General Deterrence and International Conflict: Testing Perfect Deterrence Theory

STEPHEN L. QUACKENBUSH

University of Missouri

Since general deterrence necessarily precedes immediate deterrence, the analysis of general deterrence is more fundamental to an understanding of international conflict than is an analysis of immediate deterrence. Nonetheless, despite a few exceptions, the quantitative literature has ignored the subject of general deterrence, focusing almost exclusively on situations of immediate deterrence. My purpose in this essay is to fill this evidentiary gap by subjecting a recently developed theory of general deterrence—Perfect Deterrence Theory—to a systematic test by examining general deterrence from 1816–2000. The results indicate that the predictions of perfect deterrence theory are strongly supported by the empirical record.

KEYWORDS conflict, deterrence, empirical testing of formal models, multinomial logit

Deterrence is the use of a threat (explicit or not) by one party in an attempt to convince another party not to upset the status quo. These threats have two purposes. The purpose of direct deterrence is to deter a direct attack on the defender. The goal of extended deterrence, conversely, is to deter attack on one’s allies (Snyder 1961:276–277). In addition, Morgan (1983) identifies two basic kinds of deterrence situations: immediate and general. Immediate deterrence “concerns the relationship between opposing states where at least one side is seriously considering an attack while the other is mounting a threat of retaliation in order to prevent it” (Morgan 1983:30). Classic

---

A previous version of this paper was presented at the 2002 annual meeting of the Peace Science Society, Tucson, AZ. I would like to thank Frank Zagare, Paul Senese, Paul Hensel, Sara Mitchell, Pat James, and Jerome Venteicher for helpful comments and suggestions.

Address correspondence to Stephen L. Quackenbush, Assistant Professor, Department of Political Science, 113 Professional Building, University of Missouri, Columbia, MO 65211, USA. E-mail: quackenbushs@missouri.edu
examples include the July Crisis of 1914, the Cuban Missile Crisis of 1962 and most, if not all, acute interstate crises. General deterrence, by contrast, “relates to opponents who maintain armed forces to regulate their relationship even though neither is anywhere near mounting an attack” (Morgan 1983:30). Thus, general deterrence has less to do with “crisis decisionmaking” than with everyday decisionmaking in conflictual or adversarial relationships.

Huth (1988) combines these two kinds of deterrence situations with the two types of deterrent threats to form four categories of deterrence: direct-immediate deterrence, direct-general deterrence, extended-immediate deterrence, and extended-general deterrence. Unlike many other works on deterrence, this paper focuses on direct-general deterrence.

The need for immediate deterrence indicates that general deterrence has previously failed (Danilovic 2001). If general deterrence succeeds, crises and wars do not occur. Since general deterrence necessarily precedes immediate deterrence, the analysis of general deterrence is more fundamental to an understanding of international conflict than is an analysis of immediate deterrence. Furthermore, because of selection effects, examining immediate deterrence without consideration of general deterrence can lead to misleading empirical results. As Fearon (2002:15) observes, “hypotheses that are valid for general deterrence should appear exactly reversed if we look at cases of immediate deterrence.”

Unfortunately, a divide exists between formal theories and the empirical analysis of deterrence. One reason for this divide is that while formal theories have focused on general deterrence, the quantitative literature has focused almost exclusively on immediate deterrence (Huth 1999). Important exceptions include Weede (1983), who examined extended deterrence in the cold war, and Sorokin (1994), who examined the impact of alliance formation on extended general deterrence within the Arab–Israeli conflict. Given their focus on extended deterrence, these studies are unable to provide insight on direct general deterrence. Huth and Russett (1993) did examine direct general deterrence within the context of enduring rivalries. They find that the balance of forces, arms races, and domestic conflict are all important determinants of deterrence outcomes. However, they do not directly test any specific theory of deterrence; indeed, formal theories of general deterrence have never been subjected to direct empirical testing.

Indeed, this problem has unfortunately plagued formal theories throughout political science (Green and Shapiro 1994), not only in the study of deterrence. This is unfortunate since deterrence theory appears to shed great light on the dynamics of deterrence, yet without extensive empirical analysis, one cannot know how well these theories explain real-world phenomena. One reason for this lack of testing is that, while selection of immediate deterrence cases has received a great deal of attention (for example, Huth and Russett 1984, 1988, 1990; Huth 1988), criteria for the selection of general deterrence cases have remained elusive.
This paper has two related purposes. The specific purpose is to fill this evidentiary gap by subjecting perfect deterrence theory—a recently developed theory of general deterrence—to a systematic test. I do so for several reasons. First, perfect deterrence theory (Zagare and Kilgour 2000) is supported by a formal logic with explicit theoretical expectations that facilitates empirical testing. Second, several preliminary tests of perfect deterrence theory have rendered promising, albeit provisional results (Senese and Quackenbush 2003; Quackenbush and Zagare 2006). And finally, as Huth (1999) points out, standard formulations of deterrence—to the extent that they have been explored empirically—are without compelling support.

The more general purpose is to develop the conceptualization and procedures to make such a test possible. This is necessary to bridge the divide between formal theories and quantitative analyses of deterrence. Key conceptualizations include case selection for direct general deterrence—I argue that identifying opportunity for conflict is the key. In addition, this paper offers the first direct test of incomplete information equilibrium predictions made by formal deterrence theory. Conducting such a test requires measurement of the utilities the actors have for the different outcomes that may emerge, so a modification of the measurement procedures developed by Bueno de Mesquita and Lalman (1992) is used. Also, since incomplete information equilibria depend on each state’s estimate of the opponent’s credibility, a nonlinear transformation technique is developed to estimate the credibility parameters.

To test perfect deterrence theory, I examine general deterrence from 1816–2000. After detailing the equilibrium predictions of perfect deterrence theory’s unilateral deterrence game, I more fully discuss the research design used to test them, including case selection, measurement of variables, and statistical method. In the next section, I discuss the empirical results and the theory’s ability to explain general deterrence and international conflict. The results indicate that perfect deterrence theory is well supported by the empirical record. Finally, I compare these findings to previous results supporting Bueno de Mesquita and Lalman’s (1992) international interaction game that is the basis of one of the most influential and important theories of interstate conflict.

UNILATERAL DETERRENCE GAME

Perfect deterrence theory is an axiomatically distinct theoretical alternative to classical deterrence theory, which focuses on ideas such as brinkmanship and mutual assured destruction (for example, Schelling 1960, 1966; Powell

\[1\] In addition, Danilovic’s (2002) empirical findings are consistent with the theory.
I focus on testing equilibrium predictions of the Unilateral Deterrence Game developed by Zagare and Kilgour (2000: chapter 5). Furthermore, unlike most previous tests of game-theoretic models, I focus on incomplete information conditions, since incomplete information is an important factor leading to war (Fearon 1995).

