to convince the adversary to concede, the lower cost tolerant player will be unable or unwilling to escalate higher having already failed in the attempt to win at a maximum level of escalation.

The hypothesis is tested in the context of Militarized Interstate Disputes (MIDs) (Gochman and Maoz, 1984). Militarized Interstate Disputes are defined as a "set of interactions between or among states involving threats to use military force, displays of military force or actual uses of military force. To be included, these acts must be explicit, overt, non-accidental and government sanctioned (Gochman and Maoz, 1984: 587). The military emphasis in defining a dispute was motivated by research questions pertaining to why some acute conflicts of interest evolve from the use of force as an instrument of influence during peacetime into the outbreak of war between nations. Much of the empirical work on this subject has described the process of moving from lower to higher forms of coercion as escalation (Cusack and Eberwein, 1982; Leng and Gochman, 1982; Maoz, 1983; Gochman and Maoz, 1984)."}

1 It should be noted; however, that when escalation has been defined explicitly in this literature, the
MIDs provide another context in which the dynamics of the escalation process can be analyzed. The nature of the crisis itself, which has brought the players to this stage in the conflict, has increased the probability that one or both of the players will resort to the use of force. Given the already high probability that force will be used, the MID context lends itself to an investigation of the ways in which that force is likely to be applied by different actors.

Operational Hypothesis: The Dependent Variable

The dependent variable of the hypothesis is the probability that the lower cost tolerant player achieves a higher level of escalation on a subsequent move compared to the first action taken in the dispute. However, the term "high" has little meaning in and of itself until it is compared with some other action that a player takes in the dispute. The hypothesis states that if the lower cost tolerant actor is going to escalate high from the outset of the dispute that this level of action should be the maximum escalation achieved for the weaker side.\(^2\) In other words, the highest level of action that a lower cost tolerant player commits will correspond closely with the level of escalation taken by that actor on his first move. Note that the definition of "high" escalation is not based on the level of escalation some other actor might take, e.g., i sends a nasty diplomatic note and j assassinates the beloved head of state of its rival. The term high is established by comparing that action to some other action taken by the same player later on in the dispute. What we are concerned with here is not whether the action taken imposes massive costs on the opponent. Rather, the hypothesis states that the player will achieve his highest level of action (whatever that action might be) on his very first move. The logic of this proposition

---

\(^2\) The level of escalation is conditional. Recall that all players in model IV have a mixed strategy to escalate low with some probability and to escalate high with one minus that probability. Some LCT i actors will choose to escalate low which can include imposing zero costs.
follows directly from the model. If a lower cost tolerant player has any chance of getting his demands satisfied, then his act must be convincing from the start because he cannot establish the willingness to bear costs before his time runs out.

Conceptually, the dependent variable involves a comparison between the first escalation action taken in the dispute and the highest level of escalation ultimately achieved by the weaker actor in the dispute. What is at issue is how to classify an action as a higher escalation move relative to some other action that could have been taken. It is important to stress that the hypothesis refers exclusively to the escalation behavior of the lower cost tolerant player in the dispute. Thus, the costliness of any particular action is evaluated from that actor's perspective in terms of the costs he suffers for carrying out the escalation action. It is less important theoretically to determine how the adversary evaluates the costs he suffers because of the lower cost tolerant actor's behavior or whether the recipient believes the lower cost tolerant player has escalated to a higher level compared to the initial action.

The distinction made between high and low escalation in chapter three involved an assessment of the costs associated with a particular action. Although the costs an actor suffers from escalation is based on subjective estimates, we can approximate these costs and roughly categorize actions as more or less costly. The assumption is that certain actions are considered higher escalation moves if they involve greater costs to be executed. There have been a number of different attempts to classify the level of different escalation actions (Bleichman and Kaplan, 1978; Gochman and Maoz, 1984). Since these efforts correspond closely with the way in which different escalation levels are conceptualized here, these frameworks will guide the discussion. To evaluate the relative costliness of an action, we need to anchor the scale at both the low and high end of the escalation continuum. The minimum point is represented by no escalation or zero cost imposition. An actor who fails to escalate is in effect exiting from the escalation game. Theoretically,
the maximum escalation is the complete exhaustion of one's resources in a single escalation move. This is a fairly unrealistic upper boundary and actually unnecessary to test the hypothesis since all that is required for an appropriate test of the theory is that enough actions be identified to provide some variation in the level of escalation observed. The highest level of escalation is defined when two nations go to war.

Escalation involves either a threat to use force or the actual application of sanctions. This dichotomy serves as one cut at classifying escalation actions that fall somewhere between the two extremes. The distinction between the two types of escalation action are evaluated in terms of the immediate costs that a player suffers as a result of executing the action. A threat to use force is the least costly action next to no action at all. The threat can involve any declaration of future intent such as a threat to cut off military assistance or the threat to declare war if the rival does not comply with one's demands immediately. Neither the recipient nor the sender of a threat experiences any current costs. Any costs that the sender incurs are borne out in the future either because the threat was actually carried out or because the threatening state suffers costs for failing to fulfill the promise, e.g., reputation costs (paid in the future).

A use of force involves suffering current costs. Resources are expended in moving troops, material or dropping bombs on the opponent's territory. When these resources are used to actually inflict pain on the other side (and not simply a demonstration of the capacity to do so), the cost to the escalating state is even higher.

