Strategic Interaction and the Statistical Analysis of International Conflict
Author(s): Curtis S. Signorino
Source: *The American Political Science Review*, Vol. 93, No. 2 (Jun., 1999), pp. 279-297
Published by: American Political Science Association
Stable URL: http://www.jstor.org/stable/2585396
Accessed: 10-12-2015 17:18 UTC

Strategic Interaction and the Statistical Analysis of International Conflict

CURTIS S. SIGNORINO *University of Rochester*

Although strategic interaction is at the heart of most international relations theory, it has largely been missing from much empirical analysis in the field. Typical applications of logit and probit to theories of international conflict generally do not capture the structure of the strategic interdependence implied by those theories. I demonstrate how to derive statistical discrete choice models of international conflict that directly incorporate the theorized strategic interaction. I show this for a simple crisis interaction model and then use Monte Carlo analysis to show that logit provides estimates with incorrect substantive interpretations as well as fitted values that can be far from the true values. Finally, I reanalyze a well-known game-theoretic model of war, Bueno de Mesquita and Lalman's (1992) international interaction game. My results indicate that there is at best modest empirical support for their model.

Historians from Thucydides (1972) to Taylor (1954) to Kissinger (1994) and Kagan (1995) all detail the machinations of state leaders trying to achieve their foreign policy goals through sometimes peaceful but often violent means. Not surprisingly, most international relations theories—especially those of international conflict—assume that states behave strategically. One might go so far as to say that strategic interaction is the defining characteristic of international relations.

In its most general form, strategic choice requires only that a state (or substate actor) (1) have goals and (2) in attempting to achieve them condition its behavior on the expected behavior of others. For jointly strategic interaction between two or more states, there also must be some intersubjective recognition of each's options and goals and subsequent interdependent conditioning of behavior. These goals can be as (not so) simple as mere survival or as grand as world domination. The importance of strategic interaction lies in the *interdependence* of decisions it creates among the states in the international system.[1] States do not act in a vacuum. Decisions to engage in arms production, enter into alliances, and go to war are not independent of the expected behavior of the other states in the system. Their calculations are based on what they expect other nations are currently planning to do or how they may respond to particular actions.

A wide array of theories falling under the realist, neorealist, and neoliberal rubrics make the above general assumptions. In their various manifestations, theories about power balancing, war, deterrence, alliances, crisis bargaining, and cooperation all assume that state leaders behave strategically. A large subset of these have outcomes that can be thought of as resulting from the strategic interaction of states who make choices over discrete sets of options. Examples of strategic discrete choices include whether to use force against another state; whether to give in to a deterrent; and whether to balance, bandwagon, or remain neutral in the face of a rising power.

Whether specified in prose or in math, the purpose of such theories generally is to make explicit the structure of the often very complex strategic interaction and to explain the outcomes we observe.[2] Yet, if game theory has taught us anything, it is that the likely outcome of such situations can be greatly affected by the sequence of players' moves, the choices and information available to them, and the incentives they face. In short, in strategic interaction, *structure matters*. Because of this emphasis on causal explanation and strategic interaction, we would expect that the statistical methods used to analyze international relations theories also account for the structure of the strategic interdependence. Such is not the case. This article is an attempt to remedy that—to build a bridge between international conflict models and our statistical testing of those models.

In the next section, I address more fully the problem of traditional methods used to analyze discrete data, especially binomial logit and probit. I then demonstrate how to use a game-theoretic solution concept to derive statistical strategic discrete choice models of international conflict. To help make the concepts more concrete, this technique is applied to a simple crisis game very similar in structure to a number of models in the international relations literature. Using the crisis model, I generate Monte Carlo data and examine the (in)ability of logit to model strategic interaction. Fol-

Curtis S. Signorino is Assistant Professor of Political Science, 310 Harkness Hall, University of Rochester, Rochester, NY 14627 (sign@troi.cc.rochester.edu).

This article has greatly benefited from comments provided in a number of fora: the 1998 annual meeting of the International Studies Association, the 1998 Duke University conference on "Empirically Testing Theories of International Relations," the 1997 Political Methodology Summer Conference, the 1997 annual meeting of the Midwest Political Science Association, and the Harvard Government Department's Rational Choice Discussion Group. My thanks to Neal Beck, Bruce Bueno de Mesquita, Jim Fearon, John Freeman, Chris Gelpi, Gary King, Lisa Martin, Jim Morrow, Jeff Ritter, Ken Shepsle, Alastair Smith, and Renée Smith for valuable comments, to Richard McKelvey and Tom Palfrey for answering numerous questions, to Bruce Bueno de Mesquita and David Lalman for providing their data, to John Duggan and Mark Fey for several helpful discussions, and to Branislav Slantchev for research assistance. This article previously appeared as Center for International Affairs Working Paper 98-4, Harvard University. Support from the Harvard Center for International Affairs was provided during the writing of this article.

[1] See also Smith 1997, 1–3.

[2] This stands in contrast to models whose primary purpose is to "fit" the data or forecast well, without regard to causal explanation.

lowing that, I reanalyze a well-known game-theoretic model of war, the international interaction game of Bueno de Mesquita and Lalman (1992). My results indicate that the international interaction game does not explain the variation in their data as well as originally claimed. I conclude by identifying a number of important areas for future research and suggest that structure may be a double-edged sword for positive international relations theory.

THE PROBLEM WITH TRADITIONAL METHODS OF ESTIMATION

Until recently, international relations researchers using quantitative methods had to make do with highly aggregated event data on phenomena such as war occurrence and military disputes. More than two decades ago, the Correlates of War project provided war occurrence data that was aggregated over all countries in the international system and over five-year periods. In the 1980s, the data were disaggregated to annual counts over the entire system. Because of the type of data available during this period, quantitative analyses of war occurrence generally relied on count models, most commonly the Poisson, but sometimes more flexible models were used, such as the negative binomial, hurdle Poisson, or generalized event count (King 1989a, 1989b, 1989c; Martin 1992).

More recently, the various event data sets have been disaggregated to the nation-year, dyad-year, or monad-dispute level. Typically, the disaggregation increases by at least an order of magnitude the number of observations available in a data set. One selling point when new (i.e., disaggregated) data are used is that inferences made from these data sets are often thought to be better, if only because of the increased number of data points. Indeed, more observations generally *are* better. For the more recent (primarily binary) event data, however, most international relations scholars automatically reach for logit or probit models (see, e.g., Bremer 1992; Bueno de Mesquita and Lalman 1988, 1992; Fearon 1994; Hagan 1994; Huth 1990; Huth, Gelpi, and Bennett 1993; Kim and Bueno de Mesquita 1995; Kim and Morrow 1992; Maoz and Russett 1993; Morgan and Campbell 1991; Morrow 1991; Raymond 1994; Russett 1993).

Consider the typical use of logit to analyze what is assumed to be strategic interaction, generally among more than two states.[3] The data are organized according to annual dyads of states for a given year. Depending on the study, the binary dependent variable may denote whether the two states engaged in war that year or whether deterrence was a success or failure. The independent variables may include any mix of the following: whether one or both states were democracies, whether they were allies, the extent of mutual trade, the balance of military capabilities, whether one state was rising in capabilities, whether one or both were major powers, whether they were geographically contiguous, calculated expected utilities for war, or a calculated probability of war (see previous citations). With the dependent and independent variables specified, maximum likelihood estimation is conducted using logit. The effects of the regressors on the probability of war or deterrence are analyzed, and the model may be compared to others to test which is better at explaining the outcomes.

We should be wary of such analysis for at least two reasons.[4] First, if observed actions are the result of (perhaps complex) strategic interaction, then it is unlikely that a simple logit functional form will capture the structure of that strategic interdependence—that is, the set of states interacting, their sequence of decisions, their options at decision points, the factors that influence their incentives, and the equilibrium effects of this interdependence on outcomes. Researchers have attempted to get around this by testing observable implications of theories using such variables as the concentration of military capabilities, indices of power transition, expected utilities, or expected utility-based probabilities of war. I will later show that a logit functional form with such variables is still unlikely to account for the structure and endogeneity of choices for even a relatively simple two-nation strategic model.

Second, if the interaction involves $N > 2$ states at a time, then not only is the strategic interaction reduced to a logit functional form, but also the N-nation interaction is also broken into dyads. In the typical application of logit to models of conflict, observations are assumed to be independent, conditional on the explanatory variables. Since time-series and cross-sectional interdependence are generally never incorporated, each dyad is assumed to be completely independent of any other.[5] In effect, each N-nation interaction becomes $N(N - 1)/2$ independent observations, greatly expanding the size of the data set without adding any additional information to it.

In sum, logit models of international conflict are unlikely to capture the real or theorized structure of strategic interaction. Moreover, as implemented, the independence assumptions of the statistical models are often inconsistent with strategic interdependence assumptions of the theories. Indeed, these criticisms apply not only to analyses of international conflict but also to logit and probit analyses of *any* phenomenon involving strategic interaction in international relations, comparative politics, or American politics. Because of this, we should expect, and I will later show, that logit analysis of strategic interaction can lead to parameter estimates with wrong substantive interpretations: Fitted values and predictions of outcome probabilities can be grossly incorrect, as can calculations of the effects of variables on the changes in outcome probabilities. We might also conjecture that breaking the data set into dyads, thus possibly expanding it without adding information, will affect standard errors

[3] Since my argument applies equally to both logit and probit, I henceforth refer only to logit.

[4] See also Smith 1997 for similar, independently developed arguments.

[5] For recent research on time-series cross-sectional analysis with binary data, however, see Beck, Katz, and Tucker 1998.

of estimates. I leave that question to future research. In the rest of this article, I assume that strategic interaction exists only between dyads, and I analyze solely the effects of using a logit model when two states behave strategically. To do this, however, we need a method for incorporating the structure of strategic interdependence into statistical models of conflict.

ANALYZING STRATEGIC DISCRETE CHOICE MODELS OF INTERNATIONAL CONFLICT

Continuous-variable statistical methods for modeling strategic interaction have existed for some time and have been employed by international relations researchers, particularly in analyzing arms races and superpower reciprocity (see, e.g., Goldstein 1991; Goldstein and Freeman 1990, 1991; McGinnis and Williams 1989; Ward and Rajmaira 1992; Williams and McGinnis 1988). Yet, surprisingly little has been done in political science or economics to address the problem of strategic interdependence in discrete choice models. Two recent works are the only efforts in political science to address interdependence issues for this type of data. Beck, Katz, and Tucker (1998) take a binary time-series cross-sectional approach, using panel and time-series techniques to account for the interdependence between nations and over time. Their approach is primarily data driven, that is, the goal is to provide a method that is appropriate for binary data and that allows cross-sectional and time-series interdependence to be captured in the covariance matrix or lagged dependent variables. Smith (1997) takes a slightly more theory-driven approach, modeling conflict interdependence in militarized interstate disputes using a "strategically censored" bivariate-ordered probit model. Indeed, Smith's strategic data-censoring mechanism takes us closer to fully modeling the strategic interaction that a theory might imply. However, other aspects of the strategic interdependence are aggregated into the correlation parameter of the bivariate-normal distribution. Moreover, although Smith's statistical model is consistent with at least one game-theoretic model, that model would not generally be used as the basis for a theory of conflict.[6] Both Beck, Katz, and Tucker (1998) and Smith (1997) make significant advances in their own right, but neither derives the statistical model directly from a theoretical model. To differing degrees, both aggregate some aspect of the strategic interdependence into single parameters of statistical models—models designed more for a particular type of data than for a particular theory that produces a particular type of data. More satisfactory would be a method that incorporates the strategic theory directly into the statistical model without any aggregation of the theory's components.

At first glance, one might think that traditional game-theoretic solution concepts and refinements such as Nash and subgame perfection would be useful in statistically modeling international interaction, especially since they have been increasingly used to formalize the theories of that interaction. Certainly, a number of statistical tests of model fit can be conducted using such equilibrium concepts. These are quite limited, however, and do not allow for the broader range of hypothesis tests that are generally of more interest substantively (more on this later). Traditional equilibrium concepts prove problematic in statistical analysis primarily because of the *zero-likelihood problem*.[7] As the name implies, this problem arises during maximum likelihood estimation, a method for estimating the effect parameters commonly used in hypothesis tests. Recall that maximum likelihood estimation attempts to find the set of parameters, β, that give the maximum likelihood of having generated the observed outcomes, y. The likelihood function to be maximized is defined as

$$L(\beta|y) \propto f(y_1, y_2, \ldots, y_N|x, \beta) \tag{1}$$

$$= f(y_1|x_1, \beta)f(y_2|x_2, \beta) \ldots f(y_N|x_N, \beta), \tag{2}$$

where equation 2 follows from equation 1 due to the conditional independence assumption generally made.