Figure 1 shows the structure of this game, which has two players, Challenger and Defender. At node 1, Challenger can choose whether to cooperate or defect: if he cooperates, the Status Quo remains unchanged; if he defects, Defender has an opportunity to respond. At node 2, Defender can choose whether to concede or defy. If she concedes, the outcome is Defender Concedes (DC), but if she defies, the next choice is Challenger’s. At node 3, Challenger can choose to concede, resulting in Challenger Defeated (CD), or defy, resulting in Conflict (DD).

Challenger’s utilities for outcome x are $c_x$, whereas Defender’s utilities are $d_x$, as indicated in Figure 1. Perfect deterrence theory highlights the

---

**FIGURE 1** Unilateral deterrence game. Challenger defects initially with probability $x$. Defender defies with probability $y$. 

---

2Perfect deterrence theory relies on three core models: the Unilateral Deterrence Game, the Generalized Mutual Deterrence Game, and the Asymmetric Escalation Game (Zagare and Kilgour 2000). The latter game focuses on situations of extended deterrence, and is therefore inappropriate for the current focus on direct deterrence. Furthermore, the unilateral deterrence game, unlike the mutual deterrence game, allows for differentiation between states within a dyadic relationship. Accordingly, I focus on testing equilibrium predictions of the Unilateral Deterrence Game.
importance of two variables: capability and credibility. A state’s threat is capable if the threatened party believes that it would be worse off if the threat were carried out than if it were not. For example, if Defender has a capable threat, then Challenger prefers Status Quo to Conflict.

Credible threats are believable, and in order to be believable, they must be rational to carry out (Zagare 1990). Therefore, a state that prefers conflict to backing down has a credible threat, and is said to be “Hard.” On the other hand, a state that would rather back down than fight has an incredible threat, and is said to be “Soft.”

EQUILIBRIUM PREDICTIONS

Zagare and Kilgour (2000) provide complete solutions and discussion of the Unilateral Deterrence Game. The equilibria are presented briefly in the appendix, as is an original analysis of the situation where Defender’s threat is not capable. These equilibria are the basic predictions of perfect deterrence theory to be tested. However, one cannot observe directly what equilibrium is present at any one time; only the outcome can be observed. Therefore, the equilibria must be further explicated to generate specific outcome predictions for the wide variety of conditions that may exist in reality. Three situations may arise, each generating different predictions.

The first condition that may exist is where Challenger prefers the Status Quo to Defender Concedes. Given no incentive to defect, Challenger will always cooperate, leading to the Status Quo. Secondly, if Challenger prefers Conflict to the Status Quo, then Defender lacks a capable threat. The solution presented in the appendix shows that in this case the outcome is Conflict if Defender is Hard, and if Defender is Soft the outcome is Defender Concedes.

Lastly, the most complicated scenario is when Challenger prefers Defender Concedes to Status Quo and Defender has a capable threat. In this case, only Certain Deterrence Equilibria make a unique outcome prediction (Status Quo) over their entire range of existence. Multiple outcomes emerge at intermediate (Separating Equilibrium) and low (Bluff and Attack Equilibria) levels of Defender credibility.

Under a Separating Equilibrium, the player choices are completely determined by their type. Therefore, Hard Challengers always defect at node 1. A Hard Defender also always defies, so if both players are Hard, the outcome is Conflict. However, a Soft Defender always concedes, so the outcome (with a Hard Challenger) is Defender Concedes. But if Challenger is Soft, he will always cooperate, resulting in Status Quo.

With an Attack Equilibrium, Challenger always defects initially, regardless of type. Therefore, the outcome reached is largely determined by Defender’s type. If Defender is Hard, she will defy and the outcome is Conflict if Challenger is Hard as well. But in the unlikely event that a Soft Challenger
faces a Hard Defender, then the outcome is *Challenger Defeated*. A Soft Defender always concedes, so the outcome is *Defender Concedes*.

Unfortunately, the Bluff Equilibrium does not provide unique outcome predictions for each pair of player types like the other equilibria do. A Hard Challenger always defects, as does a Hard Defender. Therefore, *Conflict* results if both players are Hard. However, a Soft Defender defies with some positive probability, so the outcome (with a Hard Challenger) will be either *Defender Concedes* (if Defender concedes) or *Conflict* (if Defender defies). A Soft Challenger also defects with some positive probability. So if a Soft Challenger faces a Hard Defender, the outcome will be either *Status Quo* (if Challenger cooperates) or *Challenger Defeated* (if Challenger defects initially). If both players are Soft almost anything can happen. If Challenger defects initially and Defender concedes, the outcome is *Defender Concedes*. But if Defender surprises Challenger by defying at node 2, then the outcome is *Challenger Defeated*.3 Finally, if Challenger cooperates, the outcome is *Status Quo*.

One additional equilibrium—Steadfast Deterrence Equilibrium—exists. As with the Certain Deterrence Equilibrium, *Status Quo* is the only possible outcome. However, this equilibrium exists at intermediate and low values of Defender credibility, and thus coexists with equilibria of other types. Therefore, through the Steadfast Deterrence Equilibrium, perfect deterrence theory predicts that the *Status Quo* is always possible as long as Defender’s threat is capable. Thus, Zagare and Kilgour’s (2000:149) claim that successful “deterrence could conceivably emerge under (almost) any conditions in a one-sided deterrence relationship.”

These three basic conditions—1) where Challenger prefers *Status Quo* to *Defender Concedes*, 2) where Defender lacks a capable threat, and 3) where Defender’s threat is capable—cover the entire range of possibilities. The outcome predictions for each scenario are summarized in Table 1 and serve as the equilibrium predictions to be tested in the empirical analysis to follow.

**RESEARCH DESIGN**

With the equilibrium outcome predictions now made, I turn to a presentation of the research design used to test these predictions. I will discuss the selection of cases, measurement of the dependent and independent variables, and the statistical method used in turn. The lack of quantitative testing of formal deterrence theory has been driven in large part by the difficulties presented by these issues. These conceptualizations are crucial not only for the specific task of testing perfect deterrence theory but also the more

---

3Since Challenger is Soft, he will always concede at node 3.
general task of making such tests possible and thereby somewhat bridging the divide between formal theories and quantitative analysis of deterrence.

