



## *Eugene: A conceptual manual*

D. Scott Bennett & Allan C. Stam

To cite this article: D. Scott Bennett & Allan C. Stam (2000) *Eugene: A conceptual manual*, International Interactions, 26:2, 179-204, DOI: [10.1080/03050620008434965](https://doi.org/10.1080/03050620008434965)

To link to this article: <https://doi.org/10.1080/03050620008434965>



Published online: 09 Jan 2008.



Submit your article to this journal [↗](#)



Article views: 606



View related articles [↗](#)



Citing articles: 15 View citing articles [↗](#)

## **EUGENE: A CONCEPTUAL MANUAL**

**D. SCOTT BENNETT\***

*Department of Political Science  
The Pennsylvania State University  
107 Burrowes Building  
University Park, PA 16802-6200  
Tel.: 814-865-6566  
Fax: 814-863-8979  
email: sbennett@psu.edu*

**ALLAN C. STAM\***

*Government Department  
Dartmouth College  
202 Silsby Hall  
Hanover, NH 03755  
Tel.: (603)646-2544  
email: allan.stam@dartmouth.edu*

*(In final form March 9, 2000)*

The study of international relations using quantitative analysis relies, in part, on the availability of comprehensive and easily manipulable data sets. To execute large-*n* statistical tests of hypotheses, data must be available on the variables of interest, and those data must be manipulated into a suitable format to allow the inclusion of appropriate control variables as well as variables of central theoretical interest. This paper introduces software designed to eliminate many of the difficulties commonly involved in constructing large international relations data sets, and with the unavailability of data on expected utility theories of war.

---

\*The order of the authors' names is alphabetical and is not intended to connote principal authorship.

In order to solve these two problems, we developed *EUGene* (the Expected Utility Generation and Data Management Program). *EUGene* is a stand-alone Microsoft Windows based program for the construction of annual data sets for use in quantitative studies of international relations. It generates data for variables necessary to incorporate key variables from implementations of the so-called "expected utility theory of war" into broader analyses of international conflict. *EUGene* is also designed to make building international relations data sets simple. It accomplishes this by automating a variety of tasks necessary to integrate several data building blocks commonly used in tests of international relations theories.

**KEY WORDS:** *expected utility; software; international relations; statistical analysis*

## INTRODUCTION

The study of international relations using quantitative analysis relies on the availability of comprehensive and easily manipulable data sets. While seemingly trivial, this statement highlights one of the many technical hurdles IR scholars face. To execute large-*n* statistical tests of hypotheses, data must be available on the variables of interest, and those data must be manipulated into a suitable format to allow the inclusion of appropriate control variables as well as variables of central theoretical interest. Frequently, however, the process of preparing data sets for analysis is cumbersome, particularly data sets with many cases and with variables that come from a variety of sources. Frequently control variables are excluded from analysis, not for theoretical or statistical reasons, but simply because cumbersome data manipulation tasks preclude optimal test design. The somewhat daunting task of preparing large data sets can have the effect of turning scholars into technicians for substantial periods of time rather than remaining focused on theory development and research design improvements. Alternately, for many scholars the barrier to entry into the realm of quantitative research is sufficiently high to preclude any sophisticated analysis at all.

One of the most important theories for which data have not been available for testing, either as a theory of interest or as an essential control variable, is the so-called expected utility theory of war (Bueno de Mesquita, 1981, 1985; Bueno de Mesquita and Lalman, 1992). While being one of the most important current theories of international conflict, Bueno de Mesquita's variant of expected utility theory has until recently only been tested on a small set of cases.<sup>1</sup> This is because expected utility data have not been available for all dyads and over the full time span of most

other international relations data sets (1816 through the 1980s or 1990s, depending on the data set). Replication of Bueno de Mesquita's data has been slow, in part because of substantial barriers inherent in replicating data construction algorithms and finding adequate computational capacity for creating these data. Perhaps just as importantly, none of the tests of other theories (democratic peace, power transition, arms races, etc.) has been able to include controls for the measures that Bueno de Mesquita claim to be powerful predictors of interstate conflict. Several of these measures may be correlated with the other variables of interest, potentially introducing substantial omitted variable bias in much of the recent quantitative work on war and dispute initiation.

This paper introduces software designed to eliminate many of the difficulties commonly involved in constructing large international relations data sets, and to make available complete data on Bueno de Mesquita and Lalman's most recent version of the so-called expected utility theory of war. We have developed *EUGene* (the Expected Utility Generation and Data Management Program) in order to solve these two problems. *EUGene* is a stand-alone Microsoft Windows based program for the construction of annual data sets for use in quantitative studies of international relations. It generates the data needed to incorporate key variables from Bueno de Mesquita and Lalman's work (1992) into broader analyses of international conflict. *EUGene's* purpose is to make the construction of international relations data sets simple. It accomplishes this by automating a variety of tasks necessary to integrate several data building blocks commonly used in tests of international relations theories.