As equation 2 displays, a requirement to conduct maximum likelihood estimation is a probability distribution $f(y_i|x, \beta)$ over the outcomes.[8] Yet, standard game-theoretic solution concepts and refinements provide little help in assigning a distribution over outcomes, since equilibria are assumed to be played with probability one, and nonequilibria are assumed to be played with probability zero.[9] And although an equilibrium in mixed strategies will result in the assignment of a probability less than one to the predicted outcomes, unless the equilibrium results in a mixture over *all* outcomes, there will still be some outcomes for which the model assigns zero probability. Since it is extremely unlikely for a model to predict perfectly every observed outcome, we can generally expect that for any value of the βs—even the maximum likelihood estimates—at least one observed outcome will have probability zero assigned to it by the model. Equation 2 shows, however, that if even a single observed outcome is not the outcome predicted by the model, then the likelihood equation will be zero. And because $L(\beta|y) = 0 \; \forall \beta$, maximum likelihood estimation will be impossible. Therefore, what is needed is some way to incorporate the structure of the strategic interaction from the game-theoretic model, but one that assigns nonzero probabilities (however small they may be) to every outcome.

In the econometrics literature, the general approach to analyzing strategic discrete choice has been to

[6] Personal communication with Alastair Smith.

[7] I have not found a documented source for this term, but I am told it is well known in the econometrics literature. (Personal communication with Mark Fey, who attributes it to Richard McKelvey, who attributes it to being well known.)

[8] Examples of $f(y_i|x, \beta)$ include the normal distribution for traditional linear regression; the Poisson, negative binomial, or generalized event count distributions for event count data; or the Bernoulli distribution for binary data.

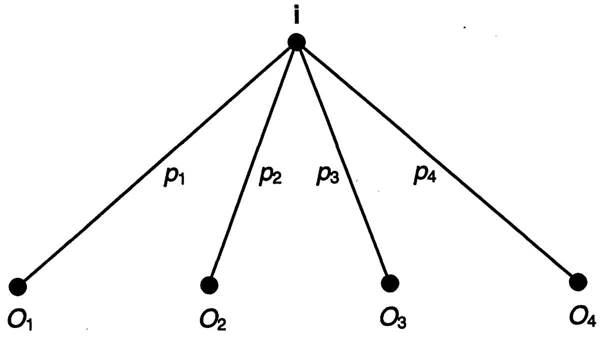
[9] In addition, except in evolutionary game theory, most say nothing about which of multiple equilibria are more likely to be played.

extend McFadden's (1974, 1976) random utility model to a strategic setting. Bjorn and Vuong (1984) and Bresnahan and Reiss (1991) use simultaneous equations models with dummy endogenous variables and assumptions of utility-maximizing behavior. McKelvey and Palfrey (1995, 1996, 1998) develop a statistical equilibrium concept for normal form and extensive form games. Because there has been a distinct move in the international conflict literature away from normal form games to extensive form games, and because the McKelvey and Palfrey agent quantal response equilibrium (QRE) is directly applicable to extensive form games, I employ it in deriving statistical strategic discrete choice models in the rest of this article. Although my argument to this point has levied no special requirements upon theories of international conflict other than an assumption of strategic interaction, hereafter I assume that the structure of the strategic interdependence embodied in a theory can be specified through an extensive form game.

At heart, the QRE is a game-theoretic solution concept, just like Nash or subgame perfection, but it and similar solution concepts are based on random utility assumptions. Depending on the implementation, these assumptions can take one of three forms: (1) part of each player's utility is unobserved, (2) players make errors in implementing their actions, or (3) players are boundedly rational (Chen, Friedman, and Thisse 1996; McKelvey and Palfrey 1995, 1996, 1998; Rosenthal 1989).[10] The basic idea is that players employ best responses to each other, conditioned on the knowledge that each will make mistakes according to some known (or assumed) distribution of errors. Specification of the distribution of those "errors" provides a statistical model of equilibrium, since it allows for nondegenerate (i.e., nonbinary) choice probabilities to be derived for the strategies players will choose. Hence, it allows for the derivation of nondegenerate probabilities for the outcomes of a game. Using a statistical equilibrium concept such as the QRE, one can derive the statistical version of a model of conflict in extensive form that directly incorporates the structure of the strategic interaction.

To understand how QRE choice probabilities are derived, it is useful to consider, first, how choice probabilities are derived in nonstrategic random utility models.[11] Figure 1 displays a discrete choice model in which individual i decides among four options (O_1, O_2, O_3, O_4). The individual's utility for option O_j is assumed to be $U^*(O_j) = U(O_j) + \varepsilon_j$, where $U^*(O_j)$ is the true utility for choice j, $U(O_j)$ is its "indirect" utility, and ε_j captures either the error in the decision making or unobserved variation in attributes or tastes. As a utility maximizer, the individual is assumed to choose the option O_j such that $U^*(O_j) > U^*(O_k)$ for all options $k \neq j$. However, because we do not observe

FIGURE 1. Discrete Choice Model with Exogenously Determined Choice Probabilities



$U^*(O_j)$ but only $U(O_j)$, we can only make probabilistic statements concerning which option the individual chooses. Multinomial logit choice probabilities, $p_j = e^{U(O_j)}/\sum_{k=1}^{4} e^{U(O_k)}$, result when the errors are assumed to be distributed type I extreme-value and multinomial probit probabilities result when the errors are assumed to be normally distributed. The widely used binomial versions of these are the special case when only two options exist. Since the utilities are assumed to be functions of exogenous variables, and since individual i's probabilities do not depend on anyone else's, the choice probabilities are exogenously determined.

The QRE is derived in a similar manner, except now we have an extensive form game with multiple decision makers, who each have their own random utilities over actions at information sets and are engaging in utility-maximizing behavior. In fact, one can roughly think of the QRE as applying the random utility model to each information set (i.e., decision node) of the extensive form game. At a given information set n, a person's utility for action a_{nj} is again assumed to be $U^*(a_{nj}) = U(a_{nj}) + \varepsilon_{nj}$, where $U^*(a_{nj})$ is the true utility for choice a_{nj}, $U(a_{nj})$ is its indirect utility, and ε_{nj} captures the error in the decision making. If a_{nj} leads to a terminal node (i.e., an outcome), then $U(a_{nj})$ is the indirect utility associated with that outcome. If a_{nj} leads to a nonterminal node, then $U(a_{nj})$ is the expected utility for taking that action. The player is again assumed to choose the option a_{nj} such that $U^*(a_{nj}) > U^*(a_{nk})$ for all options $k \neq j$ at information set n. When a distribution is specified for ε, choice probability equations can be derived for each action of each information set in the game tree.

When the errors are assumed distributed type I extreme-value, then the choice probability equations take the multinomial logit form

$$p_{nj} = \frac{e^{\lambda U_i(a_{nj})}}{\sum_{k \in A_n} e^{\lambda U_i(a_{nk})}}, \tag{3}$$

where individual i is associated with information set n, p_{nj} is the probability of action j at information set n, $U_i(a_{nj})$ is the indirect utility to i of action j, and A_n is the set of all actions at information set n. The parameter λ is inversely related to the variance of the errors.

[10] For a more detailed analysis of the statistical models that result from different theoretical sources of uncertainty in finite (or discrete) choice models, see Signorino 1998.

[11] See, for example, McFadden 1974, 1976; Maddala 1983, 59–64; and Pudney 1989, 111–9.

Therefore, smaller values of λ imply larger amounts of unobserved variation, agent error, or bounded rationality, while larger values of λ imply less error (Chen, Friedman, and Thisse 1996; McKelvey and Palfrey 1998).[12] From the bounded rationality perspective, $\lambda = 0$ corresponds to completely bounded rationality, and $\lambda = \infty$ corresponds to complete (i.e., Nash) rationality. For finite games of perfect information, $\lambda = \infty$ also corresponds to the subgame perfect equilibrium.

The solution to this system of choice probability equations, the logit quantal response equilibrium (LQRE) (McKelvey and Palfrey 1995, 1996, 1998), identifies the equilibrium probabilities for every action at every information set in the game. The equilibrium outcome probabilities are the product of the choice probabilities along the path from the first node to the respective outcome. These outcome probabilities can then be used in maximum likelihood estimation.

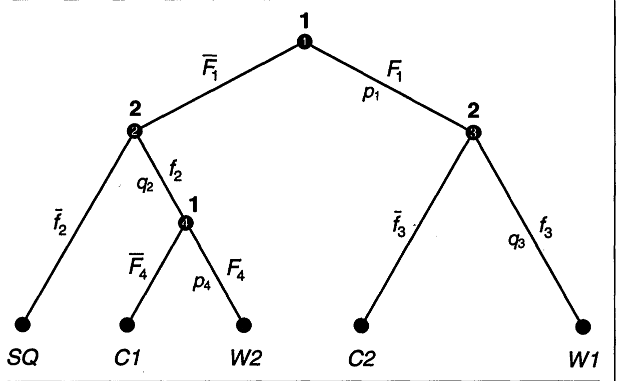
Before proceeding with a demonstration, two issues should be noted. First, the above is not just another formulation of nested logit. Since nested logit choice models involve only a single decision maker, they are nonstrategic, and their choice probabilities are exogenously determined. Although the LQRE has its roots in the same random utility framework as nested logit, the LQRE is a game-theoretic solution concept, and its choice probabilities are the equilibria of the decision makers' best responses to each other.

Second, care should be taken in transformation of the utilities. Although the LQRE is invariant to positive affine transformations of the true utilities, the $U^*(\cdot)$s, it is not invariant to scale transformations of the indirect utilities, the $U(\cdot)$s. Affine transformations of the $U^*(\cdot)$ also transform the variance of the error term, which leads to the same equilibrium choice probabilities as in the untransformed case. Transforming the indirect utilities leaves the variance untransformed, which results in different equilibrium choice probabilities. Fortunately, as I will show in the next section, this issue is generally not a problem during estimation.

A STATISTICAL MODEL OF CRISIS INTERACTION

To make this concrete and to show how one would analyze a game-theoretic model of international conflict, consider the crisis game in Figure 2, which I will also use in subsequent Monte Carlo analysis. This model has the same sequence of moves as the crisis subgame of Bueno de Mesquita and Lalman (1992, 34) and is similar in many respects to the crisis game of Kim and Bueno de Mesquita (1995, 5), the crisis bargaining game of Fearon (1994, 240), and the alliance model of Smith (1995, 407). The figure shows a potentially conflictual situation between two states, 1 and 2. State 1 must decide if it wants to fight (F_1) or not fight ($\bar{F}_1$). State 2 then decides if it wants to fight (f_2, f_3) or not fight ($\bar{f}_2$, $\bar{f}_3$). If state 1 decides not to

FIGURE 2. A Typical Bilateral Crisis Game



Note: States 1 and 2 alternate moves at decision nodes. Actions by state 1 are shown in uppercase, those by state 2 in lowercase. A bar over an action refers to the opposite of the action (e.g., not using force). The equilibrium choice probabilities used in the statistical model are denoted p_i for state 1 and q_j for state 2. Nonterminal nodes are numbered to simplify the expected utility notation and to index choice probabilities.

fight, but state 2 decides to fight, then state 1 must decide whether to respond by fighting (F_4) or not fighting ($\bar{F}_4$). If neither fights, then the outcome is the status quo (SQ). If both fight, then the outcome is war (W1 and W2). If one does not fight when the other does, then that side capitulates (C1 and C2). States 1 and 2 have utilities $U_1(\cdot)$ and $U_2(\cdot)$, respectively, for each of the outcomes. For later notational simplification, nonterminal nodes are also indexed. Strategies for state 1 are shown in uppercase, those for state 2 in lowercase.

The LQRE choice probabilities for this model are obtained by applying equation 3 at each of the information sets. Appendix A shows how to derive them from the first principles of the assumed random utility model. The resulting strategic choice probabilities are[13]

$$p_1 = \frac{e^{\lambda[q_3U_1(W1)+(1-q_3)U_1(C2)]}}{e^{\lambda[q_3U_1(W1)+(1-q_3)U_1(C2)]} + e^{\lambda\{q_2[p_4U_1(W2)+(1-p_4)U_1(C1)]+(1-q_2)U_1(SQ)\}}}; \tag{4}$$

$$q_2 = \frac{e^{\lambda[p_4U_2(W2)+(1-p_4)U_2(C1)]}}{e^{\lambda[p_4U_2(W2)+(1-p_4)U_2(C1)]} + e^{\lambda U_2(SQ)}}; \tag{5}$$

$$q_3 = \frac{e^{\lambda U_2(W1)}}{e^{\lambda U_2(W1)} + e^{\lambda U_2(C2)}}; \text{and} \tag{6}$$

$$p_4 = \frac{e^{\lambda U_1(W2)}}{e^{\lambda U_1(W2)} + e^{\lambda U_1(C1)}}. \tag{7}$$

The LQRE is the set of values for (p_1, q_2, q_3, p_4) that satisfy the above equations. Notice that, in contrast to traditional multinomial logit probabilities, here the strategic choice probabilities are often functions of each other—they are endogenous. Moreover, the probabilities have been derived directly from the strategic

[12] The type I extreme value distribution has the cumulative distribution $F(\varepsilon) = e^{-e^{-\lambda\varepsilon}}$, mean $E(\varepsilon) = \gamma/\lambda$, and variance $V(\varepsilon) = \pi^2/(6\lambda^2)$, where γ is Euler's constant (approximately 0.577).