### Case Selection

Case selection has been the biggest obstacle to the empirical analysis of general deterrence (Huth 1999). The only previous quantitative study focusing on direct general deterrence is by Huth and Russett (1993:63, emphasis added), who argue that “the population of enduring rivalries in the international system includes all dyadic relations in which a dispute created the possibility of one or both parties resorting to overt military force to achieve a gain or redress a grievance.” According to this line of reasoning, then, enduring rivalries are the proper cases for the study of general deterrence. Similarly, Diehl and Goertz focus attention on rivalries, whether or not they become enduring, and argue that “the rivalry approach provides a solution” (Diehl and Goertz 2000:91) to problems with deterrence case selection.

According to Diehl and Goertz (2000), a dyad is in a rivalry if they engage in a militarized interstate dispute. Thus, selecting all rivalries as cases of general deterrence would capture (by definition) all failures of

---

### TABLE 1 Equilibrium Outcome Predictions for the Unilateral Deterrence Game with Incomplete Information

<table>
<thead>
<tr>
<th>Existence conditions</th>
<th>Restrictions on challenger</th>
<th>Restrictions on defender</th>
<th>Predicted outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) ( c_{SQ} &gt; c_{DC} )</td>
<td>–</td>
<td>–</td>
<td>Status Quo</td>
</tr>
<tr>
<td>2) ( c_{DD} &gt; c_{SQ} )</td>
<td>–</td>
<td>( d_{DD} &gt; d_{DC} )</td>
<td>Conflict</td>
</tr>
<tr>
<td>3) ( c_{DC} &gt; c_{SQ} ) and ( c_{SQ} &gt; c_{DD} )</td>
<td>–</td>
<td>( d_{DC} &gt; d_{DD} )</td>
<td>Defender Concedes</td>
</tr>
<tr>
<td>Certain Deterrence Equilibrium</td>
<td>( p_{Def} \geq c_{t} )</td>
<td>–</td>
<td>Status Quo</td>
</tr>
<tr>
<td>Separating Equilibrium</td>
<td>( c_{s} \leq p_{Def} \leq c_{t} )</td>
<td>( c_{DD} &gt; c_{CD} )</td>
<td>Status Quo, Conflict</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( d_{DD} &gt; d_{DC} )</td>
<td>Status Quo, Defender Concedes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( d_{DC} &gt; d_{DD} )</td>
<td>Status Quo</td>
</tr>
<tr>
<td>Bluff Equilibrium</td>
<td>( c_{DD} &gt; c_{CD} )</td>
<td>( d_{DD} &gt; d_{DC} )</td>
<td>Status Quo, Conflict</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( d_{DC} &gt; d_{DD} )</td>
<td>Status Quo, Defender Concedes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( c_{CD} \geq c_{DD} )</td>
<td>Status Quo, Defender Concedes, Conflict</td>
</tr>
<tr>
<td>Attack Equilibrium</td>
<td>( c_{DD} &gt; c_{CD} )</td>
<td>( d_{DD} &gt; d_{DC} )</td>
<td>Status Quo, Defender Concedes, Challenger Defeated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( d_{DC} &gt; d_{DD} )</td>
<td>Status Quo, Defender Concedes, Challenger Defeated</td>
</tr>
</tbody>
</table>

Note: For definitions of threshold parameters \( c_{t} \), \( c_{s} \), and \( d_{n} \), see the appendix.
general deterrence. However, deterrence in dyads that have not fought would be ignored, which is particularly problematic because those are the dyads where deterrence has always worked. Furthermore, identification of the length of a rivalry, and thus the span of general deterrence, requires the assumption that the rivalry begins either with the first dispute (and thus deterrence failed when it was attempted for the first time), or at some arbitrary length of time before the first dispute and after the last.

Therefore, the rivalry approach is limited as a path to general deterrence case selection. However, one can safely assume that every state wishes to deter attacks against itself—this is the basic rationale for the maintenance of armed forces (Morgan 1983). This assumption is equivalent to the alliance portfolio literature’s assumption that every state has a defense pact with itself (Bueno de Mesquita 1975; Signorino and Ritter 1999). Hence, the difficult part of general deterrence case selection is not determining who makes deterrent threats (everyone does), but rather what states the threats are directed against. General deterrent threats are directed against any state that might consider an attack; these are states that have the opportunity for conflict.

Thus, the key to selecting cases of general deterrence is identifying opportunity for conflict. To identify cases where opportunity exists, I use the recently developed concept of politically active dyads (Quackenbush 2006a). A dyad is politically active “if at least one of the following characteristics applies: the members of the dyad are contiguous, either directly or through a colony, one of the dyad members is a global power, one of the dyad members is a regional power in the region of the other, one of the dyad members is allied to a state that is contiguous to the other, one of the dyad members is allied to a global power that is in a dispute with the other, or one of the dyad members is allied to a regional power (in the region of the other) that is in a dispute with the other” (Quackenbush 2006a:43). Quackenbush (2006a) finds that politically active dyads are able to identify opportunity as a necessary condition for international conflict, while previous measures of opportunity

---

4This is, of course, classifying militarized interstate disputes as general deterrence failures.
5This length of time could be assumed to be 10 years, for example. Thus, if two states last disputed in 1900, then the rivalry would continue to 1910. But if general deterrence is needed in 1910, why is it not needed in 1911 as well?
6In other words, every state threatens to respond if attacked, even though it may not actually carry out the threat if attacked (just as defense pacts are not always honored when tested).
7Thus, I argue that even “friendly” dyads, such as the United States and Canada, use general deterrence to regulate their relationship. Each has an incentive to prevent the other from challenging the status quo; indeed, deterrence has failed between the United States and Canada six times since 1974. Furthermore, since identifying opportunity for conflict is the key to general deterrence case selection, some may conclude that any study that uses politically relevant dyads (or politically active dyads) is an examination of general deterrence. While there is some merit to that view, most studies of international conflict do not have the same focus on the credibility and capability of threats that deterrence theory does, so such a blanket statement is not necessarily valid.
such as politically relevant dyads and regional dyads are unable to do so. Thus, we can have confidence that all politically active dyads could fight if they had the willingness to do so. The goal of deterrence is to dissuade other states from attacking the deterring state. In other words, states seek to ensure that other states—those with the opportunity to attack—do not gain the willingness to attack, and they do this through deterrence. While other empirically verified causes of war certainly impact deterrence outcomes, opportunity is the key to general deterrence case selection.