In sum, the level of escalation is evaluated in terms of the costs suffered by the escalating state for carrying out action or threat of action. The criteria used to distinguish between different levels of escalation is an assessment of the current costs the imposer suffers for executing some action. A use of force is considered a higher escalation action if it is used to punish the adversary. Threats are least costly among the various types of escalation because no current pain is experienced for issuing a threat to inflict harm. And
lastly, no escalation is least costly of all since this action is tantamount to ending or exiting from the conflict entirely.

*The Independent Variable*

The independent variable of the hypothesis is the relative cost tolerance of the players. Recall that the definition of cost tolerance refers to the maximum cost-benefit tradeoff a player is willing to make to remain in the escalation game. That is, the level of costs one is willing to suffer in a dispute depends on the value of the object at stake to a player. To create a perfect correspondence between the conceptual meaning of cost tolerance and its operationalization, it is necessary to determine how each actor evaluated the issue in the dispute for all actors in all disputes across time. There are some rather profound limitations involved with this exercise, however, there are other factors that can be used to represent cost tolerance.

In addition to the value of the issue, the other element involved in defining a player's cost tolerance threshold is the costs an actor is willing to bear. The ability of a state to impose costs is achieved primarily through a state's military force. A player's military capabilities are a combination of the various resources a state can gather for the purpose of fighting (Organski and Kugler, 1980; Small and Singer, 1982). As mentioned previously, a state's capacity to inflict damage increases as that actor's military strength increases relative to the rival. As Rosen (1972) notes, an increase in the amount of those base resources means that the same absolute loss a player suffers lessens relative to the case when it was weaker. The stronger side's ability to replenish its material more effectively also suggests that the marginal cost per use of force is declining (Rosen, 1972). All other things being equal, an increase in a player's relative military capabilities should also be associated with a relative increase in that player's cost tolerance and its willingness to remain in the escalation game (Cazzinno, 1980). This is reinforced by the fact that the opponent's capacity to inflict damage in an equivalent amount has declined, relatively
speaking. The opposite is true for the relatively weaker side. The weaker side will now be confronted with an adversary whose capacity to inflict damage is greater. Thus, the cost to the weaker side of using its own military force has just increased.

The operational hypothesis leads to the expectation that as the disparity in military capabilities increases, the probability that a weaker actor escalates to a higher level compared to the first escalation move decreases. The hypothesis is stated at the monadic level of analysis. Even though there are at least two participants for every dispute, I am only interested in the behavior of the weaker actor. The behavior in question is the first action taken by the weaker side versus the highest escalation level ultimately taken by the weaker state. Since the hypothesis is stated in probabilistic terms, what we should observe if the hypothesis is correct is that as the power disparity between the actors increases, the probability that the weaker actor achieves a higher level of escalation relative to the first action decreases.

Research Design

An appropriate test of the theory requires that the dependent variable capture different types of escalation actions. The theory discusses escalation in terms of different levels so that any action identified can be classified as involving a higher or lower level of escalation relative to the first action taken in a conflict. The Militarized Interstate Incident Data Set (Gochman and Maoz, 1984) provides information on the types of actions taken in a dispute as well as the number of incidents each party took per dispute. Escalation is represented in the Incident Data Set by a variable called a military confrontation action (MCA). There are 14 different types of MCAs which were grouped according to whether the action represented a threat, display, use of force or war.\textsuperscript{3} The theory makes a distinction between the costs associated with different categories of escalation levels but does not address the difference in cost between actions that fall within the same category.

\textsuperscript{3} This categorization scheme is found in the MID data set and was originally suggested by Stoll (1978).
According to the argument presented earlier, the highest level of escalation is war followed by a use and display of force. A threat to use force represents the lowest level of escalation. The dependent variable is measured at the ordinal level because it is possible to identify actions which are more or less escalatory but there is no underlying metric between the escalation categories to define precisely the amount of that difference. Because the hypothesis refers to dyadic disputes and the escalation behavior of the weaker actor exclusively, all disputes in which nations joined an ongoing conflict were excluded from the analysis. In addition, all minor-minor power disputes had to be excluded because no data was collected for these types of disputes.

The dependent variable requires identifying two different escalation actions taken by the weaker side in the conflict: the first escalation action taken in the dispute and the highest level of escalation achieved for the entire dispute. A dichotomous dependent variable was created to indicate whether the highest escalation action differed from the first action taken in the dispute. The dependent variable was scored 0 if the weaker side escalated to a higher level of escalation relative to the first move. The variable was coded 0 if the first and highest escalation action were taken at the same level.

The independent variable of the analysis is the relative capabilities of the two sides. The capabilities data was taken from the Correlates of War (COW) Data Set (Small and Singer, 1982). The capability variable is measured using six different indicators: military expenditures, military personnel, urban population, total population, iron/steel production and energy consumption. These indicators were combined in the COW data set to create an index which represents a state's percentage share of the world's capabilities in that year. A ratio of capabilities for the two actors was created using these percent capability scores. The ratio was created such that the stronger of the two sides was represented in the numerator. The ratio can vary between 1 (equality in capabilities) to any positive number. The ratio identifies which actor's escalation behavior is included in the various
tests and serves as the explanatory variable in the logit analysis.