[13] We need only specify one of the choice probabilities at each of the information sets, since each contains only two options.

theory. They reflect both the sequence of decisions each side must make as well as the incentives of each state for taking particular actions. For example, state 1's probability of initially fighting, p_1, depends on its expected utilities for reaching nodes 2 and 3. These expected utilities depend on the probability that state 2 will fight at those nodes, q_2 and q_3, respectively. Similarly, state 2's probability of fighting at node 2 depends on its utility for the status quo and on its expected utility for reaching node 4, which in turn depends on state 1's probability of fighting p_4 at node 4, which in turn depends on state 1's utilities for capitulating versus war.

The probability of each outcome in the crisis game is the product of the choice probabilities along the path from node 1 to that outcome:

$$p_{sq} = (1 - p_1)(1 - q_2); \quad (8)$$

$$p_{c1} = (1 - p_1)q_2(1 - p_4); \quad (9)$$

$$p_{w2} = (1 - p_1)q_2p_4; \quad (10)$$

$$p_{c2} = p_1(1 - q_3); \quad (11)$$

$$p_{w1} = p_1q_3. \quad (12)$$

As is evident, the outcome probabilities also reflect both the sequence of interaction and the endogeneity of decisions made by the two states.

To date, the LQRE has been used only to analyze experimental data for which the payoffs are fixed (Fey, McKelvey, and Palfrey 1996; McKelvey and Palfrey 1995, 1996, 1998). The goal in these studies has been to explain observed deviations from Nash and subgame perfect outcomes. In terms of estimation, the focus has been on the parameter λ, which is interpreted as a measure of the rationality of the subjects. In repeated experiments, λ is estimated for different periods of the entire trial and analyzed to see if learning takes place over the course of the experiments, that is, whether λ increases over time.

A natural extension of this to nonexperimental data is to specify the utilities in terms of explanatory variables and then estimate the effects of those variables as well as the fit of the model. To specify fully the crisis interaction model, we must specify the nations' utilities over the outcomes. In this hypothetical model, the utilities are based on each state's level of military capabilities, M_i; other assets, A_i (e.g., land, natural resources, or other nonmilitary assets that an adversary may desire); and a dummy indicator of whether both states are democracies, D_{ij}:

$$U_i(SQ) = D_{ij}; \quad (13)$$

$$U_i(Ci) = -A_i; \quad (14)$$

$$U_i(Cj) = A_j; \quad (15)$$

$$U_i(W) = P_iA_j + (1 - P_i)(-A_i - M_i); \quad (16)$$

$$P_i = \frac{M_i}{M_i + M_j}. \quad (17)$$

In words, the status quo yields no change in utility for a state unless both are democracies, in which case the utility of the status quo is raised. If i capitulates, then it loses its assets to j. The utility of war is the same regardless of who initiates the fighting. It is an expected utility based on the probability that i will win the war. If i wins, then it garners j's assets. If it is defeated, then it loses both its assets and its military capability.[14] The probability of winning a war (P_i) is defined as the ratio of i's military capabilities to the sum of i's and j's capabilities.

To estimate such a model, we assume that the utilities have been specified as above but with parameters (to be estimated) attached to each explanatory variable (e.g., $\beta_{m1}M_1$, $\beta_{m2}M_2$, $\beta_{a1}A_1$, $\beta_{a2}A_2$, β_dD_{12}). In its current form, the model is unidentified: Both λ and the β parameters cannot be estimated individually. To identify the effects parameters, λ must be constrained to some arbitrary value. Standard binomial logit models assume $\hat{\lambda} = 1$, which is what I use henceforth.

There are both advantages and limitations to the model being unidentified with respect to λ and β. Recall that the LQRE is not invariant to scale transformations of the indirect utilities. Having set $\hat{\lambda} = 1$, we might be concerned that a transformation of the indirect utilities would lead to different substantive conclusions from our estimation. Interestingly, it is precisely the unidentification of λ and β that relieves us of this problem when substantively interpreting our estimation results. Since the model is unidentified, we can only estimate the joint effect of λ and β. In the case of utilities that are linear in β, either we can reparameterize the statistical model in terms of $\beta^* = \lambda\beta$ and then estimate $\hat{\beta}^*$ or we can set $\hat{\lambda} = c$ for some arbitrary constant $c > 0$ and estimate $\hat{\beta}$. The two methods are equivalent. In the former, the joint effect of λβ is estimated directly. In the latter, the estimate $\hat{\beta}$ will compensate for having set $\hat{\lambda}$ to an arbitrary constant, and $\hat{\beta}$ will reflect the joint effect of λβ.[15]

There are two implications of this. First, although we set $\hat{\lambda} = 1$, from an estimation perspective we are not saying anything about the degree of bounded rationality of the decision makers. In other words, we are not assuming a level of rationality for them and then estimating $\hat{\beta}$. As we just noted, the estimated $\hat{\beta}$s reflect the joint effect of λ and β. Second, multiplying all utilities by a constant c (e.g., by rescaling the data) is equivalent to using the original utilities and multiplying $\hat{\lambda}$ by c. Hence, the estimated $\hat{\beta}$s in the scaled case will be $1/c$ times those in the unscaled case. Although the values of the parameter estimates may differ, the joint effect of $\hat{\lambda}\hat{\beta}$ will be the same in both cases, so the substantive interpretation of the effects of variables

[14] The assumption that i lose its military capabilities if defeated is not necessary to sustain the general results in the specification analysis section. If so desired, one could drop M_i in the $(-A_i - M_i)$ term and, instead, add an exogenous cost of war variable to the utility equation.

[15] Assuming the true parameters are λ and β and the utilities are linear in β, then setting $\hat{\lambda} = c$ will result in estimates $\hat{\beta} = (1/c)\lambda\beta$. Setting $\hat{\lambda} = 1$ therefore implies that $\hat{\beta} = \lambda\beta$.

and the probabilities of actions and outcomes will be identical. Therefore, while the equilibrium predictions of the LQRE may not be invariant to scale transformations of the indirect utilities, our substantive interpretations of estimation results will be.

The above does point out, however, the main limitation imposed by the unidentification of λ and β. Namely, having conducted our analysis, we can tell whether the predicted behavior is more or less rational, but we cannot tell whether that is due to an intrinsic characteristic of rationality in the decision makers (λ) or to the size of the stakes (as a result of the βs). For most questions of interest in international relations, this is not problematic, since in the LQRE, higher stakes produce behavior that is more rational (à la Nash). Moreover, all random utility models, including binomial logit and probit, face the same problem of identifying the effects and variance parameters, so this issue is not unique to using the LQRE solution concept in estimation.

Given a method for determining the probabilities over outcomes, the machinery for estimating the parameters is much the same as in standard random utility models. In fact, given the LQRE choice probabilities, we can analyze any of the actions or outcomes in the model, depending on the availability of data for dependent variables. The important point is that we incorporate the theorized strategic interaction into the statistical model.

For example, assuming data are available on all outcomes, we can let $y_{i,j}$ be a dummy that is one when $j \in \{sq, c1, w2, c2, w1\}$ is the outcome for observation i, and zero otherwise. The likelihood function to be maximized is

$$L = \prod_{i=1}^{n} p_{sq}^{y_{i,sq}} p_{c1}^{y_{i,c1}} p_{w2}^{y_{i,w2}} p_{c2}^{y_{i,c2}} p_{w1}^{y_{i,w1}}. \tag{18}$$

If, instead, data are only available on whether war occurred, then we let y_i be a dummy that is one if war occurred for observation i, and zero if it did not. Since the dependent variable represents wars started by either side (i.e., W1 or W2), the probability of war is $p_w = p_{w1} + p_{w2}$, and our likelihood function is

$$L = \prod_{i=1}^{n} p_w^{y_i}(1 - p_w)^{1-y_i}. \tag{19}$$

Finally, outcome data simply may not be available. Let us assume in this case that data are available only on nations' actions, such as whether state 1 initiated the use of force (F_1 in Figure 2).[16] We let y_i be one if state 1 initiated the use of force, and zero otherwise. The probability that nation 1 initiates force is p_1. Therefore, the likelihood function to be maximized in this case is

$$L = \prod_{i=1}^{n} p_1^{y_i}(1 - p_1)^{1-y_i}. \tag{20}$$

In each of these cases, the parameter estimates are then found by maximizing the likelihood with respect to those parameters, applying parameter constraints if necessary.[17] In practice, this can be done using any maximum likelihood estimation or constrained maximum likelihood estimation software that allows the user to define a procedure for the log-likelihood function. Such a procedure would generally take the data and the current iteration's parameter values, insert those into equations 4–17, and calculate the probability of each observation's actual action or outcome, depending on the type of dependent variable employed. The procedure would then either return a vector of the logarithm of those outcome probabilities or the sum of that vector, depending on the software used.

SPECIFICATION ANALYSIS: HOW WELL DOES BINOMIAL LOGIT MODEL STRATEGIC INTERACTION?

In view of the widespread use of logit to analyze models of strategic interaction, an important question concerns how well it actually accounts for strategic interaction.[18] In the current context, analyzing the misspecification of logit presents some difficulties. Analysis of omitted variable bias in OLS models is typically conducted by comparing estimators of the same parameter in two equations, one in which all relevant variables are present, and one in which a relevant variable is missing. Analysis of simultaneous equations bias employs the same system of equations and compares parameters estimated using OLS versus those using 2SLS. In such cases, specification analysis employs equations of the same functional form and compares estimators of the same parameter of interest under different conditions.

An appropriate analogy for our current situation would be if the "truth" were represented by a simultaneous equations model, but instead we regressed a different set of explanatory variables (or functions of the true exogenous variables) on one of the system's endogenous variables. In the case of comparing a logit versus strategic specification, the analysis of bias is made difficult for at least two reasons. First, in a logit specification, an analyst may be forced to use "aggregate" variables for concepts that could be directly modeled in the strategic specification, such as the relationship of a balance of power to the utility for war. Therefore, a strategic model and a typical binomial logit test of it may not even include the same set of regressors. Second, the logit and strategic models have different functional forms, which makes it difficult to

[16] I am indebted to Bruce Bueno de Mesquita for suggesting this approach. See Bueno de Mesquita, Morrow, and Zorick (1997) for a method of performing logit on actions at a decision node, using the subgame perfect equilibria of the subgames following those actions.

[17] For example, in the crisis model we would need to constrain $\beta_{m1}\beta_{m2} \geq 0$. Otherwise, we might obtain negative values for probabilities.

[18] To be precise, by "traditional logit" I mean binomial logit with linear latent utilities. I thank an anonymous reviewer for identifying this.

compare coefficients even when they are associated with the same regressor. For example, if we include a variable "democracy" in both the logit and strategic regressions, then we cannot compare their coefficients directly, since each coefficient must be interpreted in the context of its model's functional form. We *can* relate the coefficients to a common quantity of interest, such as the probability of war occurring. When we move from comparing estimates to comparing estimators, however, the math involved becomes much more complex than in standard analyses of bias.

Fundamentally, what we really want to know is whether binomial logit will give us a correct picture of the world we are analyzing—a world that we assume involves strategic decision making. To examine this, we can conduct a simple Monte Carlo analysis. We can generate the data for our Monte Carlo world using the statistical version of the two-nation crisis game as the "truth." Then, typical logit analyses can be run using that data. Finally, knowing what generated the data, we can assess the substantive interpretation of the logit parameter estimates as well as logit's ability to model the true probability of war. It should be noted that, as specified, the crisis game's utilities are not complex. In fact, arguably it is *too simple* a model of international conflict. Therefore, if logit can adequately model strategic interaction, this should not be a hard test.

To generate Monte Carlo data for the crisis game, the military capabilities (M_i) and assets (A_i) of each nation are assigned uniformly distributed random draws between 0 and 100. For the joint democracy dummy variable, D_{ij}, random draws are made with a 0.2 probability of both being democracies. The parameters of each variable are assumed to be $\beta_{m1} = \beta_{m2} = \beta_{a1} = \beta_{a2} = 1$ and $\beta_d = 20$. Since the democracy variable is a dummy, this puts it in the same range as the others and provides a substantial (but not overwhelming) incentive for two democracies to seek the status quo. The outcomes are determined using the underlying random utility model of the LQRE, that is, at each information set, the nation maximizes the $U^*(a_{nj}) = U(a_{nj}) + \varepsilon_{nj}$ over its possible actions, A_n, at that information set, where $U(a_{nj})$ is determined by the utilities of the outcomes or the expected utilities based on them, and ε_{nj} is drawn from a type I extreme value distribution with $\lambda = 1$. For the data to be used in the strategic random utility model regression, the dependent variable is set to an index (1–5) that corresponds to one of the outcomes in Figure 2. For the logit regression data, the dependent variable is set to one if the outcome is either W1 or W2, zero otherwise. For each observation, we then have M_1, M_2, A_1, A_2, and D_{12} and the observed outcome. A total of $N = 1{,}000$ observations were generated.