Some readers might be concerned that there is no attempt to determine whether the challenger actually intends to attack, and thus, whether the lack of an attack can meaningfully be considered general deterrence success. This issue does not actually pose a problem for the analysis here. The predictions being tested, from Table 1, are about particular game outcomes, not a dichotomous measure of deterrence “success” and “failure.” Furthermore, these predictions explicitly cover the case where the challenger has absolutely no interest in attacking. While the three non-status quo outcomes are essentially different categories of general deterrence failure, the status quo outcome is not necessarily the result of successful deterrence. For example, if Challenger prefers Status Quo to Defender Concedes, the only rational outcome is Status Quo. Although Challenger’s decision to not challenge the status quo in this case is not really “successful deterrence,” it is predicted by perfect deterrence theory.

For the cases that are selected (because they are politically active), I employ a directed-dyad-year unit of analysis. Within a directed dyad, the direction of interaction is important; for example, United States→Japan is one directed dyad and Japan→United States is another. The Unilateral Deterrence Game theoretically differentiates between the roles played by each state in a dyad, so employing directed-dyads enables empirical differentiation between the states in a dyad as well. Furthermore, since each state in a politically active dyad has the opportunity for conflict with the other, each state has an opportunity to challenge the status quo. Therefore, each state is considered to be a potential challenger. For example, in the United States→Japan directed dyad, the United States is Challenger and Japan is Defender, while in the Japan→United States directed dyad, Japan is Challenger and the United States is Defender.

Because most international relations data are based on annual observations, the year is the time period used for the cross-sectional time series data analysis conducted here. Therefore, each politically active directed-dyad-year constitutes an observation. Since general deterrence deals with the outbreak, rather than the continuation, of international conflict, I eliminate dyad-years marked by a conflict continuing from the previous year as well as ‘joiner’ dyads. Furthermore, I drop the directed dyad B→A in years with a continuing conflict in the directed
Testing Perfect Deterrence Theory

dyad A→B.\(^8\) This results in a total of 384,865 politically active directed-dyad-years for the time period 1816–2000.\(^9\)

Dependent Variable

The dependent variable for this analysis codes which outcome—Status Quo, Defender Concedes, Conflict, or Challenger Defeated—of the Unilateral Deterrence Game occurred for each observation. These game outcomes are determined by using version 3 of the militarized interstate dispute (MID) data set (Ghosn, Palmer, and Bremer 2004). If no dispute was initiated for a given directed dyad year, the deterrence outcome is coded as Status Quo.

The coding of the other deterrence outcomes is not as straightforward, but it is possible by utilizing information from both the outcome and the hostility level of the MID data set. The outcome variable of the MID data categorizes whether the outcome of a dispute is victory, yield, stalemate, compromise, released, unclear, or joins ongoing war. More information is needed, however, in order to capture the deterrence outcome rather than just the dispute outcome. This additional information is provided by the hostility level of each state in a MID, which is coded using a five-point scale. The hostility level indicates whether no militarized action was taken by that state, a threat to use force was made, a display of force, an actual use of force, or war.

Dispute initiation is equated to Challenger’s defection at node 1 of the unilateral deterrence game. However, to determine what outcome—Defender Concedes, Challenger Defeated, or Conflict—is reached following Challenger’s initial defection, it must be determined which player, if any, conceded. Unfortunately, the MID data do not indicate the sequence of events within a dispute. Nonetheless, the outcome and the hostility level variables can be used to determine what game outcome best corresponds to each dispute.

If both sides use force, or the outcome is “joins ongoing war,” the game outcome is coded Conflict. If mutual use of force did not occur, then a “victory by A” or “yield by B” is coded as Defender Concedes, whereas a “victory by B,” a “yield by A,” or a “released” outcome is coded as Challenger Defeated.\(^10\) If there was no mutual use of force and the outcome was “unclear,” “stalemate,” or “compromise,” then I compare hostility levels to determine the outcome.

---

\(^8\)For detailed discussions of directed-dyads and joiners, see Bennett and Stam (2000c).

\(^9\)Because of missing data, the number of observations actually used in the analyses is reduced further. EUGene, version 3.1 (Bennett and Stam 2000a) was used to construct the data set.

\(^10\)The released outcome is “identified whenever the seizure of material or personnel culminates with their release from captivity” (Jones, Bremer, and Singer 1996:180). In game terms, Challenger backed down in the end. Analyses (not reported here) reveal no significant differences between dropping released cases or including them.
If one state escalates to a higher hostility level than the other, it is reasonable to infer that the other state has conceded at some point rather than to defy by continuing to escalate. Thus, if Defender’s highest hostility level is lower than Challenger’s, then the outcome is Defender Concedes. But if Defender escalates to a higher hostility level, the outcome is Challenger Defeated. Finally, if both states escalate to the same hostility level, the outcome is coded Conflict.

Variables for the individual outcomes are combined into a single categorical variable. The resulting variable, called game outcome, is the dependent variable for the analyses to follow.

Independent Variables

In order to test the predictions summarized in Table 1 for each observation, player types and credibilities must be measured. This requires measurement of the utility that each player has for the various outcomes in each dyad year, a quite difficult task. Indeed, it has never been done with respect to deterrence theory.

The utilities for each outcome of a deterrence situation can be specified in terms of the costs and benefits of that outcome. For the two outcomes following a concession—Defender Concedes and Challenger Defeated—the utility comes from the size of concession and the cost associated with that outcome. The utility for Conflict is a lottery between receiving one’s demand and conceding the stakes to the other player, weighted by the probability of victory.

Following this logic, each player’s utility for each outcome can be expressed in mathematical form. Accordingly, Challenger’s utility for each outcome is operationalized as

\[ c_{SQ} = u_c(SQ) \]  
\[ c_{DC} = u_c(\Delta_c - \phi_p c) \] 

11Similar game outcome coding procedures have been employed by Bueno de Mesquita and Lalman (1992) and Bennett and Stam (2000b).

12Following Bueno de Mesquita and Lalman (1992:43), I assume that the cost of conceding “after absorbing a first-strike blow is larger, the greater the power of the rival:” \( \gamma_c(1 - p_c) \).

13I assume that domestic costs (\( \phi_p c \)) rise with the subjective probability of victory, since “the more powerful participant in a dyadic relationship bears a greater burden for finding a peaceful resolution of differences. Domestic populations are assumed to dislike bullying and to punish those who employ violence if their power should have made them persuasive enough to resolve their disputes without it” (Bueno de Mesquita and Lalman 1992:46).
Testing Perfect Deterrence Theory

and Defender’s utilities are operationalized similarly, where \( \Delta_x \) is the size of the demand made by state \( x \), \( p_x \) is the probability \( x \) wins, \( \phi_x \) is the domestic political cost associated with the use of force by \( x \), \( \gamma_x \) is the cost borne by \( x \) upon giving in, and \( \omega_x \) is the cost borne by \( x \) in a conflict.\(^{14}\)

Bueno de Mesquita (1975) developed a technique to measure state’s utilities based on their portfolios of alliance commitments. This technique has been refined over time, and the procedures outlined by Bueno de Mesquita and Lalman (1992: appendix 1) serve as the baseline for the measurement of the individual components of equations 1 through 4 used here. States’ utilities are measured using the similarity of foreign policy portfolios and risk attitudes (Bueno de Mesquita 1985). The probabilities of victory are estimated using relative power and estimates of the likelihood of intervention by third parties.\(^{15}\) In addition, several important changes have been made to their procedures; details of these changes are provided in the appendix, along with an explanation of the nonlinear method developed to measure each state’s credibility parameter.