*Data Analysis*

The hypothesis was tested using logit estimations of the probability that the weaker side would achieve a higher level of escalation after its first action. The dependent variable was scored 0 if the weaker side took a higher escalation action and 1 if both escalation action were taken at the same level. Logit is an appropriate statistical technique for two reasons. First, the dependent variable is dichotomous and techniques such as ordinary least squares will produces inefficient estimators. Second, a control variable is included in the analysis which is measured at the ordinal level. The control variable is the first level of escalation taken by the weaker side. The first action taken does not serve as an explanatory variable because the theory does not predict the level of escalation that the weaker side will choose as its first action. The theory does state that given the first action taken by the weaker side (whatever that action might be), the weaker nation is less likely to escalate beyond the level of that first action. This control variable is necessary since there are constraints that operate on states depending on the first level of action taken. For instance, states that begin with a use of force can only go to war to achieve a higher level of escalation. On the other hand, states that begin with a threat to use force can escalate up to three levels higher.

The ratio of military capabilities is the explanatory variable and is used to predict the probability that the weaker side will achieve a higher level of escalation. There are 39 cases in which a weaker side took a higher escalation action and 80 cases in which the first escalation move was the highest the weaker side attained for the entire dispute. The results of the analysis are presented in Table 5.1. The results indicate that a moderate to weak relationship exists between the disparity in capabilities and a decrease in the likelihood that the weaker actor escalates to a higher level relative to the first action.
### TABLE 5.1

**Logit Estimates for Weaker State's Escalation Behavior**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.5139</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(1.03)</td>
<td></td>
</tr>
<tr>
<td>Ratio</td>
<td>-0.006</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td></td>
</tr>
<tr>
<td>First Esc. Action</td>
<td>-1.63*</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.333)</td>
<td></td>
</tr>
</tbody>
</table>

Number in parentheses are standard errors of the estimate

N = 121

*p < .0001

% correctly predicted
Null = 67%
Logit = 78%
The inverse relationship between the two variables is consistent with the hypothesis, however, the standard error of the estimate is a little more than one-half the size of the coefficient. The naive model which simply predicts the outcome based on the modal category, correctly predicts that the weaker side will not achieve a higher level of escalation 67% of the time. The logit model correctly predicts 79% of the cases (a 12% improvement over chance), however, the improvement in the prediction rate is driven almost exclusively by the control variable, the first action taken in the dispute.4

One potential reason for the modest to weak findings is that the ratio variable used to predict the behavior of the weaker side must be rather discriminatory given the distribution of the cases on the dependent variable. The naive model is already predicting 67% of the cases correctly in the absence of any information on the actor's capabilities ratios. The inability of the capability ratio variable to discriminate strongly between the cases is corroborated by an inspection of the ratio variable's contribution to the change in the probability of observing the weaker side taking a higher escalation action. By manipulating the values of the ratio variable and holding the control variable constant at its mean, I derive the change in the probability of observing a successful trial of the dependent variable (a decline in the probability that the weaker side escalates to a higher level). When the capability ratio changes from one of rough equality to cases where the stronger side has twice the military capabilities of the weaker actor, the probability that the weaker side escalates to a higher level decreases by 2%, from .36 to .34. When the capabilities of the actors changes from parity to cases in which the stronger side is 100 times stronger than the opponent, the probability that the weaker side escalates higher than the initial action decreases by only 13%, from .36 to .23.

The evidence presented thus far suggests there is only a modest to weak

4 Because the dependent variable is based partly on the first action taken in the dispute, the two variables are highly intercorrelated.
relationship between the variables. Because the logit analysis collapses a large amount of information into single estimates, an inspection of a contingency table serves as a useful diagnostic to analyze the percentage of cases that fall within the predicted categories. To create a crosstabulation matrix, I collapsed the continuous ratio measure into four categories based on the interquartile range. Table 5.2 presents the categorical ratio measures crosstabulated against a variable defined as the escalation level. No escalation represents cases in which the first and highest action were taken at the same level of escalation. The other two categories are cases in which the highest escalation was one or two levels above the first action taken by the weaker side. The hypothesis leads to the expectation that the percentage of cases in the no escalation category should increase as we move across the ratio categories from left to right (rough parity to extreme disparity). The proportion of cases falling within the one or two higher escalation categories should decrease moving left to right.

This expectation is only partially borne out. The contingency table analysis shows that when the two sides were roughly equal in power, the slightly weaker actor did not escalate to a higher level on any subsequent move in 71% of the cases. It is only when we reach the last column, where the disparity between the actors is greater than 60:1, that the increase in the frequency of cases is in the hypothesized direction. The hypothesis predicted that the number of cases would be increasing as the power disparity between the disputants increases. This is only true for gross inequalities in power. The same conclusion holds for cases in which the weaker actor escalated to one level higher following the initial escalation move. As the disparity between the actors increases, the probability that the weaker actor escalates to a higher level increases until we reach the final column. The only cell in the category of two higher escalation actions that has a higher frequency of cases than expected are cases when the stronger side has anywhere from a 12:1 to 60:1 advantage over its weaker opponent. In fact, the percentage of cases falling
**TABLE 5.2**