With the Monte Carlo data in hand, one strategic and three logit regressions were run. Using the strategic choice probabilities derived in the last section, a strategic regression was run to determine whether it could recover the true parameters. As Table 1 shows, the "strategic" estimates are very close to the true parameter values. While this should not surprise us, since the data were generated using random utility behavior applied to the bilateral crisis game, the results provide some confidence that the estimation procedure is working correctly and that the parameters are identified.

TABLE 1. Results of Strategic versus Typical Logit Models

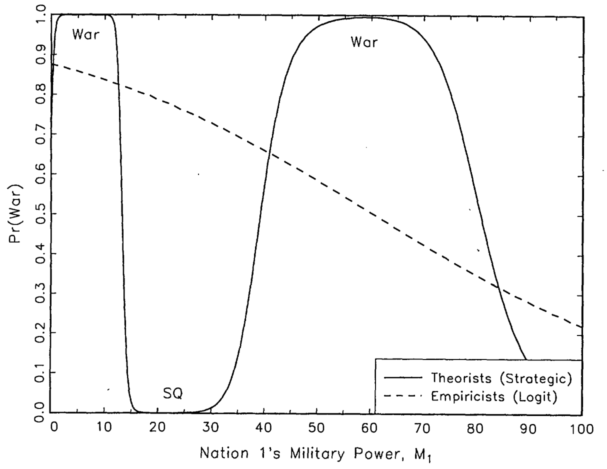
	Strategic	Naive	Balance of Power	Joint Utility
Constant		.76	−1.45	−.69
		(.29)	(.13)	(.10)
Military_1	.98	−.03		
	(.06)	(.003)		
Military_2	.99	−.03		
	(.06)	(.003)		
Assets_1	.98	.02		
	(.06)	(.003)		
Assets_2	.99	.02		
	(.06)	(.003)		
MilCon			3.20	
			(.27)	
$U_1(W)U_2(W)$				−.22
				(.015)
Democracy_{12}	19.26	−1.15	−1.08	−1.91
	(1.21)	(.21)	(.19)	(.30)
ln L	−140.6	−505.1	−580.8	−351.6

Note: The N = 1,000 observations were generated using the simple crisis game. The Strategic model uses the LQRE probabilities. The Naive, Balance-of-Power, and Joint Utility models are standard logit models. Standard errors are in parentheses.

The first binomial logit model analyzed is representative of a naive analyst who, having obtained a new data set with variables for war, capabilities, assets, and democracy, decides to conduct a simple logit regression of M_1, M_2, A_1, A_2, and D_{12} on war "to see what the data reveal." Much to the analyst's delight, the results of this regression, denoted as Naive in Table 1, are all highly statistically significant. Substantively, the estimates imply that an increase in military capabilities decreases the probability of war, an increase in assets increases the probability of war, and two democracies are unlikely to go to war. No doubt the researcher would note the counterintuitive finding that another nation's increase in military power actually decreases one's own probability of going to war with that nation. Since it is unrealistic to suggest that nations divest themselves of their assets, the analyst's policy prescription would be that nations should increase their military capabilities as much as possible: More military power leads to less war for everyone. One could probably find justification for this in one of the various theories in international relations. The question is whether the substantive inferences are correct, given the true model that generated the outcomes.

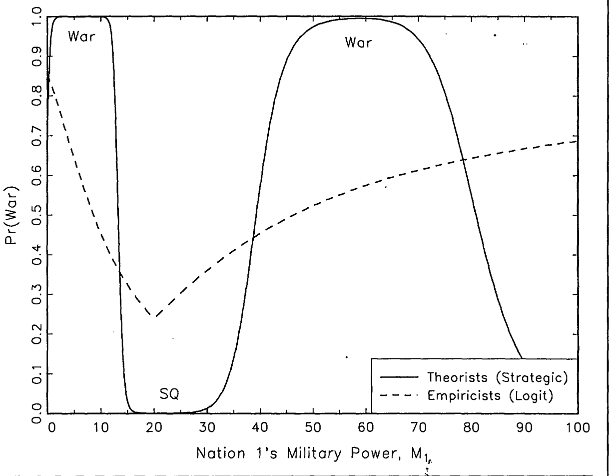
The graph in Figure 3 displays the true (i.e., strategic) and logit probabilities of war, setting $M_2 = 20$, $A_1 = 40$, $A_2 = 40$, and $D_{12} = 0$ and letting M_1 vary over its observable range. As the logit estimates indicated, the logit probability of war decreases as state 1's military power increases. Yet, Figure 3 also suggests that the naive analyst's policy prescription is a recipe for disaster, since the logit curve clearly does not reflect

FIGURE 3. Strategic versus Logit Probabilities of War: Naive Model



Note: The solid line shows the "true" strategic probability of war based on the simple crisis game, which was used to generate the data. The dashed line shows what the Naive logit model would predict in this case.

FIGURE 4. Strategic versus Logit Probabilities of War: Balance of Power Model



Note: The solid line shows the "true" strategic probability of war based on the simple crisis game. The dashed line shows what the Balance-of-Power logit model would predict in this case.

the true probability of war. It is true that at certain levels of military power, nation 1 will decrease the probability of war by increasing its military capabilities, but only up to a point. If, for example, nation 1 is at $M_1 = 25$ (with a high probability of the status quo), then increasing its military capabilities will lead to an equilibrium with a high probability of war. Note, too, that fitted values and calculated effects of M_1 on the probability of war will at times be grossly wrong—fitted values sometimes by more than 0.7 (on a scale of 0 to 1). To be fair, the naive logit model does accurately capture the effect of democracy on war.

Some may argue that I have stacked the deck with the preceding example; researchers generally would not regress the individual variables on war but instead would attempt to identify observable implications of a theory and then use indices or (perhaps nonlinear) functions of the above variables to get at the strategic interaction. For example, suppose a more sophisticated analyst were interested in testing hypotheses concerning the effects of balance of power and power preponderance on war. A common way of doing this has been to use a measure of the concentration of military capabilities among states. Using the same data as in the previous analysis, this analyst constructs a variable, *MilCon*, for the concentration of the two states' capabilities and regresses it and the democracy variable on war.

The logit estimates for the Balance of Power model were shown in Table 1, and the probability of war is plotted in Figure 4. The estimates, which are again highly significant, indicate that an increase in capability concentration leads to an increased probability of war. In fact, the figure does show that a decrease in capability concentration (e.g., when $M_1 = M_2 = 20$) leads to a decrease in war. It is not always true, however, that an increase in capability concentration heightens the probability of war. At $M_1 = 60$, raising capability concentration by increasing M_1 decreases the probability of war. As in the previous case, the logit curve for the probability of war is often quite divergent from the true probability of war. Nevertheless, the effect of democracy is correct.

Finally, consider one last fairly sophisticated analysis of the data. Suppose a third analyst hypothesizes that the more two nations jointly value war, the more likely they are to go to war. This has quite an intuitive appeal—in fact, Bueno de Mesquita and Lalman (1988) conduct an analysis exactly along these lines.[19] Let us assume this third analyst uses the same data as before and constructs a joint utility of war variable, $U_1(W)U_2(W)$, that employs the utility for war equation (equation 16) used in generating the data.

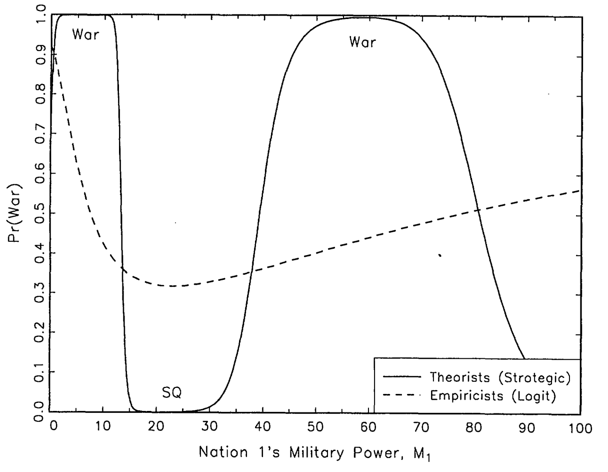
As seen in Table 1, the results for the Joint Utility regression are, yet again, highly significant. Moreover, as in the first analysis, they are also quite counterintuitive. The logit estimates imply that the more two states jointly value war, the less likely war is. Yet, as Figure 5 indicates, as in the two previous cases, the logit curve does not accurately reflect the true probability of war. As before, the effect of democracy is correct.

The question posed at the beginning of this section was: How well does traditional logit model strategic interaction? At least for the simple crisis interaction model here, the answer appears to be: Not very well at all.[20] Perhaps more troubling are the highly significant results in each case, which would be interpreted by the

[19] Bueno de Mesquita and Lalman (1988) actually regress their operationalized "P(War)" on war, but P(War) is defined in terms of the product of the two sides' expected utilities for challenging each other.

[20] The one exception is the effect of democracy, which has been correct in each case. My conjecture is that, since democracy appears linearly and in only a single outcome utility (i.e., SQ), it has a monotonic effect on the probability of war, which the logit analyses accurately capture.

FIGURE 5. Strategic versus Logit Probabilities of War: Joint Utility Model



Note: The solid line shows the "true" strategic probability of war based on the simple crisis game. The dashed line shows what the Joint Utility logit model would predict in this case.

typical researcher as supporting one model or another. Hence, out of a single data set, support could be "found" for a number of different theories of international conflict—all of which are wrong.

REANALYZING THE INTERNATIONAL INTERACTION GAME

One of the best-known models in the international conflict literature is Bueno de Mesquita and Lalman's (1992) international interaction game. Pedagogically, it is an excellent candidate for the application of a strategic statistical model. It is a rare example of a well-specified game-theoretic model of international conflict that has been empirically tested, with considerable attention paid to the operationalization of the expected utilities involved.[21] Moreover, the authors have made their data available for use by other scholars.

The international interaction game is also an excellent candidate because, although Bueno de Mesquita and Lalman find considerable empirical support for it, the structure of the strategic interaction is never directly or completely incorporated into the statistical models used to test it. Moreover, two limitations to their tests raise questions concerning the empirical support. Since we now have the means to do so, we can join the theory and estimation and provide a better test of this important model.

This section proceeds as follows. First, the international interaction game is presented in a way that is true to the original specification of Bueno de Mesquita and Lalman but that allows us to develop LQRE choice probabilities and to test nested models of the game. Second, I will identify and attempt to correct (or at least mitigate) two limitations of the original tests in Bueno de Mesquita and Lalman, and I will then reanalyze the data using subgame perfection, as they did. Finally, because there are limitations to the types of statistical analysis that can be done using subgame perfection, I will analyze the international interaction game using a statistical strategic model based on LQRE choice probabilities.

The International Interaction Game

The model that I will use here is, for purposes I will detail later, a generalization of the game specified in Bueno de Mesquita and Lalman. Figure 6 displays the perfect information version of the international interaction game. Here, two states choose between making demands of each other and using force if demands are made. If neither makes a demand on the other, then the game results in the status quo (SQ). If nation 1 makes a demand and nation 2 concedes to it, then the result is acquiescence by 2 (Acq_2). Similarly, if nation 2 makes a demand and nation 1 concedes, then the result is acquiescence by 1 (Acq_1).

Nodes 5 and 6 of the game identify "crisis subgames" (Bueno de Mesquita and Lalman 1992, 30–4). Here, demands have been levied by both players, and each must decide whether to use force in attempting to obtain its demands. Negotiation (Nego) results when both decide not to use force. Capitulation by 2 (Cap_2) results when nation 1 uses force and 2 backs down. Similarly, capitulation by 1 (Cap_1) results when nation 2 uses force and 1 backs down. Finally, when both states use force, the result is war; it is differentiated by whether 1 or 2 initiates the use of force (War_1 and War_2, respectively).