Once the utilities, credibility parameters (\( p_{\text{Def}} \) and \( p_{\text{Ch}} \)), and threshold values (\( c_t \), \( c_s \), and \( d_n \)) are measured, the equilibrium outcome predictions for each politically active directed dyad-year are made according to the criteria specified in Table 1. This results in four dummy variables: \textit{predicted Status Quo}, \textit{predicted Defender Concedes}, \textit{predicted Challenger Defeated}, and \textit{predicted Conflict}. Each of these variables equals 1 if the corresponding outcome is predicted in equilibrium for the current observation, and 0 otherwise.

Since not all of the outcome predictions in Table 1 are unique, these variables are not mutually exclusive. That is, a value of “1” for \textit{predicted Status Quo} does not indicate that each of the other variables equals “0.” For those observations where multiple outcomes are predicted in equilibrium, the variable for each predicted outcome takes the value “1.” Therefore, each of these four variables will be included in the multinomial logit models;

\[ c_{\text{CD}} = u_c[\Delta_d - \gamma_c(1 - p_c)] \tag{3} \]
\[ c_{\text{DD}} = p_c[u_c(\Delta_c - \phi_c p_c - \omega_c(1 - p_c)) + (1 - p_c)[u_c(\Delta_d - \phi_c p_c - \omega_c(1 - p_c))] \tag{4} \]

\(^{14}\)These utilities for the various outcomes are equivalent to the utilities of some of the outcomes specified by Bueno de Mesquita and Lalman (1992). Specifically, \( c_{\text{SQ}} = U^A(SQ) \), \( c_{\text{DC}} = U^A(CapB) \), \( c_{\text{CD}} = U^A(CapA) \), and \( c_{\text{DD}} = U^A(WarA) \). The only exception is that no distinction is made here between the costs borne by the attacker and target in a conflict—they are assumed to be equal (\( \omega \)). Although Bueno de Mesquita and Lalman distinguished between the two cost terms theoretically, in their empirical testing the cost terms were assumed to be the same (\( \alpha = \gamma \)).

\(^{15}\)Complete details of the measurement procedures are provided by Bueno de Mesquita and Lalman (1992), as well as Bennett and Stam (2000a; 2005), who provide useful clarifications. I use the utility components drawn directly from Bennett and Stam’s EUGene software.
together, they serve as the primary independent variables for the prediction of game outcomes.

CONTROLS

These independent variables constitute the equilibrium outcome predictions of perfect deterrence theory. However, a number of alternative explanations of international behavior exist. Therefore, several control variables, representing the major foci of recent conflict studies, are used to test the robustness of the results obtained. Some of these variables (particularly relative power and S score) are components of the utility functions that form the basis of the equilibrium predictions to be tested. However, it is possible that these variables exert an independent influence on deterrence outcomes in addition to their influence on equilibrium predictions. Furthermore, since these individual components are combined in a different, nonlinear fashion within the unilateral deterrence game to generate equilibrium predictions, multicollinearity between the equilibrium predictions and the control variables is not present.\(^\text{16}\)

BALANCE OF FORCES

There is little dispute that relative power has an important effect on international conflict behavior. Most significantly, its importance has been highlighted in previous empirical studies of deterrence (for example, Huth 1988). It is therefore important to control for the effect of the balance of forces in the present analysis. I use the composite indicator of national capabilities from the Correlates of War project (Singer, Bremer, and Stuckey 1972) to measure military capabilities for each state. To determine the balance of forces in a dyad, I create a ratio of Challenger’s capabilities to the total capabilities of the dyad. The final variable, balance of forces, ranges from 0 (when Challenger is weak compared to Defender) to 1 (when Challenger is very strong compared to Defender).

S SCORE

Signorino and Ritter’s (1999) S score measures the similarity of foreign policy positions between states. Although S is an important component of the measurement of utilities employed here, it is possible that similarity in foreign policy positions has an independent effect on the outcome of international interactions. Therefore, I include it as a separate variable, which ranges

\(^{16}\text{Correlation between outcome predictions and control variables ranges from 0.002 to 0.25.}\)
from –1 to 1, with positive values indicating increasingly similar alliance portfolios and negative values representing increasingly dissimilar portfolios.

**Distance**

Geographic proximity has repeatedly been found (for example, Bremer 1992) to be an important predictor of international conflict. To control for the effects of proximity, the distance between states in a dyad is measured. I take the natural logarithm of the distance between capital cities, except for the USSR and United States when other cities are included, and states with land borders are considered to be zero miles apart (Bennett and Stam 2000a).

**Democracy**

The democratic peace literature (for example, Russett and Oneal 2001) has repeatedly found that democracies have seldom if ever fought one another. Furthermore, Schultz (2001) has found that democracy has an important effect on credibility. It is therefore important to control for the effect of joint democracy on deterrence outcomes. To measure democracy, I use the polity 2 variable, which ranges from −10 to 10, of the Polity IV data (Marshall and Jaggers 2002). I use the minimum democracy score for the dyad as a variable in the empirical analyses.

**Peace years spline**

The final control variable is of more a methodological than substantive character. Beck, Katz, and Tucker (1998) have argued that it is important for studies of conflict using pooled dyadic time series to account for time dependence within dyads. In other words, while the standard statistical assumption is that each observation is independent, observations of different years of the same dyad are not truly independent. I account for time dependence by employing Beck, Katz, and Tucker’s method of including peace years and three cubic spline variables that account for time dependence.