The Level of Escalation Achieved by the Weaker State in a Militarized Interstate Dispute

<table>
<thead>
<tr>
<th>Power Ratio*</th>
<th>Ratio &lt; 2.44</th>
<th>Ratio ≥ 2.44 and &lt; 12.03</th>
<th>Ratio ≥ 12.03 and &lt; 59.31</th>
<th>Ratio ≥ 59.31</th>
</tr>
</thead>
<tbody>
<tr>
<td>Escalation</td>
<td>Escalation on first and last move</td>
<td>Escalation on first and last move</td>
<td>Escalation on first and last move</td>
<td>Escalation on first and last move</td>
</tr>
<tr>
<td>Same level</td>
<td>71%</td>
<td>63%</td>
<td>50%</td>
<td>83%</td>
</tr>
<tr>
<td>One level higher</td>
<td>16%</td>
<td>27%</td>
<td>33%</td>
<td>10%</td>
</tr>
<tr>
<td>Two levels higher</td>
<td>13%</td>
<td>10%</td>
<td>17%</td>
<td>7%</td>
</tr>
<tr>
<td>N=31</td>
<td>N=30</td>
<td>N=30</td>
<td>N=30</td>
<td>N=30</td>
</tr>
</tbody>
</table>

*Power Ratio = Disparity in military capabilities of stronger state over weaker state, e.g., a power ratio of two means that stronger state is twice as strong as the opponent
within the cells of the third column are all inverse from the theoretical expectations of the model. This suggests that there may be some systematic pattern to the cases that fall within this category.

To check this suspicion, an analysis of the cases that comprised this third column revealed that approximately 25% of the cases were drawn from World War II. This raises the possibility that a complex set of alliance patterns which would produce non-independent cases, could be depressing the strength of the findings. I assumed that by failing to incorporate the capabilities that an alliance partner could bring to bear in the conflict on behalf of the weaker side, may mean that some cases which appeared to be disparate in capabilities would, following an adjustment of the ally's capabilities, be more likely to fall into categories of rough parity. As a first test, a logit analysis was re-run excluding the eleven World War II cases. The sign on the coefficient of the ratio variable was in the hypothesized negative direction, however, the statistical significance of the variable actually decreased compared to the original test of the data. Second, the contingency table analysis did not reveal any significant difference between an analysis with or without the WWII cases. Except for the WWII cases, there did not appear to be anything unusual about the cases that made up the 12:1 to 60:1 power disparity dyads. The results of the empirical analysis do not completely disconfirm the hypothesis, however, the results are much weaker than expected. Further investigation is required to determine what factors are contributing to the uneven results across the cases.

Conclusion

In the beginning of this chapter, I noted that there are few theoretical inquiries that specify the way in which states are expected to escalate in a conflict and far fewer empirical tests on the subject. This chapter presented an empirical test of one of the derivations of the fourth model which identified the level to which a lower cost tolerant actor was expected to escalate in a conflict. In brief, the argument states that the lower cost tolerant actor should
be more likely to escalate to the maximum of his capacity on the first move. Lower cost tolerant actors are aware that their chances of prevailing against a higher cost tolerant opponent in a long, gradual game of escalation are fairly low. Thus, one of the only means available to a lower cost tolerant actor who wants to achieve his demands is to convince an adversary that he can withstand the costs of escalation by executing a bold, decisive escalation action that approaches the maximum of his capacity on the first move (the opponent does not know of course whether or not that maximum has actually been achieved).

The empirical support for the hypothesis is less than overwhelming. While there is a slight tendency for the weaker side to fail to escalate to a higher level relative to the first action taken in a dispute, the relationship is not particularly strong excepts in cases of gross capability disparities. One potential reason for the weak findings is that the key explanatory variable, military capabilities, does not explicitly incorporate the capabilities that a potential ally could bring to bear on behalf of the weaker side (Sabrosky, 1980; Bueno de Mesquita, 1981; Levy, 1983; Siverson and Tennefoss, 1984; Kim, 1990). If this assumption is correct, then a weak player with the backing a major allies, would be expected to escalate to higher levels as the disparity in capabilities decreases.

Another potential reason for the modest findings is that the relationship between capabilities and cost tolerance is not as closely associated as assumed. It may also be the case that the two concepts are related but only for certain capability thresholds not currently specified by the theory. Perhaps there are a larger number of instances in which the weaker actor is high cost tolerant rather than low cost tolerant. If this is the case, then the former would also be expected to achieve higher levels of escalation following the first move. This interpretation is consistent with empirical findings that the level of "resolve" (similar conceptually to the notion of cost tolerance) is independent from and a better predictor of an actor's escalation behavior than are relative capabilities (Lebow, 1981; Maoz, 1983; 1985).
However, the results clearly indicate that the relationship holds for a subset of cases but not for the entire sample. It may be that the theory is not specifying the relationships properly, however, one must conclude based on the empirical test in chapter four on extended deterrence crises and the solid evidence in support of the theory, that dismissing the theory is premature. It may be the case that variables not controlled for in this study such as the location of the conflict or the alliance patterns between states, could be depressing the strength (Most and Starr, 1980; Cusack and Eberwein, 1982; Siverson and Tennesfoss, 1984).