Each state's utilities over the outcomes are defined similarly to Bueno de Mesquita and Lalman:

$$U_i(SQ) = \beta_{i1}U_i(SQ'); \tag{21}$$

$$U_i(Acq_j) = \beta_{i2}U_i(\Delta_i); \tag{22}$$

$$U_i(Acq_i) = \beta_{i3}U_i(\Delta_j); \tag{23}$$

$$U_i(Nego) = \beta_{i4}P_iU_i(\Delta_i) + \beta_{i5}(1 - P_i)U_i(\Delta_j); \tag{24}$$

$$U_i(Cap_j) = \beta_{i2}U_i(\Delta_i) + \beta_{i6}\phi_iP_i; \tag{25}$$

$$U_i(Cap_i) = \beta_{i3}U_i(\Delta_j) + \gamma_i(1 - P_i); \tag{26}$$

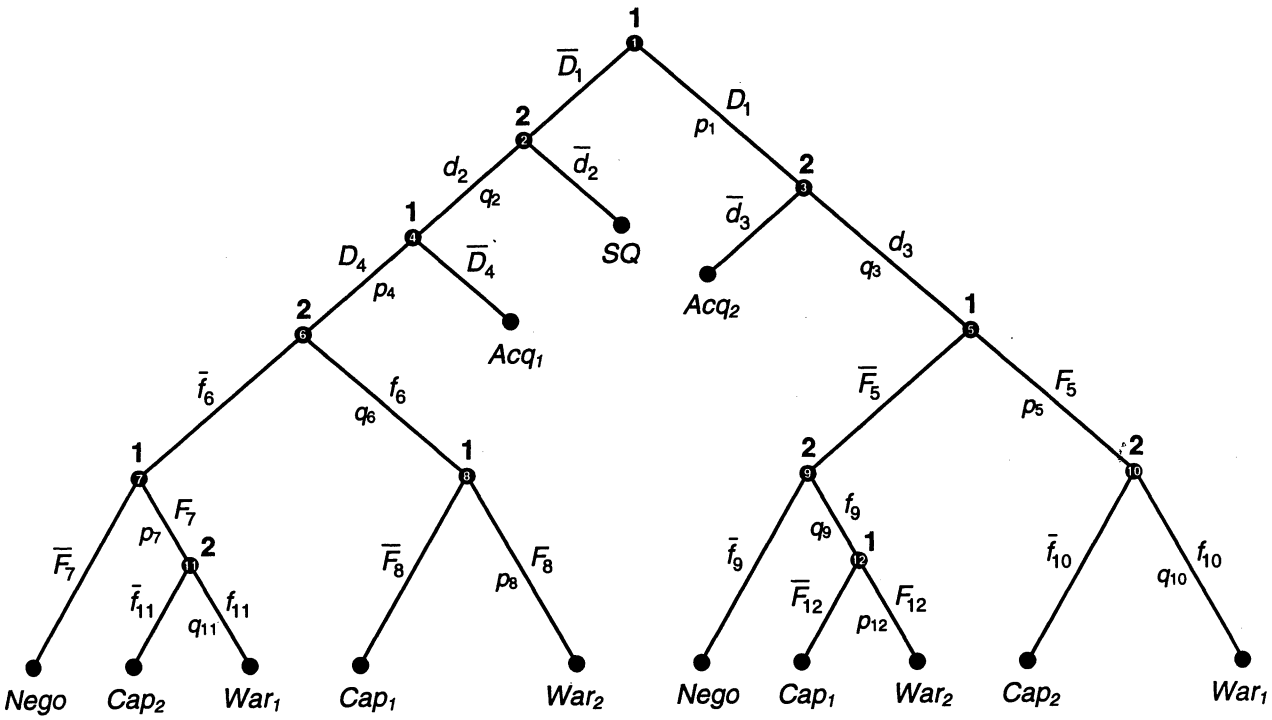
$$U_i(War_i) = \beta_{i4}P_iU_i(\Delta_i) + \beta_{i5}(1 - P_i)U_i(\Delta_j) + \beta_{i6}\phi_iP_i + \alpha_i(1 - P_i); \tag{27}$$

$$U_i(War_j) = \beta_{i4}P_iU_i(\Delta_i) + \beta_{i5}(1 - P_i)U_i(\Delta_j) + \beta_{i6}\phi_iP_i + \tau_i(1 - P_i). \tag{28}$$

$U_i(SQ')$ is i's utility for the status quo, $U_i(\Delta_i)$ is i's utility for its own demand, $U_i(\Delta_j)$ is i's utility for state j's demand, P_i is i's subjective probability that it can win a war against j, ϕ_i is the domestic political cost of using force, and γ_i is the cost of giving in after being

[21] I am not aware of another case in which researchers both formally specify their model and then test it by operationalizing the utilities involved. Because of the difficulty in obtaining data representative of utilities, tests of formal models of conflict (see, e.g., Fearon 1994; Kim and Morrow 1992) have generally been conducted more indirectly, through tests of the observable implications.

FIGURE 6. The International Interaction Game of Bueno de Mesquita and Lalman (1992) and Corresponding Choice Probabilities Used in the Strategic Statistical Model



Note: States 1 and 2 alternate moves at decision nodes. Actions of state 1 are shown in uppercase, those of state 2 in lowercase. A bar over an action refers to the opposite of the action (e.g., not making a demand). The equilibrium choice probabilities used in the statistical model are denoted by p_i for state 1 and q_j for state 2. Nonterminal nodes are numbered to simplify the expected utility notation and to index choice probabilities.

attacked. The utilities $U_i(War_i)$ and $U_i(War_j)$ are differentiated by the cost terms α_i and τ_i, which are the costs of attacking and being a target, respectively. Note that for the international interaction game, there are two sets of these utilities: one for nation 1 and one for nation 2. Finally, the generalization to which I referred above is the addition of the β_{ij} coefficients to each term. These allow us to specify "nested" versions of the utilities later. The utilities in Bueno de Mesquita and Lalman are the special case in which $\beta_{i1} = \beta_{i2} = \beta_{i3} = \beta_{i4} = \beta_{i5} = 1$, and $\beta_{i6} = -1$. Based on these utilities, Bueno de Mesquita and Lalman specify in detail the conditions leading to a subgame perfect equilibrium (SPE) of the international interaction game.

Comparing Observed Outcomes to Subgame Perfect Predictions

To test the international interaction game, Bueno de Mesquita and Lalman use data on 707 European dyads from 1815 to 1970. The utilities for demands, the utilities for the status quo, and the subjective probabilities of winning a war are all operationalized based on functions of states' alliance commitments and their national material capabilities. The outcomes are operationalized using the militarized interstate disputes data set. For details concerning these operationalizations, see Bueno de Mesquita and Lalman (1992, Appendix 1).[22]

Ideally, for any given European dyad, one should be able to take the two states' operationalized utilities over the outcomes and determine the SPE of the game. Although that would not be equivalent to developing a strategic statistical model, one then could at least compare the expected outcomes to the empirical outcomes for all the cases in the data set. To their credit, Bueno de Mesquita and Lalman (1992, 72–92) attempt to do just that. As alluded to earlier, however, there are two limitations to their tests. The first concerns how they determine expected outcomes in the face of unavailable data. The second concerns how aggregating outcomes can yield misleading results. I address these in turn and then reanalyze the international interaction game on its own terms, that is, using subgame perfection. Not only does this provide a better test of the international interaction game in the framework that Bueno de Mesquita and Lalman originally conceived it, but it also allows us to compare the results using the SPE solution concept to those using the LQRE solution concept.

[22] I do not address issues concerning operationalizations here, since my purpose is to reexamine how well their theoretical model fares using their data when the structure of the strategic interaction is incorporated into the statistical model.

Limitation 1: Unavailable Cost Data and Nonassessable Equilibrium Conditions

In addition to the utilities of states' demands and of the status quo, utility equations 21–28 contain a number of cost terms: domestic costs of using force (ϕ_i), the costs of giving in after being attacked (γ_i), the costs of attacking another state in war (α_i), and the costs of being attacked (i.e., the target) in a war (τ_i). Although Bueno de Mesquita and Lalman operationalize the domestic costs of using force, they do not have data on the other costs. Because of that, Bueno de Mesquita and Lalman (1992, 76 fn 10; 298) acknowledge that they cannot empirically assess certain conditions necessary to determine the game's SPE, that is, the model's expected outcome. Some conditions involving the above cost terms, such as $U_i(Acq_i) > U_i(Cap_i)$, can still be evaluated because, by definition, one has to be less than or equal to the other. Others do not have such a priori restrictions on the ordering.[23] To forge ahead with their empirical tests, Bueno de Mesquita and Lalman (1992, 76 fn. 10) are forced to ignore nonassessable conditions in the calculation of the SPE expected outcomes. Having said that, their tables of results appear to test whether the SPE predictions explain the observed outcomes for War_1, Acq_2, SQ, and Nego individually (Bueno de Mesquita and Lalman 1992, 77–89). Moreover, it appears that considerable support is found for the international interaction game.

But the question must be raised: What is the effect of ignoring the equilibrium conditions that cannot be evaluated due to the lack of cost data? What is really being tested? Take the case of testing whether the model's prediction of War_1 is related to whether war is empirically observed (Bueno de Mesquita and Lalman 1992, 77, Table 3.3). The necessary and sufficient conditions for War_1 to be a SPE are that $U_1(Cap_1) > U_1(War_2)$, $U_1(War_1) > U_1(Acq_1)$, $U_2(Cap_1) > U_2(Nego)$, and $U_2(War_1) > U_2(Acq_2)$ (Bueno de Mesquita and Lalman 1992, 72). Yet, because the cost data are unavailable, only one of these conditions can be evaluated directly: $U_2(Cap_1) > U_2(Nego)$.[24] The result is that Bueno de Mesquita and Lalman are forced to drop equilibrium conditions in determining the expected outcomes, which they label "War_A" (War_1 in Figure 6). What they do not make clear is that dropping individual equilibrium conditions creates a reduced set of conditions that are consistent with the SPE conditions for multiple outcomes.

For the sake of argument, let us assume that all the above conditions could be evaluated except for $U_1(Cap_1) > U_1(War_2)$. In this case, the reduced set of conditions from dropping $U_1(Cap_1) > U_1(War_2)$ is consistent not only with War_1 as the SPE but also with SQ or Nego as the SPE. Hence, we should expect the dummy predictor variable of War_1 (used in the empirical analysis) to contain "1"s not only when War_1 is the SPE but also for a subset of the cases in which SQ and Nego are the SPE. As the number of nonassessable conditions increases, so does the inclusion of other outcomes in the predictor variable. It is not clear, then, what is really being tested by labeling this War_1 and regressing it against the observed outcomes for war. This same argument applies to the other tests conducted in Bueno de Mesquita and Lalman (1992, 77–89).

Limitation 2: Tests Based on Aggregated Outcomes

The second limitation of these tests concerns how the expected outcomes are compared to the observed outcomes. Recall that there are seven observed outcomes: SQ, Acq_1, Acq_2, Nego, Cap_1, Cap_2, and War.[25] Assuming that the first limitation identified above were not an issue, a correct comparison of the actual versus SPE expected outcomes would be to construct a 7×7 cross-tabulation. Entries in the table along the main diagonal would represent observed outcomes that were correctly predicted. Entries in the off-diagonal cells would represent observed outcomes that were not correctly predicted. As long as no marginal counts (or frequencies) equal zero, a χ^2 statistic can be calculated to determine whether the expected and observed outcomes are independent of each other. Since this is not always an accurate measure of how well the model fits the data, one might also calculate an indicator of model fit, such as the percentage of the outcomes correctly predicted.

Such a table is not presented by Bueno de Mesquita and Lalman. Instead, the tests are generally conducted for each individual outcome by aggregating the seven outcomes into two: whether or not a particular outcome occurred. The problem with aggregating a 7×7 table into a 2×2 table is that the latter can give the misleading impression that the model has done a good job of explaining the variation in the outcomes when in fact it has not. For example, suppose we aggregate the 7×7 table into a 2×2 table with rows representing whether the SQ occurred and the columns representing whether the model predicted the SQ. I will use the notation "(actual, predicted)" to refer to the cells of this table. χ^2 and goodness of fit statistics will generally be favorable when most of the elements fall along the main diagonal, which is comprised of the (not SQ, not SQ) and the (SQ, SQ) cells. While the latter cell represents a real success for the model in terms of predicting an outcome, the former is more complicated—and therein lies the problem.

In the 2×2 table, the (not SQ, not SQ) cell can hide a large amount of error. It includes not only the six other "correctly predicted" cells along the main diagonal of the 7×7 table but also the 30 other "mispredicted" cells in which SQ did not occur and was not

[23] For the a priori ordering restrictions, see Bueno de Mesquita and Lalman 1992, 47, Table 2.3.

[24] In some cases, some (but not all) of the above conditions can be evaluated indirectly, as implications of the remaining assessable conditions.

[25] There are eight outcomes in the model, but War_1 and War_2 are not differentiated empirically, so their model predictions are combined when comparing against observed outcomes.