Statistical Method: Multinomial Logit

Because the dependent variable, *game outcome*, has more than two unordered categories, the appropriate statistical method to use is multinomial logit. Since there are four categories, three equations are estimated. These equations independently estimate the effects of the independent variables on producing outcomes *Defender Concedes*, *Challenger Defeated*, and *Conflict* relative to the *Status Quo*. 
Two alternative models are available for empirical analysis of the multiple, unordered categories of the dependent variable employed here. The first of these is strategic probit, which Signorino and Tarar (2006) utilized for a unified theory and test of extended immediate deterrence. Strategic probit provides a method for integrating the theoretical and empirical models, while accounting for nonmonotonic influences of independent variables resulting from strategic interaction (Signorino 1999, 2003). Thus, strategic probit would provide an excellent avenue for empirically determining the impact of factors such as balance of forces, arms races, and domestic conflict on general deterrence. However, since strategic probit uses a different solution concept (quantal response equilibrium vs. perfect Bayesian equilibrium) and different information conditions (complete vs. incomplete information) than perfect deterrence theory, an empirical analysis using strategic probit would not be a test of perfect deterrence theory’s equilibrium predictions, which is the purpose of this paper.

The second alternative would be a nested logit model. Nested logit models are models of a decision process that is made in stages in which the decisions at later stages are limited by those made in earlier stages. This seems to fit nicely with the game structure examined here, where each node of the Unilateral Deterrence Game would correspond to a different stage of the nested logit model. Unfortunately, nested logit is also not appropriate here, for two primary reasons. First, as with the strategic probit, it would not provide a test of the equilibrium predictions of perfect deterrence theory. Secondly, nested logit models are unable to account for strategic interaction between multiple decisionmakers, and are thus only appropriate for analyses of a single decisionmaker’s nested decisions (Signorino 2003).

While the multinomial logit model is also a decision model that does not account for strategic interaction, that does not pose a problem for this analysis. Strategic interaction between Challenger and Defender is taken into account through perfect deterrence theory, which makes equilibrium outcome predictions prior to the empirical analysis. Thus, the theory is the “decisionmaker” (deciding which outcome(s) is predicted for each observation), and the purpose of this analysis is to test how well the theory makes those decisions.

RESULTS

I begin testing the equilibrium predictions of perfect deterrence theory outlined in Table 1 through a base model—incorporating only the four outcome prediction variables and the peace years spline variables—shown in Table 2. This model fits the data significantly better than a model including only the constant term.
Beginning with predicted Status Quo, one can see that the coefficient for this variable is insignificant for all but the Conflict equation. This indicates that predicting the status quo has no significant impact on the occurrence of Defender Concedes or Challenger Defeated, while it increases the probability of Conflict. This latter result is quite surprising; if robust, this finding would be problematic for the theory.

The impact of each of the non-status quo outcome predictions on the occurrence of game outcomes is in line with expectations. The prediction of Defender Concedes makes Defender Concedes significantly more likely to occur, and significantly increases the probability of Challenger Defeated and Conflict. Similarly, predicted Challenger Defeated makes all non-status quo outcomes, most importantly Challenger Defeated, more likely to occur. The prediction of Conflict also has a significant impact in each equation. In other words, if Conflict is predicted, all non-status quo outcomes (most importantly, Conflict) are more likely to occur.

To examine the robustness of these results against alternative explanations, the full multinomial logit model was run including the control variables. The full model, shown in Table 3, produces a highly significant fit to the data. The results of the full model indicate that the predictions of perfect deterrence theory are quite robust.

Unlike the base model, predicted Status Quo is found to reduce the probability of any non-status quo outcome, although the coefficient for the Challenger Defeated equation is not significant. Thus, although the results for this variable were somewhat contrary to expectations in the base model, after other factors are controlled for the expected pattern emerges.
The other key variables continue to be positive and significant across all three equations. Therefore, perfect deterrence theory’s predictions of Status Quo, Defender Concedes, Challenger Defeated, and Conflict are all strongly supported by the empirical record, and these predictions are robust against alternative explanations.

The results for the control variables are mostly in line with previously established expectations. The balance of forces has a positive coefficient throughout, although it is not significant for the Challenger Defeated equation. Distance, democracy, and peace years all reduce the probability of each non-status quo outcome, although democracy is surprisingly not significant for the Challenger Defeated equation. The S score only has a significant impact on Challenger Defeated, which is in the expected negative direction.

These results for the prediction of game outcomes are more easily understood through examination of predicted probabilities. The value for the predicted outcome is set to 1, the values for the other outcome predictions is set to 0, and the control variables are set to their means.
Table 4 shows the impact of outcome predictions on the predicted probabilities of outcomes. To ease interpretation, the *Status Quo* column of Table 4 shows the predicted probability of each outcome given a prediction of *Status Quo*. The other columns show the change in the predicted probability of a given outcome when a non-status quo outcome is predicted. For example, predicting *Defender Concedes* decreases the probability of *Status Quo* by 0.19%, increases the probability of *Defender Concedes* by 85.71%, and increases the probability of *Challenger Defeated* and *Conflict* by 50.00% each. Prediction of any non-status quo outcome reduces the probability of the *Status Quo*, although not by the largest of margins. The prediction of a concession by Defender greatly increases the probability of a *Defender Concedes* outcome (an increase of over 85%), the prediction of a defeat for Challenger greatly increases the probability of *Challenger Defeated* (an increase of 300%), and the prediction of *Conflict* has a large impact on the probability of *Conflict* (an increase of 250%). Each of these results is in line with expectations.

Astute readers will note that although predictions of *Defender Concedes* and *Conflict* have their largest impact on the appropriate outcome, they also greatly increase the probability of other non-status quo outcomes. This should actually be expected, since each of the equations is a comparison between a non-status quo outcome and the status quo. When a non-status quo outcome is predicted, it should make the status quo less likely to occur. If the status quo is less likely, then all non-status quo outcomes (even those not specifically predicted) are more likely.17

---

**Note:** These predicted probabilities were calculated using the results of the full model (Table 3). The value for each predicted outcome was set to 1, the values for the other outcome predictions set to 0, and the control variables were set to their means.

17Thus, the key results are the direction and significance of the outcome predictions on those outcomes (e.g., of predicted *Defender Concedes* in the *Defender Concedes* equation). Note that each of these coefficients is positive and highly significant.
Nonetheless, prediction of *Challenger Defeated* has a larger impact on the *Defender Concedes* and *Conflict* outcomes than on the *Challenger Defeated* outcome. This indicates that the model has some difficulty in differentiating between concessions by Challenger and Defender. However, Wald tests for combining outcomes confirm that the model is able to distinguish between each of the outcomes, again significant at the <0.0001 level.

A further test of the model’s empirical fit is provided by block tests of statistical significance, as shown in Table 5. These tests clearly demonstrate that the full model provides a better fit than the null model. Furthermore, removing any single equilibrium prediction produces a significantly worse fit to the data than the full model, as does removing the set of all four equilibrium predictions. Thus, we can have confidence that each outcome prediction of perfect deterrence theory provides an important contribution to explaining the dynamics of deterrence.