The way in which states escalate sometimes leads to complex patterns of dispute resolution. It is important to gain an understanding of these patterns particularly since some types of escalation appear to end in violence more consistently than others (Leng and Wheeler, 1979; Leng, 1983; Greffenius, 1990; Wilkenfeld, 1991). The interconnections between process and outcomes are complex indeed but our understanding of these patterns will not progress as systematically as it could, unless the relationships between decision makers' beliefs, motivations and escalation behavior are derived from a well specified theory.
Chapter VI
Conclusion

Overview

Escalation is an integral part of the conflict landscape in the international arena. Actors adopt escalation strategies to obtain something of value which can only be achieved by convincing the other side that it is in his best interest to meet the coercive state's demands. To convince the other side through escalation is to exploit the adversary's fear of the threat of future pain and suffering (Schelling, 1966). Despite the prevalent use of escalation tactics in all types of conflict interactions surprisingly little is known about the dynamics of the escalation process.

The fundamental research question of this thesis inquired as to the conditions under which actors were expected to escalate in conflict. This seemingly straightforward question has generated nothing short of a plethora of different answers. Escalation has been hypothesized to occur only when a state has superior strength; only when states are weaker than the opponent; when they can expect to win; when they want to stave off defeat; for purposes of revenge or anger or states escalate because the opponent escalated. Often these relationships are stated unconditionally, e.g., escalation is best explained as a hostility spiral (Osgood, 1962; Holsti, North and Brody, 1964; Milburn, 1977). The list of variables that purportedly drive the escalation dynamic could be extended even further. However, the main point is that a great deal of the diversity in opinion on escalation is partly attributable to failure to develop systematic theories that address the process of escalation and the lack of empirical tests which can be used as evidence in support or against the theory specifying the link between a variable and its affect on escalation. As a result, the number of variables considered to have a strong impact on the escalation behavior of states continues to grow without any way to discern the plausibility of these relationships.
In this thesis, I have developed a theory of the escalation process which is based on the view that escalation is part of a larger process of interstate bargaining. The bargaining conceptualization of escalation is useful because it recognizes explicitly that the conflict between two states is over some issue. Viewed from this perspective, the incentive for states to escalate is the desire to convince the adversary to back down from his demands. In turn, the value that players attach to the issue at stake defines the relative willingness of a state to suffer the costs of escalation. The maximum costs a player is willing to pay to gain some prize is defined as a player's cost tolerance and this is one of the key concepts of the theory. The importance of the cost tolerance concept is that the costs that players are willing to absorb to play the escalation game are endogenous.

The game theoretical models developed to represent the escalation process were built in four different stages. One advantage that formal theory brings to the research enterprise is an increase in analytical rigor. The simplification process exposes the basic interconnections between the variables and clarifies the logic underlying the model's development. This approach proved to be extremely useful for a number of different reasons. A useful theory must produce a set of falsifiable hypotheses that can be compared against the historical record. Each of the models developed throughout the thesis produced a set of falsifiable expectations which identified different conditions under which states were expected to escalate in conflict. Throughout the third chapter, the parameters of the various hypotheses were given empirical content to demonstrate the kinds of variables present in the empirical world that reflect the abstract relationships proposed in the model. The theory of escalation was developed so that the interactions among the variables specified in the various hypotheses were applicable across a wide variety of different international relations contexts.

Each model produced some insight regarding the dynamics of the escalation process. As the Vietnam example used throughout the discussion illustrates, some of the
models were better equipped than others in addressing the conditions under which states were expected to escalate. The conceptual limitations of each model were addressed by extending the model in ways that could shed light on the important research questions. All of the models were generalizations of their predecessors and the derivations from different model versions produced important refinements on the conditions under which each of the hypotheses were expected to hold.

The first model produced an important but fairly intuitive conclusion. The player who can best withstand the costs of escalation should win the game. If we evaluate the ability to suffer costs in terms of an actor's level of resources, we would expect, on average, that stronger and more powerful nations tend to prevail in the majority of international conflicts. And this conclusion seems to hold up quite well against the empirical evidence (Ray, 1990). The derivation from model I is also important in terms of providing a general expectation that can be compared against the hypotheses derived from later model versions utilizing a more realistic set of assumptions.

While the outcome predicted from model I is consistent with many events, model I also suggested that escalation should never occur, a prediction inconsistent with empirical observation (Smoke, 1977; Gochman and Maoz, 1984). The complete information assumption in model I means that the players can determine the winner and the loser prior to the genesis of the escalation game. As a consequence, the loser submits to the demands of the winner immediately. President Johnson would have never escalated American involvement in Vietnam; the Soviets would have never installed strategic missiles in Cuba in 1962 and so on.

Clearly, however, if we are to gain any insight into why states escalate, (and not just who is most likely to prevail), the assumption of complete information must be relaxed (Wagner, 1982; Powell, 1987; Nalebuff, 1991; Kilgour and Zagare, 1991). Once incomplete information was incorporated into the basic model structure, models II-IV
produced important refinements on the conclusion that the stronger state always wins. Under certain conditions, players can take advantage of their opponent's uncertainty over the outcome in an effort to manipulate the opponent into believing that he cannot outlast the escalating player in a game of escalation. It is of course true by definition that players who think they can win the conflict at an acceptable cost are the only players that escalate in a conflict. However, we must be careful not to equate a greater willingness to suffer costs with the player that always prevails. Under certain conditions, bluffing one's willingness to play the escalation game to unacceptable cost levels to the opponent can succeed (Powell, 1987).