TABLE 2. Results of the International Interaction Game, Assuming Subgame Perfection and Unknown Cost Parameters

	Sets of Possible Predicted Outcomes						
Actual Outcome	∅[a]	SQ	Nego	Acq_2 Nego	Acq_1 Acq_2 Nego	Acq_1 Acq_2 Nego War_1	SQ Acq_1 Acq_2 War_1
SQ	21	1	3	57	3	115	0, 38
Acq_1	0	0	0	1	0	0, 6	0, 1
Acq_2	4	1	0	0, 21	0	0, 62	0, 13
Nego	2	0	3	0, 9	0	0, 11	10
Cap_1	0	0	1	5	0	18	2
Cap_2	6	0	1	17	0	74	12
War	20	1	5	36	0	0, 111	0, 16
				1% ≤ CP ≤ 41%			

Note: Cells with two entries represent cases in which (1) the actual outcome is included in the set of possible predicted outcomes, but (2) because of the missing cost data, we cannot state with certainty whether the actual outcome is predicted by the international interaction game. In these cases, the first and second entries are the minimum and maximum numbers, respectively, of correct predictions possible. Based on this, the percentage of cases correctly predicted (CP) is between 1% and 41%.
[a]∅ predictions are those cases in which the empirically calculated utilities violated the a priori ordering restrictions, yielding "no prediction" from the model.

predicted: (Nego, Cap_1), (Acq_2, War), and so on. By examining the 7 × 7 contingency table, one can tell immediately how well the model fares. In a 2 × 2 aggregation of it, we cannot tell whether the (not SQ, not SQ) elements represent successful predictions of other outcomes or mispredictions of non-SQ outcomes. The fact that Bueno de Mesquita and Lalman's 2 × 2 tables for Acq_2, SQ, and Nego have 62%, 63%, and 84% of their elements, respectively, in the (not, not) cell, while only successfully predicting 91 cases (13%), raises the question of how many of the (not, not) cases are also successful predictions.

Bounds on the International Interaction Game's Explanatory Power

I now reanalyze the international interaction game, controlling for the above limitations. In the previous section, it was noted that the model's explanatory power can be determined by first constructing a 7 × 7 cross-tabulation of the actual versus predicted outcomes and then calculating the percentage of outcomes correctly predicted. Traditional statistical models produce single expected outcomes. The international interaction game also yields a single SPE outcome for any ordering of the outcomes.[26] Dropping equilibrium conditions, however, means that we cannot differentiate between multiple expected outcomes, all of which are consistent with the available information on outcome orderings. For example, dropping $U_1(Cap_1) > U_1(War_2)$ from the War_1 SPE conditions results in a subset of assessable conditions for which SQ, Nego, and War_1 are all consistent as SPE. A consequence is that we cannot determine whether certain observed outcomes were "correctly predicted." For the given example, if the actual outcome is Acq_1, we can definitely say that the model mispredicted the outcome, since the actual outcome is not in the set of possible predicted outcomes: {SQ, Nego, War_1}. If the actual outcome is War_1, then we have no way of knowing whether the model correctly predicted the outcome. On the one hand, War_1 could have been predicted by the model. On the other hand, the nonassessable conditions may have been such that SQ or Nego is the true predicted outcome. In this case, if War_1 is the actual outcome, all we can say is that the set of possible predicted outcomes contains the actual outcome.

While this introduces uncertainty concerning the fit of the model, we can at least put bounds on that fit. Depending on the actual outcome and the set of possible predicted outcomes, three situations may arise. First, regardless of the number of possible predicted outcomes, if the observed outcome is not contained in the set of possible predicted outcomes, then we can definitely count that as a model misprediction. Second, there is also no uncertainty if the model predicts only a single outcome and it matches the observed outcome—we can count that as a correct prediction. Uncertainty arises in the third situation, when the actual outcome is contained in the set of possible predicted outcomes and that set contains multiple outcomes. In this case, since the actual outcome may or may not have been predicted, we count it as a misprediction for the lower bound and as a correct prediction for the upper bound.

Table 2 displays the cross-tabulation of the actual outcomes versus the sets of possible SPE predicted outcomes given the available data.[27] Cells with two entries represent cases in which the actual outcome is included in a set of multiple possible predicted outcomes. In these cases, the first and second entries are the minimum and maximum numbers, respectively, of

[26] I assume that ties do not exist in the ordering of outcomes.

[27] In some cases, the empirically calculated utilities violated the a priori ordering restrictions, yielding "no prediction" from the model. In Table 2 these are denoted as ∅ predictions.

correct predictions possible. As Table 2 shows, without the cost data, all we can say about the model is that the percentage of cases correctly predicted (CP) is between 1% and 41%. As a reference, a null model that always predicted SQ would correctly predict 34% of the observations. So, if we were to give the international interaction game the greatest benefit of the doubt possible, it would predict 21% better than the modal category.

A STATISTICAL MODEL OF THE INTERNATIONAL INTERACTION GAME

The preceding analyses allowed us to answer only one particular question: How well does the model fit the data? In conducting empirical tests, we are often also interested in testing hypotheses concerning particular variables, groups of variables, and alternative models. The preceding method, while useful for comparing actual versus predicted outcomes, does not allow us to address any of these issues. Using a statistical equilibrium solution concept does.

Using the extensive form game in Figure 6, the LQRE outcome probabilities of the international interaction game can be derived in the same manner as for the bilateral crisis game; they are listed in Appendix B. The utilities are again given by equations 21–28. As I mentioned previously, the β_{ij} parameters in the utility equations allow us to compare nested models. I analyze three models based on different parameterizations of β_{ij}, γ_i, α_i, and τ_i.

1. Null Model: All parameters (β_{ij}, γ_i, α_i, τ_i) of the utility terms are restricted to zero. This model corresponds to the decision maker at each node choosing an option with probability 0.5, for example, by flipping a coin. The log-likelihood value of this model is then compared with the log-likelihoods of the other models.
2. Bueno de Mesquita and Lalman Bounded Rationality (BdM&L-BR) Model: The β_{ij} parameters of the utility terms are restricted to the values in the Bueno de Mesquita and Lalman model, as previously identified: $\beta_{i1} = \beta_{i2} = \beta_{i3} = \beta_{i4} = \beta_{i5} = 1$, and $\beta_{i6} = -1$. The cost parameters γ_i, α_i, and τ_i are estimated subject to the Bueno de Mesquita and Lalman constraints: $\gamma_i, \alpha_i, \tau_i \leq 0$, and $\tau_i \leq \alpha_i$.
3. Unrestricted Model: All parameters are now estimated, allowing us to assess the different effects of the utility terms in each nation's utility functions. The β_{ij} parameters are estimated subject to the constraints that $\beta_{i1}, \beta_{i2}, \beta_{i3}, \beta_{i4}, \beta_{i5} \geq 0$ and $\beta_{i6} \leq 0$. The cost parameters, γ_i, α_i, and τ_i, are estimated subject to the constraints $\gamma_i, \alpha_i, \tau_i \leq 0$ and $\tau_i \leq \alpha_i$.[28]

The exact same data are used for this analysis as for the analysis under subgame perfection. For the null model, no real estimation is needed. We simply plug in the above parameter values and data, calculate the probabilities of the observed outcomes, and calculate the log-likelihood and percentage of outcomes correctly predicted. For the BdM&L-BR and unrestricted models, constrained maximum likelihood estimation was conducted similarly to the estimation of the bilateral crisis game. Table 3 displays these results.

The null model represents one null hypothesis: All coefficients are zero. Again, one interpretation of this is that decision makers randomly choose their options at each node with probability 0.5. In this case, the terminal probabilities are: Pr(SQ) = 0.25, Pr(Acq_1) = 0.13, Pr(Acq_2) = 0.25, Pr(Nego) = 0.09, Pr(Cap_1) = 0.06, Pr(Cap_2) = 0.08, and Pr(War) = 0.14. The log-likelihood value for choosing actions this way is $\ln L^0 = -1292.7$. This value is of no real interest in itself, but it will be useful in comparisons to the log-likelihoods of the other two models.

Since we no longer have a probability zero or one prediction for each outcome, determining the percentage correctly predicted is done slightly differently here than in the analysis under subgame perfection. The method most often used in this type of situation is to identify the outcome with the highest probability as the "predicted outcome" and then compare those predicted outcomes to the actual outcomes.[29] In the case of the null model, this is further complicated by the fact that both the SQ and Acq_2 are predicted as having the highest probability of occurring: Pr(SQ) = Pr(Acq_2) = .25. Therefore, two values are reported in Table 3: $CP = 14\%$ if we take Acq_2 as the predictor, or $CP = 34\%$ if we take SQ as the predictor.

In the BdM&L-BR model, the values of the β_{ij} parameters are restricted to those in Bueno de Mesquita and Lalman (1992), but the parameters γ_i, α_i, and τ_i are estimated, since the cost data are not available. This allows us to constrain as many parameters as possible to those of Bueno de Mesquita and Lalman's theory, while estimating parameters in lieu of the missing cost data. As Table 3 displays, the estimates of four of the six cost parameters are at the constraint boundaries: $\gamma_2 = \alpha_1 = \alpha_2 = \tau_1 = 0$.[30] The BdM&L-BR model does slightly better than the null model in terms of the log-likelihood value ($\ln L = -1286.5$), but the percentage correctly predicted ($CP = 23\%$) is slightly worse than in the null model (assuming SQ is always predicted). As we saw previously, the percentage correctly predicted for Bueno de Mesquita and Lalman's

[28] The constraints are imposed to keep the parameters and utilities consistent with Bueno de Mesquita and Lalman's theory. In some situations, parameters should be left free, for example, to determine whether effects are in the expected direction, but in this case it does not make sense to think of costs adding to a nation's utility or of a nation's demands being negatively related to its utility. Maximum likelihood estimation was also conducted without constraints on the parameters. A number of the parameter estimates had values prohibited under Bueno de Mesquita and Lalman's constraints. In general, however, the results were not substantively different from those presented here. Results are available upon request from the author.

[29] As a measure of model fit, the log-likelihood value is the better indicator in this type of situation. Nevertheless, one can think of the percentage correctly predicted here as using a rule of thumb to determine a single prediction for each observation. The values of CP reported in Table 3 should then be interpreted as a measure of goodness of fit subject to the outcome selection criterion.

[30] Because the estimates are at the constraint boundaries, the standard errors of the estimates cannot be calculated. I denote these cases with a period.

TABLE 3. Results of the Strategic Statistical Model of the International Interaction Game

Term		Null	Bueno de Mesquita and Lalman Bounded Rationality Model: Nation 1	Nation 2	Unrestricted Model: Nation 1	Nation 2
$U(SQ')$	β_{i1}	0	1	1	0 ·	.87* (.24)
$U(\Delta_i)$	β_{i2}	0	1	1	.29 (.18)	0 ·
$U(\Delta_j)$	β_{i3}	0	1	1	2.55* (.38)	.91* (.15)
$P_iU(\Delta_i)$	β_{i4}	0	1	1	0 ·	0 ·
$(1 - P_i)U(\Delta_j)$	β_{i5}	0	1	1	2.68* (.66)	1.07* (.22)
ϕ_iP_i	β_{i6}	0	−1	−1	0 ·	0 ·
$(1 - P_i)$	γ_i	0	−1.50* (.53)	0 ·	−.42 (.54)	0 ·
	α_i	0	0 ·	0 ·	0 ·	0 ·
	τ_i	0	0 ·	−.39 (.26)	0 ·	0 ·
	$\ln L =$	−1,292.7	−1,286.5		−1,207.9	
	$2(\ln L - \ln L^0) =$				169.5	
	$df =$				14	
	$p =$				0.000	
	$CP =$	14%,[a] 34%[b]	23%		29%	

Note: $N = 707$. The null model restricts all coefficients to zero and corresponds to state leaders who make their decisions at each node by flipping a coin. The BdM&L-BR model restricts the β_{ij} coefficients to those assumed in Bueno de Mesquita and Lalman (1992) and estimates the cost parameters, γ_i, α_i, and τ_i, in lieu of the missing cost data. The unrestricted model allows the terms in each nation's utility functions to have different effects. Standard errors are in parentheses. Parameter estimates at their constraint boundary have standard errors denoted by a period. *$p < .01$.
[a] Acq_2 predicted.
[b] SQ predicted.

model under subgame perfection falls between 1% and 41%.

There are a number of obvious limitations to the null model and to the BdM&L-BR model. Thus far, we have made assumptions concerning the β_{ij} coefficients and compared how well the various models fit the data. Substantively, we may want to estimate the effect of all the explanatory variables and identify the extent to which that improves model fit. In addition, a valid point of contention would be that the BdM&L-BR model is not the same as in Bueno de Mesquita and Lalman (1992). Although the β_{ij} are restricted at the levels of those in Bueno de Mesquita and Lalman (1992), a different equilibrium solution concept is being used (i.e., LQRE versus subgame perfection). Moreover, the restriction of both the β_{ij} and λ implies some fixed level of bounded rationality, hence the appellation here.

The unrestricted model allows us to address all these issues, since all β_{ij} and cost parameters are estimated. Recall that subgame perfection is a special case of the LQRE for finite games of perfect information (i.e., when $\lambda = \infty$). Therefore, if the model is correctly specified, and if the states are highly rational in their decision making, then we should expect to see estimates of β_{ij} and the cost parameters that are very high in magnitude, representing the joint effect of a high λ and the true parameters.

Table 3 shows mixed results for the unrestricted model. On the one hand, there are a number of statistically significant parameters, and we can compare the relative size of the effects for nation 1 and nation 2. For example, it appears that the utility of the other nation's demand, $U(\Delta_j)$, has a larger effect in nation 1's utility calculations than in nation 2's. The same can be said for the effect of $(1 - P_i)U(\Delta_j)$. On the other hand, of the eighteen parameters estimated, eleven are at the zero boundary, individually closer to the null model parameters than to the Bueno de Mesquita and Lalman parameters. Furthermore, the remaining parameters are not so high in magnitude that we can infer both that the model is correctly specified and that the decision makers are highly rational.