One unusual feature of this test of perfect deterrence theory is that the theory does not provide unique outcome predictions in most cases; rather, because of multiple equilibria, two or three different outcomes are possible in many observations. Table 6 helps to clarify the empirical implications of

| TABLE 5 Block Tests of Statistical Significance |
|------------------|----------|----------|-------------|
| Variable removed     | df       | Log-likelihood | Probability     |
| (Null Model, constant only) | –       | 0.153,372.32  | –               |
| (Full Model)         | 36       | 0.136,088.77  | <0.0001 (vs. null model) |
| All 4 equilibrium variables | 12     | 0.139,388.39  | <0.0001 (vs. full model) |
| Predicted Status Quo | 3        | 0.136,13.57   | 0.022           |
| Predicted Defender Concedes | 3     | 0.136,398.44  | <0.0001         |
| Predicted Challenger Defeated | 3    | 0.137,479.94  | <0.0001         |
| Predicted Conflict   | 3        | 0.136,99.36   | <0.0001         |

| TABLE 6 Comparison of Predicted and Observed Outcomes |
|----------------|----------|-----------|---------------|
| Predicted outcome(s) | Observed outcome | Percent correct |
| | SQ | DC | CD | DD | Total | Non-SQ |
| SQ | 282,875 | 633 | 158 | 466 | 284,132 | 99.56 | – |
| SQ, DC | 21,391 | 110 | 21 | 41 | 21,563 | 99.71 | 63.95 |
| SQ, DC, CD | 3,435 | 51 | 10 | 51 | 3,547 | 98.56 | 54.46 |
| SQ, DC, DD | 5,750 | 30 | 12 | 25 | 5,817 | 99.48 | 82.09 |
| SQ, CD | 6,698 | 59 | 14 | 28 | 6,799 | 98.67 | 13.86 |
| SQ, DD | 11,503 | 53 | 15 | 31 | 11,602 | 99.82 | 31.31 |
| DC | 36,377 | 148 | 31 | 78 | 36,634 | 0.40 | 57.14 |
| DD | 14,683 | 54 | 17 | 17 | 14,771 | 0.12 | 19.32 |
| Total | 382,712 | 1,138 | 278 | 737 | 384,965 | 86.26 | 48.66 |

Note: Outcomes are abbreviated as follows: SQ: Status Quo; DC: Defender Concedes; CD: Challenger Defeated; DD: Conflict.
these nonexclusive equilibrium predictions. Each row shows the number of observations that correspond to each group of predictions from Table 1, separated by the outcome actually observed. Thus, for example, there are 21,563 observations where the predicted outcome is either Status Quo or Defender Concedes; of these, 21,391 actually result in the status quo, and a further 110 result in a concession by Defender, for a total accuracy rate of 99.71%. One can see that each of the predictions is about 99% accurate, except for the unique predictions of Defender Concedes and Conflict, which are less than 1% accurate.

Clearly, the accuracy of these latter predictions is problematic. As Table 1 makes clear, these unique predictions of Defender Concedes or Conflict occur when Challenger prefers Conflict to Status Quo. In that case, Challenger’s only rational course of action is to challenge the status quo, leading to one of those outcomes. This is the prediction of perfect deterrence theory; however, it must also be the prediction of any theory, since that is the only logical result of those preferences. This result is clearly a product of error in the measurement of utilities. For instance, despite the maintenance of the status quo throughout, Conflict is the predicted outcome for the United States→Canada dyad for the period 1920–1941, because throughout, the United States is measured to prefer Conflict to the Status Quo. While the United States almost certainly did not truly have that preference, it is the best estimate produced by the measurement of utilities employed here, based on the method pioneered by Bueno de Mesquita and Lalman (1992). Bueno de Mesquita and Lalman (1992:298) acknowledge that the procedures they developed “may introduce considerable measurement error,” concluding that this “suppresses . . . our results.” Hence, future research should seek to improve our measurement of utilities; once that is done, one would expect support for perfect deterrence theory to be even stronger.

Since non-status quo outcomes are rare events, one could achieve 99% accuracy by simply predicting the Status Quo for every observation. However, in so doing, one would achieve 0% accuracy of non-status quo predictions. As one can see from Table 6, the non-status quo predictions here are about 49 percent accurate (with the accuracy of individual predictions between 14 and 82 percent accurate). Thus, perfect deterrence theory provides a proportional reduction in the error of non-status quo predictions of 20 percent. Although improvement is always possible, these results nonetheless support perfect deterrence theory.

These results for perfect deterrence theory can be put in perspective by comparing them with an alternative theory of international conflict. Bueno de Mesquita and Lalman (1992) constructed the International Interaction Game to analyze relations between states. Although this is not the way they present it, theirs is essentially a theory of general deterrence. Furthermore, it has been extensively tested by Bennett and Stam (2000b) in
a multinomial logit analysis quite similar to that conducted here. Therefore, it is useful to compare the results found here to the results obtained by Bennett and Stam.

Only five outcomes emerge as equilibrium predictions of the International Interaction Game: status quo, acquiescence by A, acquiescence by B, negotiation, and war. However, six different outcomes of the IIG—status quo, acquiescence by B, negotiation, capitulation by A, capitulation by B, and war—are realized. Thus, one equilibrium prediction (acquiescence by A) never occurs in reality, and two other outcomes that do occur (capitulation by A and capitulation by B) are never predicted in equilibrium. In the Unilateral Deterrence Game, however, all four possible outcomes occur in both predictions and the historical record.

In addition, the multinomial logit results obtained here provide much stronger support for perfect deterrence theory than the support Bennett and Stam (2000b) found for the International Interaction Game. As Bennett and Stam (2000b:473) note, only two of the equilibrium predictions “are helping to differentiate non-status quo outcomes from the status quo.” Compare this to the Unilateral Deterrence Game, where each outcome prediction helps to differentiate between the status quo and non-status quo outcomes. If perfect deterrence theory is more consistent with the empirical record than “one of the most important theories of international conflict” (Bennett and Stam 2000b:451), then one must take note.

CONCLUSIONS

A divide exists in the literature between formal theories and quantitative analyses of deterrence, largely because formal theories of general deterrence have never been subjected to direct empirical testing. This paper has sought to bridge this divide in two ways.

First was to develop the conceptualization and procedures to make such direct empirical testing possible. Key conceptualizations include case selection for direct general deterrence—I argue that identifying opportunity for conflict is the key. In addition, procedures to measure the actors’ utilities for the different outcomes were developed. Also, since incomplete information equilibria depend on each state’s estimate of the opponent’s credibility, a nonlinear transformation technique was developed to estimate the credibility parameters.