The theory also provides some insight on why certain patterns of escalation emerge (Pruitt, 1969; Leng, 1979; 1993; Gochman and Leng, 1983; Wilkenfeld, 1991). If all that was involved in the decision to escalate was a desire to pummel the adversary into submission, then it is likely that states would use high levels of coercion immediately if they believed that there was a decent chance that the opponent would be unable to counter the cost imposition. However, states also have an incentive to minimize the price necessary to convince the opponent to give in to one's demands. The ability of players to utilize a strategy that entails the minimum possible costs is a function of a player's beliefs about the type of adversary he faces. Some states may have a lesser ability to maximize the chance of getting their demands satisfied while minimizing the price that has to be paid to secure the prize.

The hypotheses, whether they refer to the conditions under which states are expected to escalate or the level to which states are expected to escalate, are all formulated at a level of generality such that the hypotheses derived from the models can be tested empirically in different contexts in which escalation is a real possibility. Three of these hypotheses were tested in two different international contexts: extended deterrence crises and militarized interstate disputes.
Empirical Evidence

The empirical findings presented in this thesis provide fairly solid support for the theory, however, the strength of the findings is mixed and somewhat weaker in the context of militarized disputes. The results of the empirical test using extended deterrence crises clearly show that states’ decisions to escalate a conflict to higher levels is a function of the anticipated reaction of the opponent. It is not, however, the desire to avoid a vicious action-reaction dynamic that prevents players from engaging in higher forms of coercive behavior (Young, 1968; Ikle, 1971). Players' decisions to exit from the escalation game depend critically on the beliefs that players hold about the value the opponent attaches to his own concession as well as the level of destruction that the adversary can impose throughout the process of reaching some kind of agreement. It may be the case that the internal political needs of states' decision makers provide an impetus for challenging another actor to a fight (Jervis, 1976; Lebow, 1981; Russett, 1990; Morgan and Bickers, 1991), however, what the theory suggests and the empirical evidence clearly shows is that the willingness of low cost tolerant actors to escalate to higher levels, irrespective of the initial motivation of a challenge, is closely related to the perception of the havoc the adversary is willing and able to wreak on the player daring it to a game of escalation. In order for these variables to have an affect on the actor's escalation behavior, actors must be able to perceive the values that these variables assume for the opponent (Blainey, 1973; Jervis, 1976; George, 1980; Lebow, 1981).

A different hypothesis was tested in the context of Militarized Interstate Disputes. The hypothesis pertained to the level to which a lower cost tolerant player was likely to escalate when the protagonist was believed to be of a higher cost tolerance. Cost tolerance was measured in terms of military capabilities and the operational hypothesis lead to the expectation that as the disparity between the two sides increased, the weaker actor was expected to escalate to the maximum of his capacity on the first move. The findings
confirmed that there was an inverse relationship between increasing power disparities and a decrease in the likelihood that weaker actors to achieve their highest level of hostility after the first action. The findings, however, were not uniform across all cases. More work needs to be done to determine the reasons for the patterns of escalation that emerged from the analysis.

Limitations of the Current Study and Future Research

The empirical analyses conducted in this thesis suggest that the theory of escalation developed is a useful approach in gaining some insight on the causes of interstate escalation. However, I am also mindful of some of the limitations of the current work that must be addressed in future research.

One limitation of the study is that while treating the concession strategy as an all or nothing choice provides a useful first cut at the problem, the failure to specify how concessions and escalation interact to affect patterns of dispute resolution may be obscuring one important factor acting on states' incentives to remain in the escalation game. For instance, what effect does a partial concession have on the subsequent escalation behavior of an actor? Both of these strategies are important tools of crisis management and the interdependent affects of these tactics on one another is bound to lead to some refinements concerning the levels to which actors are willing to escalate (George et al., 1971; Lebow, 1981; Craig and George, 1990).

Similarly, treating escalation as a dichotomy has provided a great deal of insight into the conditions under which states are expected to escalate, however, the range of useful and important questions that can be addressed in the research is somewhat limited by the assumption of a discrete strategy space. As it currently stands, the factors that affect escalation within a dispute are all fairly static. Once a continuum of escalation is incorporated into the model, hypotheses can be derived which state the relationship
between more dynamic factors such as choices over the rate of escalation and the affects of these choices on the opponent's willingness to remain in the escalation game.

Finally, more empirical work is necessary before the strength of the theory can be assessed. Since the hypotheses derived from the various models identify relationships among the variables that are generalizable across other contexts, empirical tests of these propositions in the areas of arms races and war should provide important additional evidence for or against the theory.

This thesis is the first stage in the development of a more comprehensive and complete theoretical understanding of the dynamics of the escalation process. The theory developed here has proven to be a useful framework that can be built upon and extended in some of the ways mentioned previously. Because the escalation process can produce serious pain and destruction for all parties involved in a conflict, responsible scholarship demands that we inform those making decisions why simple-minded dictums such as "avoiding another Vietnam" does not shed light on which foreign policy adventures can be entered into with sound policy results and those that should be avoided if we are unwilling to pay the price.
References


Huntington, Samuel P. "Arms Races: Prerequisites and Results." *Public Policy* 7 (1958): 539-471.