We may still ask how well the unrestricted model fits the data. The log-likelihood ratio test shows that the amount of variation explained by the unrestricted model over the null model is highly significant ($p = 0.000$) and due to systematic variation. Yet, its ability to explain the variation in outcomes is still fairly weak. Note that the log-likelihood value ($\ln L = -1207.9$) is still relatively close to that of the null model ($\ln L^0 = -1292.7$). Using the aforementioned criterion for determining predicted outcomes, the unrestricted model predicts 29% of the outcomes correctly, which is better than the BdM&L-BR model and closer to the upper bound of the Bueno de Mesquita and Lalman model under subgame perfection. Nevertheless, predicting the modal category, SQ, would yield a higher percentage correctly predicted. Thus, even the unrestricted model does not fare well.

These results, in combination with those assuming subgame perfection, suggest there is less support for the international interaction game than Bueno de Mesquita and Lalman claim.[31] It is difficult to say where the international interaction game needs the most improvement, since it has so many "moving parts." At least three general areas might be beneficial for future research: (1) the game structure, (2) the equilibrium solution concept, and (3) the data. Concerning the first, the game structure simply may not be appropriate. The outcomes may reflect uncertainty not captured by the model of perfect information analyzed here. Moreover, as Bueno de Mesquita and Lalman (1992, 281) note themselves, it is a dyadic structure, and more than two parties may be involved in the strategic interaction, such as alliance partners.

Although the issue is rarely addressed, the solution concept is also an important part of the model. Traditional formal specifications of international conflict assume that states behave perfectly rationally. Yet, experimental economists have developed alternative equilibrium solution concepts (e.g., the LQRE) precisely because of observed deviations from Nash behavior. It may be that a solutions concept such as the LQRE, which allows for a range of rationality or "error-proneness" in behavior, is more appropriate for modeling international interaction, but that remains to be seen.

Finally, the data employed here have gone through a number of thoughtful but complex operationalizations that must be included under the larger rubric of "the model." Data on militarized interstate disputes are categorized in a particular way to obtain the outcomes used as the dependent variable. Data on alliance commitments and national material capabilities are transformed into measures of risk propensity and utilities for demands and the status quo. All of these require modeling assumptions that may or may not be true.[32] Lastly, cost data surely would be an important addition to any model such as this. The analyses in this section have attempted to give Bueno de Mesquita and Lalman the most benefit of the doubt by estimating the cost parameters that yield the best model fit, but this is never a substitute for the real data. Unfortunately, this type of cost data is difficult, if not impossible, to obtain.

[31] See also Smith 1997 for a similar assessment but using a different model.

[32] For example, all the explanatory variables rely on the use of τ_b as a measure of alliance policy similarity. Signorino and Ritter (1997) identify a number of problems with using τ_b for this purpose and suggest an alternative measure of policy portfolio similarity.

CONCLUSION

Although strategic interaction is at the heart of most international relations theory, it has largely been missing from much empirical analysis in the field. This article attempts to bridge the gap between the theory and estimation of a large class of models of international conflict. The LQRE solution concept applied here is a promising approach for incorporating the structure of theorized strategic interdependence into statistical models. Using this method, I showed that failure to incorporate the strategic interaction in statistical models can lead to parameter estimates with incorrect substantive interpretations as well as fitted values that are at times wildly wrong. The method was also used to reanalyze Bueno de Mesquita and Lalman's (1992) international interaction game.

The techniques applied here are relatively new, and much remains to be done in applying, refining, and extending them. One area of considerable importance is the analysis of misspecification. I have shown cases in which logit produces dramatically incorrect results. It would be interesting to know whether there are plausible general conditions in which the logit parameter estimates have the correct substantive interpretation and the logit probabilities closely approximate the strategic choice probabilities. In short, when does structure *really* matter in statistical analysis? Under what conditions can we make do with logit?

In addition, throughout the analyses here, we have not had to worry about multiple equilibria. The games employed are games of perfect information, which resulted in a single subgame perfect equilibrium or logit quantal response equilibrium for any value of the utilities (but assuming no ties). To determine probabilities over outcomes for the likelihood function of our statistical model, there was only one set of equilibrium probabilities to consider. For games of imperfect information, however, multiple equilibria will often arise. An important question then is: Given multiple equilibria, how are we to assign probabilities over outcomes? This is a surprisingly little-studied area in political science or in econometrics, but it must be addressed if we are to conduct similar statistical analyses of more realistic models of international interaction.

Finally, the main point of this article has been that structure matters not only in our theories of strategic interaction but also in our statistical tests of those theories. This, however, may present a double-edged sword for positive international relations theory more broadly. On the one hand, it implies that we must ensure that our statistical models are consistent with our theories. We know that small changes to a theory (e.g., the number of players, the sequence of their moves, the choices and information available to them, and their incentives) can have large consequences in what the theory predicts. If a theory is vague, then it is unclear what statistical model would be consistent with that theory. Therefore, if we want to ensure consistency between a theory and a statistical model, we must be as precise as possible in the specification of the theory. Given the requirement for theoretical precision, how are we to specify and test strategic theories without doing so formally? The statistical analyses conducted here required that the interaction be specified as an extensive form game. Although this may not be the only way to formalize and test strategic interaction, even methods using simultaneous systems with dummy variables (Bjorn and Vuong 1984; Bresnahan and Reiss 1991) require a high degree of mathematical specificity.

On the other hand, although the call for increased formalization of theories may be welcomed by many (but certainly not all) positivists, the importance of structure also seems to cut the other way. Consider the typical derivation and analysis of a positive theory. One major assumption generally held—indeed, held throughout this article—is that the structure of the model remains constant across all observations in the data. In other words, although the utilities may vary from case to case, it is assumed that the same "game" structure is being played in every situation across time. It does not seem unreasonable to suspect, however, that the *true* game structure changes over time and place. If even small changes in structure can make a large difference in likely outcomes, and if the true structure of the strategic interaction changes from observation to observation in our data, then what are we to make of any statistical results predicated on the assumption of a fixed game? The reaction to this should *not* be to use more poorly specified theories and statistical tests, which suffer from all the maladies mentioned so far. Rather, the problem of a changing structure would seem to be an even more difficult version of the specification problem identified previously. Nevertheless, it is an issue that those interested in positive international relations theory must address if ultimately we are to make progress in this general endeavor.

APPENDIX A: DERIVATION OF THE STRATEGIC CHOICE PROBABILITIES FOR THE CRISIS GAME

Derivation of the LQRE choice probabilities for the crisis game shown in Figure 2 is similar to the derivation of multinomial logit probabilities when assuming random utility. The random utility assumptions were identified previously in this article. See McFadden (1974) and Maddala (1983, 60–1) for similar derivations but of multinomial logit probabilities. I provide a derivation from first principles of only two of the choice probabilities, since the other two follow in the exact same manner. In the following, although it is an abuse of notation, I will refer to the (expected) utility of nodes rather than strategies. For example, $U_2(4)$ will refer to the utility to state 2 of reaching node 4 or, equivalently, of choosing to fight at node 2. This greatly simplifies the notation.

The probability p_4 that state 1 will choose to fight at node 4 is

$$\begin{aligned} p_4 &= \Pr[U_1^*(W2) > U_1^*(C1)] \\ &= \Pr[U_1(W2) + \varepsilon_{w2} > U_1(C1) + \varepsilon_{c1}] \\ &= \Pr[\varepsilon_{c1} < U_1(W2) - U_1(C1) + \varepsilon_{w2}]. \end{aligned} \tag{A-1}$$

Let $f(\varepsilon_{w2}, \varepsilon_{c1})$ and $F(\varepsilon_{w2}, \varepsilon_{c1})$ be the joint pdf and cdf of $(\varepsilon_{w2}, \varepsilon_{c1})$, and let $F_{w2}(\varepsilon_{w2}, \varepsilon_{c1}) = dF(\varepsilon_{w2}, \varepsilon_{c1})/d\varepsilon_{w2}$. Then

$$\begin{aligned} p_4 &= \int_{-\infty}^{\infty} \int_{-\infty}^{U_1(W2)-U_1(C1)+\varepsilon_{w2}} f(\varepsilon_{w2}, \varepsilon_{c1})\, d\varepsilon_{c1}\, d\varepsilon_{w2} \\ &= \int_{-\infty}^{\infty} F_{w2}[\varepsilon_{w2}, U_1(W2) - U_1(C1) + \varepsilon_{w2}]\, d\varepsilon_{w2}. \end{aligned} \tag{A-2}$$

This holds for any joint density $f(\varepsilon_{w2}, \varepsilon_{c1})$. Here, we now assume ε is independently and identically distributed according to a type I extreme value density: $f(\varepsilon) = \lambda \exp(-\lambda\varepsilon - e^{-\lambda\varepsilon})$, and $F(\varepsilon) = \exp(-e^{-\lambda\varepsilon})$, with $E[\varepsilon] = \gamma/\lambda$, and $V[\varepsilon] = \pi^2/(6\lambda^2)$, where γ is Euler's constant, approximately equal to 0.577.

With that assumption, equation A-2 becomes

$$\begin{aligned} p_4 &= \int_{-\infty}^{\infty} f(\varepsilon_{w2}) F[U_1(W2) - U_1(C1) + \varepsilon_{w2}]\, d\varepsilon_{w2} \\ &= \int_{-\infty}^{\infty} \exp\{-\lambda\varepsilon_{w2} - e^{-\lambda\varepsilon_{w2}}\} \\ &\qquad \exp\{-e^{-\lambda[U_1(W2)-U_1(C1)+\varepsilon_{w2}]}\}\, d\varepsilon_{w2} \\ &= \int_{-\infty}^{\infty} \exp\{-\lambda\varepsilon_{w2} - e^{-\lambda\varepsilon_{w2}}[1 \\ &\qquad + e^{-\lambda[U_1(W2)-U_1(C1)]}]\}\, d\varepsilon_{w2} \\ &= \int_{-\infty}^{\infty} \exp\{-\lambda\varepsilon_{w2} \\ &\qquad - e^{-\lambda\varepsilon_{w2}}\left[\frac{e^{\lambda U_1(W2)} + e^{\lambda U_1(C1)}}{e^{\lambda U_1(W2)}}\right]\}\, d\varepsilon_{w2}. \end{aligned}$$

Now, let $z = \ln[(e^{\lambda U_1(W2)} + e^{\lambda U_1(C1)})/e^{\lambda U_1(W2)}]$. Then

$$\begin{aligned} p_4 &= \int_{-\infty}^{\infty} \exp\{-\lambda\varepsilon_{w2} - e^{-\lambda\varepsilon_{w2}}e^{z}\}\, d\varepsilon_{w2} \\ &= \int_{-\infty}^{\infty} \exp\{-\lambda\varepsilon_{w2} - e^{-\lambda\varepsilon_{w2}+z}\}\, d\varepsilon_{w2}. \end{aligned}$$

Letting $\varepsilon^* = \lambda\varepsilon_{w2} - z$,

$$\begin{aligned} p_4 &= \int_{-\infty}^{\infty} (e^{-z}e^{z}) \exp\{-\lambda\varepsilon_{w2} - e^{-(\lambda\varepsilon_{w2}-z)}\}\, d\varepsilon_{w2} \\ &= e^{-z} \int_{-\infty}^{\infty} \exp\{-(\lambda\varepsilon_{w2} - z) - e^{-(\lambda\varepsilon_{w2}-z)}\}\, d\varepsilon_{w2} \\ &= e^{-z} \int_{-\infty}^{\infty} \exp\{-\varepsilon^* - e^{-\varepsilon^*}\}\, d\varepsilon^* \\ &= e^{-z} \int_{-\infty}^{\infty} f(\varepsilon^*)\, d\varepsilon^* \\ &= e^{-z} \\ &= \frac{e^{\lambda U_1(W2)}}{e^{\lambda U_1(W2)} + e^{\lambda U_1(C1)}}. \end{aligned} \tag{A-3}$$

To derive the probability that state 2 chooses to fight (f_2) if it is at node 2, we now have to consider not only the utilities of outcomes but also the expected utility of reaching the nonterminal node 4. A slightly condensed derivation of the probability that state 2 chooses to fight is given as follows:

$$\begin{aligned} q_2 &= \Pr[U_2^*(4) > U_2^*(SQ)] \\ &= \Pr[U_2(4) + \varepsilon_4 > U_2(SQ) + \varepsilon_{sq}] \\ &= \Pr[\varepsilon_{sq} < U_2(4) - U_2(SQ) + \varepsilon_4] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{U_2(4)-U_2(SQ)+\varepsilon_4} f(\varepsilon_4, \varepsilon_{sq})\, d\varepsilon_{sq}\, d\varepsilon_4 \\ &= \int_{-\infty}^{\infty} F_4[\varepsilon_4, U_2(4) - U_2(SQ) + \varepsilon_4]\, d\varepsilon_4 \\ &= \int_{-\infty}^{\infty} f(\varepsilon_4)F[U_2(4) - U_2(SQ) + \varepsilon_4]\, d\varepsilon_4 \\ &= \frac{e^{\lambda U_2(4)}}{e^{\lambda U_2(4)} + e^{\lambda U_2(SQ)}}. \end{aligned} \quad \textbf{(A-4)}$$