Users of the program simply specify the type of data set they would like to create by selecting from a series of drop-down menus. Choices include the unit of analysis, population of cases, variables to include, and output format. The program assembles a data set according to these user specifications and outputs it for analysis in other statistical software packages. The program also creates command files that make reading the data into other statistical programs automatic. Users of *EUGene* do not need to be able to write a single line of computer code in order to merge data, read data from input files of varying formats, or convert data into common units of analysis. By reducing the time necessary to carry out routine data set construction tasks, *EUGene* allows users to proceed more rapidly to the analysis stage, and allows scholars to spend more time on theory development and asking new research questions than on data management. Below, we describe *EUGene*, its functions and operation, and key details of its data generation algorithms. *EUGene* is available as freeware

from the American Political Science Association Conflict Processes Section's home page at [www.eugenesoftware.org](http://www.eugenesoftware.org).

## SIMPLIFYING DATA MANAGEMENT

The first of *EUGene*'s two major goals is to facilitate analysis in international relations by minimizing the relatively unproductive time scholars must spend on routine tasks of data set construction and manipulation. An obvious but critically important first step in any quantitative data analysis is to create an accurate and appropriate data set. Such data set construction is frequently an onerous and time-consuming task. *EUGene* serves as a tool to facilitate and simplify the process of merging and creating data sets in international relations, especially data sets created with the directed dyad-year as the unit of analysis. Scholars have increasingly come to use data sets based on the dyad-year to conduct quantitative analyses. This is because dyadic interaction lies at the heart of strategic international behavior, and because it is possible to combine explanations from multiple levels of analysis in one quantitative study. In particular, *directed* dyad-year studies can include variables from the individual level (e.g., polity type), dyadic level (e.g., balance of forces, distance), and system level (e.g., polarity, system uncertainty) in studies of conflict (see, e.g., Bremer, 1992; Huth, Bennett and Gelpi, 1992; Huth, Gelpi and Bennett, 1993; Maoz and Russett, 1993; Oneal and Russett, 1997; Beck, Katz and Tucker, 1997; Ray and Wang, 1998; Bennett and Stam, 2000c).<sup>2</sup> Most scholars rely on annual data both because data are widely available at this level of temporal aggregation, and because the year represents a natural political break due to budget cycles, electoral cycles, and the presence of winter that in many areas hampers military action.<sup>3</sup>

Unfortunately, several aspects of creating dyadic data sets make the task difficult for many researchers. On the independent variable side, creating dyadic data sets involves merging data and renaming variables, typically from multiple monadic data sets that require conversion to a dyadic form. Even before this step, appropriate sets of dyad-years must be created, a task requiring users to create dyads from lists of states while verifying that both members of the dyad are indeed system members when they are coded as dyads. Given that there are over 1.2 million dyad-years in the international system between 1816 and 1997, this is a task that must be automated, and must be done accurately. On the dependent variable side, the most common data sets with international conflict events (the Correlates of War Militarized Interstate Dispute data set and Interstate

War data set) have not been organized in dyadic form and must be converted into dyadic interactions. The necessary merges and conversions are not always straightforward, requiring users to make important decisions about coding that are sometimes not recognized, and serve as a significant barrier to wider theory development, theory testing, and graduate student training.

Because it has been difficult to easily create quantitative IR data sets that incorporate many variables, scholars have in the past been forced to spend nearly as much time creating their data sets as they do on theory development and analysis (and sometimes more). This slows the research process, and sometimes tends to turn researchers into technicians rather than scholars. *EUGene* was created as a menu-driven tool for Windows to make data set creation easier, with several specific goals in mind that we detail below. *EUGene* allows users to:

- Construct data sets with different units of analysis, and where appropriate convert input data of varying units of analysis to the selected unit, at the click of a button.
- To choose variables for inclusion in final data sets from a variety of input sources without manually writing code to merge various input data.
- To easily select subsets of data based on common criteria such as political relevance, time period, or great power status.
- To make clear the variety of critical but often unstated assumptions about the construction of key dependent variables and the inclusion of problematic cases that go into the construction of international relations data sets, and force users to make informed decisions about these items.
- To facilitate replication by providing a single program for data set creation that will produce the same results for all users, eliminating the problem of hidden or forgotten steps typically encountered when attempting replication.