Appendix 1

Model II: One-sided Incomplete Information

Player i's decision rule: c_i > a_i*

p = probability that j is LCT
1 - p = probability that j is HCT
a = the value to i of the compromise outcome minus two escalation costs
b = the value to i of j's concession minus one escalation cost
c = the value to i of i's concession

(1) p*(a) + (1-p)*(b) = c
(2) pa + b - pb = c
(3) pa - pb = c - b
(4) p*(a - b) = c - b
(5) p = (c - b)/(a - b)

*The subscripts on i's payoffs are dropped.
p = prior that i is LCT
1-p = prior that i is HCT
P(ple) = probability that escalating player is LCT. Call this Q
P((1-p)le) = probability that escalating player is HCT. This is 1.0.
a = value to j of i's concession minus two escalation costs (i is LCT)
d = value to j of i's concession minus three escalation costs (i is HCT)
b = the value to j of conceding to i's demands minus one escalation cost
*Subscripts on j's payoffs are dropped after the first equation

\[
\begin{align*}
\frac{P(ple)p}{P(ple)p + P((1-p)le)(1-p)} \ast (aj) + 1 - \frac{P(ple)p}{P(ple)p + P((1-p)le)(1-p)} \ast (dj) &= (bj) \\
\end{align*}
\]

\[
\begin{align*}
(1) \quad \frac{Q_p}{Q_p + (1-p)} \ast aj + 1 - \frac{Q_p}{Q_p + (1-p)} \ast dj &= bj \\
(2) \quad \frac{Q_{pa}}{Q_p + (1-p)} + \frac{d - Q_{pd}}{Q_p + (1-p)} &= b \\
(3) \quad \frac{Q_{pa} + d - Q_{pd}}{Q_p + (1-p)} &= b
\end{align*}
\]
\[
\begin{align*}
(4) \quad & \frac{Q_{pa} - Q_{pd}}{Q_p + (1-p)} = (b - d) \\
\end{align*}
\]

\[
\begin{align*}
(5) \quad & Q_{pa} - Q_{pd} = (b - d)Q_p + (1-p) \\
(6) \quad & Q_{pa} - Q_{pd} = Q_{pb} - Q_{pd} + (b - d)(1 - p) \\
(7) \quad & Q_{pa} - Q_{pd} - Q_{pb} + Q_{pd} = (b - d)(1 - p) \\
(8) \quad & Q_{pa} - Q_{pb} = (b - d)((1 - p) \\
(9) \quad & Q_p(a - b) = (1 - p)(b - d) \\
(10) \quad & Q = (1 - p)(b - d)/(p)(a - b)
\end{align*}
\]
Appendix 3
Model IV: A Low Cost Tolerant i’s Mixed Strategy

*Subscripts on player j’s payoffs, b, f, l, s, e, g, n, t, are dropped.

(1) \[ \begin{align*}
\text{ps}(x^*(b) + (1 - x)^*(f)) + p(1 - s)((x)^*(l) + (1 - x)^*(s)) &= \\
\text{ps}(x^*(e) + (1 - x)^*(g)) + p(1 - s)((x)^*(n) + (1 - x)^*(t))
\end{align*} \]

(2) \[ \begin{align*}
\text{ps}(bx + f + fx) + (p - ps)(lx + s - sx) &= ps(ex + g - gx) + (p - ps)(nx + t - tx)
\end{align*} \]

(3) \[ \begin{align*}
\text{psbx} + \text{psf} - \text{psfx} + \text{plx} + \text{ps} - \text{psx} - \text{pslx} - \text{ps} + \text{pssx} &= \\
\text{psex} + \text{psg} - \text{psg} + \text{px} + \text{pt} - \text{ptx} - \text{psnx} - \text{p} + \text{pst} + \text{p}
\end{align*} \]

(4) \[ \begin{align*}
(\text{psbx} - \text{psfx} - \text{pslx} + \text{pssx}) + (\text{plx} - \text{psx}) + (\text{psf} - \text{pss}) + \text{ps} &= \\
(\text{psex} - \text{psg} + \text{px} + \text{pt} - \text{ptx} + \text{psnx} - \text{p} + \text{p})
\end{align*} \]

(5) \[ \begin{align*}
(\text{psbx} - \text{psfx} - \text{pslx} + \text{pssx} - \text{psex} + \text{psgx} + \text{p} - \text{p}) + (\text{plx} - \text{psx} - \text{p} + \text{pt}) &= \\
(\text{psg} - \text{pst} - psf + \text{pss}) + (\text{pt} - \text{p})
\end{align*} \]

(6) \[ \begin{align*}
\text{psx}(b - f - l + s - e + g + n - t) + px(1 - s - n + t) &= ps(g - t - f + s) + p(t - s)
\end{align*} \]

(7) \[ \begin{align*}
px((s^*(b - f - l + s - e + g + n - t) + (1 - s - n + t)) &= p((s^*(g - t - f + s) + (t - s))
\end{align*} \]

(8) \[ \begin{align*}
x &= \frac{p((s^*(g - t - f + s) + (t - s))}{p((s^*(b - f - l + s - e + g + n - t) + (1 - s - n + t))}
\end{align*} \]
Appendix 4
Model IV: A High Cost Tolerant i’s Mixed Strategy

\[ x' = \text{probability that i escalates high} \]
\[ (1 - x') = \text{probability that i escalates low} \]
\[ p = \text{probability that i is low cost tolerant} \]
\[ (1 - p) = \text{probability that i is high cost tolerant} \]
\[ s = \text{probability that j is low cost tolerant} \]
\[ (1 - s) = \text{probability that j is high cost tolerant} \]
*Subscripts on player j’s payoffs, a, e, k, r, e, g, m, t are dropped.
**x’ and (1 - x’) are referred to as x and (1 - x) after the first equation.