The expected utility to 2 of reaching node 4 depends on whether state 1 will fight: $U_2(4) = p_4U_2(W2) + (1 - p_4)U_2(C1)$. Hence, the probability that state 2 will choose to fight at node 2 is

$$q_2 = \frac{e^{\lambda[p_4U_2(W2)+(1-p_4)U_2(C1)]}}{e^{\lambda[p_4U_2(W2)+(1-p_4)U_2(C1)]} + e^{\lambda U_2(SQ)}}. \quad \textbf{(A-5)}$$

The probability that state 2 chooses to fight at node 3 and the probability that state 1 decides to fight at node 1 are derived in exactly the same manner as above, giving

$$q_3 = \frac{e^{\lambda U_2(W1)}}{e^{\lambda U_2(W1)} + e^{\lambda U_2(C2)}} \quad \textbf{(A-6)}$$

$$p_1 = \frac{e^{\lambda[q_3U_1(W1)+(1-q_3)U_1(C2)]}}{e^{\lambda[q_3U_1(W1)+(1-q_3)U_1(C2)]} + e^{\lambda\{q_2[p_4U_1(W2)+(1-p_4)U_1(C1)]+(1-q_2)U_1(SQ)\}}}. \quad \textbf{(A-7)}$$

APPENDIX B: UTILITIES AND STRATEGIC CHOICE PROBABILITIES FOR THE INTERNATIONAL INTERACTION GAME

This appendix provides the nations' utilities for the nodes associated with the international interaction game shown in Figure 6 and the resulting LQRE choice probabilities. With some abuse of notation, let $U_i(n)$ be the utility of nation i for the outcome at node n. When n is a nonterminal node, $U_i(n)$ represents the expected utility for node n over the actions available at that node. The terminal node utilities were previously given in the specification of the international interaction game. The expected utilities over the nonterminal nodes are

$$\begin{aligned} U_i(2) &= q_2U_i(4) + (1 - q_2)U_i(SQ); \\ U_i(3) &= q_3U_i(5) + (1 - q_3)U_i(Acq_2); \\ U_i(4) &= p_4U_i(6) + (1 - p_4)U_i(Acq_1); \\ U_i(5) &= p_5U_i(10) + (1 - p_5)U_i(9); \\ U_i(6) &= q_6U_i(8) + (1 - q_6)U_i(7); \\ U_i(7) &= p_7U_i(11) + (1 - p_7)U_i(Nego); \\ U_i(8) &= p_8U_i(War) + (1 - p_8)U_i(Cap_1); \\ U_i(9) &= q_9U_i(12) + (1 - q_9)U_i(Nego); \\ U_i(10) &= q_{10}U_i(War) + (1 - q_{10})U_i(Cap_2); \\ U_i(11) &= q_{11}U_i(War) + (1 - q_{11})U_i(Cap_2); \\ U_i(12) &= p_{12}U_i(War) + (1 - p_{12})U_i(Cap_1). \end{aligned}$$

All choice probabilities are with respect to the positive action (versus "not" action), as in Figure 6. Let p_i be the probability that nation 1 chooses the positive action at node i. Similarly, q_j represents the probability that nation 2 chooses the positive action at node j. Finally, λ is set to one for identification of the coefficients during estimation. Nation 1's LQRE choice probabilities are then

$$\begin{aligned} p_1 &= \frac{e^{U1(3)}}{e^{U1(3)} + e^{U1(2)}}; \\ p_4 &= \frac{e^{U1(6)}}{e^{U1(6)} + e^{U1(Acq1)}}; \\ p_5 &= \frac{e^{U1(10)}}{e^{U1(10)} + e^{U1(9)}}; \\ p_7 &= \frac{e^{U1(11)}}{e^{U1(11)} + e^{U1(Nego)}}; \\ p_8 &= \frac{e^{U1(War)}}{e^{U1(War)} + e^{U1(Cap1)}}; \\ p_{12} &= p^8; \end{aligned}$$

and nation 2's LQRE choice probabilities are

$$\begin{aligned} q_2 &= \frac{e^{U2(4)}}{e^{U2(4)} + e^{U2(SQ)}}; \\ q_3 &= \frac{e^{U2(5)}}{e^{U2(5)} + e^{U2(Acq2)}}; \\ q_6 &= \frac{e^{U2(8)}}{e^{U2(8)} + e^{U2(7)}}; \\ q_9 &= \frac{e^{U2(12)}}{e^{U2(12)} + e^{U2(Nego)}}; \\ q_{10} &= \frac{e^{U2(War)}}{e^{U2(War)} + e^{U2(Cap2)}}; \\ q_{11} &= q_{10}. \end{aligned}$$

Again, the probability of each terminal node is then given by the product of the probabilities of actions leading to it:

$$\begin{aligned} \Pr(SQ) &= (1 - p_1)(1 - q_2); \\ \Pr(Acq_1) &= (1 - p_1)q_2(1 - p_4); \\ \Pr(Acq_2) &= p_1(1 - q_3); \\ \Pr(Nego) &= (1 - p_1)q_2p_4(1 - q_6)(1 - p_7) + p_1q_3(1 - p_5)(1 - q_9); \\ \Pr(Cap_1) &= (1 - p_1)q_2p_4q_6(1 - p_8) \\ &\quad + p_1q_3(1 - p_5)q_9(1 - p_{12}); \\ \Pr(Cap_2) &= (1 - p_1)q_2p_4(1 - q_6)p_7(1 - q_{11}) \\ &\quad + p_1q_3p_5(1 - q_{10}); \\ \Pr(War_1) &= (1 - p_1)q_2p_4(1 - q_6)p_7q_{11} + p_1q_3p_5q_{10}; \\ \Pr(War_2) &= (1 - p_1)q_2p_4q_6p_8 + p_1q_3(1 - p_5)q_9p_{12}. \end{aligned}$$

REFERENCES

Beck, Nathaniel, Jonathan Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42 (October): 1260–88.

Bjorn, Paul A., and Quang H. Vuong. 1984. "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation." Social Science Working Paper 537. California Institute of Technology.

Bremer, Stuart. 1992. "Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War, 1816–1965." *Journal of Conflict Resolution* 36 (June): 309–41.

Bresnahan, Timothy F., and Peter C. Reiss. 1991. "Empirical Models of Discrete Games." *Journal of Econometrics* 48 (April/May): 57.

Bueno de Mesquita, Bruce, and David Lalman. 1988. "Empirical Support for Systemic and Dyadic Explanations of International Conflict." *World Politics* 41 (October): 1–20.

Bueno de Mesquita, Bruce, and David Lalman. 1992. *War and Reason*. New Haven, CT: Yale University Press.

Bueno de Mesquita, Bruce, James D. Morrow, and Ethan R. Zorick. 1997. "Capabilities, Perception, and Escalation." *American Political Science Review* 91 (March): 15–27.

Chen, Hsiao Chi, James W. Friedman, and Jacques-Francois Thisse. 1997. "Boundedly Rational Nash Equilibrium: A Probabilistic Choice Approach." *Games and Economic Behavior* 18 (January): 32–54.

Fearon, James D. 1994. "Signaling versus the Balance of Power and Interests." *Journal of Conflict Resolution* 38 (June): 236–69.

Fey, Mark, Richard D. McKelvey, and Thomas R. Palfrey. 1996. "An Experimental Study of the Constant-Sum Centipede Games." *International Journal of Game Theory* 25 (3): 269–87.

Goldstein, Joshua S. 1991. "Reciprocity in Superpower Relations: An Empirical Analysis." *International Studies Quarterly* 35 (June): 195–209.

Goldstein, Joshua S., and John R. Freeman. 1990. *Three-Way Street: Strategic Reciprocity in World Politics*. Chicago: University of Chicago Press.

Goldstein, Joshua S., and John R. Freeman. 1991. "U.S.-Soviet-Chinese Relations: Routine, Reciprocity, or Rational Expectations?" *American Political Science Review* 85 (March): 17–35.

Hagan, J. D. 1994. "Domestic Political Systems and War Proneness." *International Studies Review* 38 (October): 183–207.

Huth, Paul. 1990. "The Extended Deterrent Value of Nuclear Weapons." *Journal of Conflict Resolution* 34 (June): 270–90.

Huth, Paul, Christopher Gelpi, and D. Scott Bennett. 1993. "The Escalation of Great Power Militarized Disputes." *American Political Science Review* 87 (September): 609–23.

Kagan, Donald. 1995. *On the Origins of War and the Preservation of Peace*. New York: Anchor Books.

Kim, Woosang, and Bruce Bueno de Mesquita. 1995. "How Perceptions Influence the Risk of War." *International Studies Quarterly* 39 (March): 51–65.

Kim, Woosang, and James D. Morrow. 1992. "When Do Power Shifts Lead to War?" *American Journal of Political Science* 36 (November): 896–922.

King, Gary. 1989a. "Event Count Models for International Relations." *International Studies Quarterly* 33 (June): 123–47.

King, Gary. 1989b. *Unifying Political Methodology*. Cambridge: Cambridge University Press.

King, Gary. 1989c. "Variance Specification in Event Count Models." *American Journal of Political Science* 33 (June): 762–84.

Kissinger, Henry. 1994. *Diplomacy*. New York: Simon & Schuster.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Maoz, Zeev, and Bruce Russett. 1993. "Normative and Structural Causes of Democratic Peace, 1946–1986." *American Political Science Review* 87 (September): 624–38.

Martin, Lisa. 1992. *Coersive Cooperation*. Princeton, NJ: Princeton University Press.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, ed. Paul Zarambka, New York: Academic Press.

McFadden, Daniel. 1976. "Quantal Choice Analysis: A Survey." *Annals of Economic and Social Measurement* 5 (Fall): 363–90.

McGinnis, Michael D., and John T. Williams. 1989. "Change and Stability in Superpower Rivalry." *American Political Science Review* 83 (December): 1101–23.

McKelvey, Richard D., and Thomas R. Palfrey. 1995. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior* 10 (July): 6–38.

McKelvey, Richard D., and Thomas R. Palfrey. 1996. "A Statistical Theory of Equilibrium in Games." *Japanese Economic Review* 47 (2): 186–209.

McKelvey, Richard D., and Thomas R. Palfrey. 1998. "Quantal Response Equilibria for Extensive Form Games." *Experimental Economics* 1 (1): 9–41.

Morgan, T. Clifton, and Sally H. Campbell. 1991. "Domestic Structure, Decisional Constraints, and War." *Journal of Conflict Resolution* 35 (June): 187–211.

Morrow, James D. 1991. "Alliances and Asymmetry: An Alternative to the Capability Aggregation Model of Alliances." *American Journal of Political Science* 35 (November): 904–33.

Pudney, Stephen. 1989. *Modelling Individual Choice: The Econometrics of Corners, Kinks, and Holes*. Oxford: Basil Blackwell.

Raymond, G. A. 1994. "Democracies, Disputes, and Third-Party Intermediaries." *Journal of Conflict Resolution* 38 (March): 24–42.

Rosenthal, R. W. 1989. "A Boundedly-Rational Approach to the Study of Noncooperative Games." *International Journal of Game Theory* 18 (3): 273.

Russett, Bruce. 1993. *Grasping the Democratic Peace*. Princeton, NJ: Princeton University Press.

Signorino, Curtis S. 1998. "Statistical Analysis of Finite Choice Models in Extensive Form." Paper presented at the 1998 annual meeting of the American Political Science Association, Boston.

Signorino, Curtis S., and Jeffrey M. Ritter. N.d. "Tau-b or Not Tau-b: Measuring the Similarity of Foreign Policy Positions." *International Studies Quarterly*. Forthcoming.

Smith, Alastair. 1995. "Alliance Formation and War." *International Studies Quarterly* 39 (December): 405–25.

Smith, Alastair. 1997. "Political Selection: The Effect of Strategic Choice on the Escalation of International Crises." Paper presented at the 1997 annual meeting of the Midwest Political Science Association, Chicago.

Taylor, A. J. P. 1954. *The Struggle for Mastery in Europe*. Oxford: Oxford University Press.

Thucydides. 1972. *History of the Peloponnesian War*. New York: Penguin Books.

Ward, Michael D., and Sheen Rajmaira. 1992. "Reciprocity and Norms in U.S.–Soviet Foreign Policy." *Journal of Conflict Resolution* 36 (June): 342–68.

Williams, John T., and Michael D. McGinnis. 1988. "Sophisticated Reaction in the U.S.–Soviet Arms Race: Evidence of Rational Expectations." *American Journal of Political Science* 32 (November): 968–95.