Using these procedures, I have endeavored to fill a large void in the literature by testing perfect deterrence theory, focused on the theory’s Unilateral Deterrence Game with incomplete information. I used multinomial logit methods to examine the prediction of particular game outcomes. The results indicate that the theory is well supported by the historical record.
The results of this research have important policy implications as well. The policy recommendations of perfect deterrence theory and classical deterrence theory are diametrically opposed on many issues (Zagare 2004). For example, classical deterrence theory argues that national missile defense undermines the stability of deterrence (for example, Powell 2003), whereas an application of perfect deterrence theory demonstrates that national missile defense enhances deterrence stability (Quackenbush 2006b). However, policy discussions in academia and government are generally based on classical deterrence theory. Given its strong empirical support, coupled with logical and empirical limitations of classical deterrence theory (Zagare 1996), perfect deterrence theory would provide a much better basis for analyzing various aspects of national security policy. This could be done by applying perfect deterrence theory to case studies of particular events (for example, the analysis of the war in Kosovo by Quackenbush and Zagare 2006) or to the analysis of particular policy issues (for example, the analysis of national missile defense by Quackenbush 2006b).

In addition, perfect deterrence theory provides a natural basis for further empirical research on the dynamics of deterrence. First, future research focused on improving measurement of utilities would be particularly pertinent. Also, direct tests of further formal theories—not only theories of deterrence—are of vital importance for improving our understanding of international interactions. In addition, this analysis was focused on only one of the theory’s three core models, applied across all observations. Thus, this analysis may have underestimated the theory’s explanatory power. One avenue for future research is therefore to conduct a more nuanced test incorporating models of mutual and extended deterrence. And finally, further theoretical expansions of perfect deterrence theory—such as an incorporation of more nuanced bargaining elements like endogenous stakes—would greatly improve our understanding of the dynamics of deterrence.

REFERENCES


APPENDIX

I briefly detail the equilibria and measurement details here; complete discussions are provided in an online appendix at http://web.missouri.edu/~quackenbushs/

Unilateral Deterrence Game Equilibria

Zagare and Kilgour (2000) model incomplete information by allowing each player to be uncertain of the other player’s type. Challenger estimates the probability that Defender is Hard, \( p_{\text{Def}} \), and Defender estimates the probability that Challenger is Hard, \( p_{\text{Ch}} \) (where \( 0 < p < 1 \) in each case). They find four major types of perfect Bayesian equilibria. Each equilibrium specifies values for four important parameters: \( x_H \) (\( x_S \)) = probability that a Hard (Soft) Challenger defects at node 1, and \( y_H \) (\( y_S \)) = probability that a Hard (Soft) Defender defies at node 2. Certain Deterrence equilibria exist where \([x_H, x_S; y_H, y_S] = [0, 0; 1, \ast]\) when \( p_{\text{Def}} \geq c_t \); Steadfast Deterrence equilibria exist where \([x_H, x_S; y_H, y_S] = [0, 0; 1, u]\) when \( p_{\text{Def}} < c_t \); Separating equilibria exist where \([x_H, x_S; y_H, y_S] = [1, 0; 1, 0]\) when \( c_s \leq p_{\text{Def}} \leq c_t \); Bluff equilibria exist where \([x_H, x_S; y_H, y_S] = [1, v; 1, u]\) when \( p_{\text{Def}} < c_s \) and \( p_{\text{Ch}} < d_n \); Attack equilibria exist where \([x_H, x_S; y_H, y_S] = [1, 1; 1, 0]\) when \( p_{\text{Def}} < c_s \) and \( p_{\text{Ch}} \geq d_n \). The region of existence for each equilibrium is defined by a combination of three critical points. These critical points, defined below, are necessary to empirically generate the equilibrium predictions tested here:

\[
\begin{align*}
    c_t &= \frac{c_{DC} - c_{SQ}}{c_{DC} - c_{DD+}}, \\
    c_s &= \frac{c_{DC} - c_{SQ}}{c_{DC} - c_{CD}}, \\
    d_n &= \frac{d_{CD} - d_{DC}}{d_{CD} - d_{DD+}}.
\end{align*}
\]

Measurement Details

Three important changes are made here to Bueno de Mesquita and Lalman’s measurement procedures. First, I use Signorino and Ritter’s (1999) measure of foreign policy similarity, \( S \), as the basic building block for these measurements of utility. Secondly, I constrain the political cost term to be non-negative to prevent this “cost” term from providing an artificial boost to the utility for Conflict. Therefore, I use the equation \( \phi_k = \max[u_k(SQ), 0] \).

The third important change in Bueno de Mesquita and Lalman’s measurement techniques made here deals with the costs borne by a state in giving in (\( \gamma_S \)) compared to the costs borne by a state in conflict (\( \omega_C \)). While Bueno de Mesquita and Lalman simply assume that \( \gamma_S = \omega_C = 1 \), this sets the cost of backing down equal to the cost of fighting. Unfortunately, this assumption makes Conflict preferred to concession in almost all circumstances. In other
words, if it is assumed that $\gamma_x = \omega_x$, then almost all possible threats are credible. Clearly this is a problem; states often prefer to back down rather than fight, since the costs borne by a state in giving in are generally less than the costs associated with conflict. Therefore, it is assumed here that $\gamma_x = 0.25$ and $\omega_x = 1$.

With these modifications to the procedures laid out by Bueno de Mesquita and Lalman (1992), each state’s utilities for each outcome are measured. The individual components of these utilities are calculated using EUGene (Bennett and Stam 2000a). These raw utility values (which can range from −3 to 2) are normalized to the interval from 0 to 1.

Since I am testing predictions based on incomplete information conditions where player types are not common knowledge, Challenger’s estimate of the probability that Defender is Hard ($p_{\text{Def}}$) and Defender’s estimate of Challenger’s type ($p_{\text{Ch}}$) are needed. I develop a nonlinear transformation of the difference between the utility for conflict and a concession, which I call $\Delta U = U(\text{conflict}) - U(\text{back down})$, to the interval between 0 and 1 while also accounting for the impact of risk propensity. These credibility estimates for each player are measured as

\[
p_{\text{Ch}} = \left[ \frac{e^{3(c_{\text{DD}} - c_{\text{CD}})}}{1 + e^{3(c_{\text{DD}} - c_{\text{CD}})}} \right]^{(1 + R_{\text{Ch}})}, \quad p_{\text{Def}} = \left[ \frac{e^{3(d_{\text{DD}} - d_{\text{DC}})}}{1 + e^{3(d_{\text{DD}} - d_{\text{DC}})}} \right]^{(1 + R_{\text{Def}})}.
\]