(1) \[ (1 - p)(s)((x')(a) + (1 - x')(e)) + (1 - p)(1 - s)((x')(k) + (1 - x')(r)) = \]
\[ (1 - p)(s)((x')(d) + (1 - x')(h)) + (1 - p)(1 - s)((x')(n) + (1 - x')(v)) \]

(2) \[ (s - ps)(ax + ec - ex) + (1 - s - p + ps)(kx + r - rx) = \]
\[ (s - ps)(dx + h - hx) + (1 - s - p + ps)(nx + v - vx) \]

(3) \[ (sax + se - sex - psax - pse + psex) + (kx + r - rx - skx - sr + srz - pkx - pr + prx + psx + prs - pssx) = \]
\[ (sdx + sh - shx - psdx - psh + pshx) + (nx + v - vx - snx - sv + svx - pnx - pv + pvx + pnx + psx - psvx) \]

(4) \[ (-psax + pse + psx - psx - psx) + (sax - sex - skx + srz) + (kx - r) + (prx - pkx) + (se - sr) + (psr - pse) + (-pr + r) = \]
\[ (-psdx + psh + pnx + psx) + (sdx - shx - snx + svx) + (nx - vx) + (pvx - pnx) + (sh - sv) + (psv - psv) + (sv - svx) \]

(5) \[ (psex + psx - psx - psx - psx - psh + psn + psvx) + (sax - sex - skx + srz - sdx + shx + snx + svx) + (kx - r - nx + vx) + prx - pkx + pnx + pnx) = \]
\[ (sh - sv - se + sr) + (psv - psh + psr + pse) + (sv - sv) + (v - v) \]

(6) \[ psx(e + k - r - a + d - h - n + v) + sx(a - e - k + r - d - h + n - v) + x(k - r - n + v) + px(r - k - v + n) = \]
\[ s(h - v - e + r) + ps(v - h - r + e) + p(r - v) + (v - r) \]

\[ x' = \frac{p((s)(v - h - r + e) + (r - v)) + s(v - h - e + r) + (v - r)}{ps(e + k - r - a + d - h - n + v) + s(a - e - k + r - d - h + n - v) + p(r - k - v + n) + (k - r - n + v)} \]
Appendix 5
Model IV: Player j's Revised Estimates of i's Type

\[ P(\text{ple}) = \text{Probability that high escalating player is LCT. Call this w.} \]
\[ P((1-p)\text{le})) = \text{Probability that high escalating player is HCT. Call this z.} \]

*subscripts are dropped after equation.

\[
\begin{align*}
\frac{P(\text{ple}) \cdot p}{P(\text{ple}) \cdot p + P((1-p)\text{le})) \cdot (1-p)} & \quad \text{*(bj) + 1} & \frac{P(\text{ple}) \cdot p}{P(\text{ple}) \cdot p + P((1-p)\text{le})) \cdot (1-p)} & \quad \text{*(kj)} = \\
\frac{P(\text{ple}) \cdot p}{P(\text{ple}) \cdot p + P((1-p)\text{le})) \cdot (1-p)} & \quad \text{*(dj) + 1} & \frac{P(\text{ple}) \cdot p}{P(\text{ple}) \cdot p + P((1-p)\text{le})) \cdot (1-p)} & \quad \text{*(mj)}
\end{align*}
\]

\[
\begin{align*}
\frac{wp}{wp + z(1-p)} & \quad \text{*(b) + 1} & \frac{wp}{wp + z(1-p)} & \quad \text{*(k) =} \\
\frac{wp}{wp + z(1-p)} & \quad \text{*(d) + 1} & \frac{wp}{wp + z(1-p)} & \quad \text{*(m)}
\end{align*}
\]

\[
\begin{align*}
\frac{wpb + k - wpk}{wp + z(1-p)} & = \frac{wpd + m - wpm}{wp + z(1-p)}
\end{align*}
\]
\[(wpb + k - wpk)^*((wp + z(1-p)) = (wpd + m - wpm)^*((wp + z(1-p))\]

(1) dividing both sides by \(wp + z(1-p)\)
\[wpb + k - wpk = wpd + m - wpm\]

(2) collecting terms, and subtracting;
\[(wpb - wpk - wpd + wpm) = (m - k)\]

(3) taking out common terms,
\[wp (b - k - d + m) = (m - k)\]

(4) dividing by \(p(b - k - d + m)\)
\[
\frac{(m - k)}{p^*(b - k - d + m)}\]

(5) \(w = \frac{(m - k)}{p^*(b - k - d + m)}\)