### Unit of Analysis

The first choice made by users when creating a new data set in *EUGene* is the most fundamental of all research design decisions, namely the unit of analysis. *EUGene* allows users to choose to create data sets with the country-year, directed dyad-year, directed dispute-dyad, and directed dispute dyad-year as the unit of analysis.<sup>4</sup> By selecting these units of analysis, users can examine monadic time-series (by creating a country-year data

set), examine dispute initiation from a condition of peace or examine the duration of peace (by creating a directed dyad-year data set), examine the escalation of disputes (by creating a directed dispute-dyad data set), or examine the evolution and duration of disputes over time (by creating a directed dispute-dyad-year data set).

### Variables

*EUGene* allows users to specify the variables that are to be included in the output data set by clicking on a set of check boxes. The program as distributed allows users to choose from a set of over 60 variables from several of the most important international relations data sets. In Table I we detail the sources of the data *EUGene* assembles.

Users can select from over 60 variables when assembling a data set. These include Polity III democracy scores and ancillary components (Jaggers and Gurr, 1995), Correlates of War project capability data (Singer, Bremer and Stuckey, 1972), data on interstate distances, tau-b scores, risk attitude data, contiguity data, region, peace years (Beck, Katz and Tucker, 1997), expected utility values and international interaction game equilibrium predictions (Bueno de Mesquita and Lalman, 1992), and COW Militarized Interstate Dispute data (Jones *et al.*, 1996). The ability to include many variables in the output data set easily makes *EUGene* useful for scholars pursuing a variety of research agendas, not just those exploring applications of expected utility theory. For example, variables are available that figure into the democratic peace theory, power transition theory, balance of power theory, and theories of system structure, to name just a few.

### Population of Cases

*EUGene* allows users to specify the scope of the output data set as being either all dyads, or one of a number of subsets of countries and years. Users can specify a particular range of years for output (e.g., just 1945–1992, or 1816–1914), and can select from common subsets of countries (e.g., all dyads, politically relevant dyads, major power dyads, contiguous states, or a user-selected list such as rivals). Alternatively, users may generate all dyad-years and include variables in the output to allow selection at a later time. Being able to select these commonly-used subsets of cases allows users to conduct comparative analyses and explore the sensitivity of their results to factors like era or region.

**TABLE I**  
**Currently Included Variables and Data Sources**

Variable	Unit of Analysis	Source	Original Citation
Capabilities	country-year	COW national capabilities index (CINC)	Singer, Bremer and Stuckey (1972)
Contiguity/Distance	dyad-year	Contiguity from COW contiguity data set; Distance computed in EUGene	Small and Singer (1982)
Dyad Duration	dyad-year	Calculated in EUGene	Bueno de Mesquita (1981)
Expected Utility (War Trap)	dyad-year	Calculated in EUGene	Bueno de Mesquita (1985), Bueno de Mesquita and Lalman (1992)
Expected Utility (War and Reason)	directed dyad-year	Calculated in EUGene	Bueno de Mesquita and Lalman (1992)
International Interaction Game Equilibria	directed dyad-year	Calculated in EUGene	Small and Singer (1982)
Home Region	country-year	COW system membership and region list	
Relevant Region of Conflict	directed dyad-year	Calculated in EUGene	Bueno de Mesquita (1981)
Major Power Status	country-year	COW system membership list	Small and Singer (1982)
MID Data (initiation, joiner, MID Number, hostility level, etc.)	directed dyad-year	COW MID data set (v2.1)	Jones, Bremer and Singer (1996)
Peace Years	dyad-year	Calculated in EUGene	Beck, Katz, and Tucker (1997)
Political Relevance	dyad-year	Calculated in EUGene	Maoz and Russett (1993)
Regime/Polity Characteristics (democ, autoc, etc.)	country-year	Polity III	Jagers and Gurr (1995)
Uncertainty (based on region)	country-year	Calculated in EUGene	Bueno de Mesquita (1975); Bueno de Mesquita and Lalman (1992)
Risk Attitude (EUGene)	country-year	Calculated in EUGene	Bueno de Mesquita (1985)
Risk Attitude ( <i>War Trap Revisited Data</i> )	country-year	Bueno de Mesquita (through Tolstoy software)	Bueno de Mesquita (1985); Horn (1990)
Tau-B	dyad-year	Calculated in EUGene	Bueno de Mesquita (1975); Bueno de Mesquita (1981)
Tau-B with System Leader	country-year	Calculated in EUGene	Lemke and Reed (1998)



### Case Inclusion Criteria and Assumptions

A variety of other under-discussed issues arise in the context of creating dyad-year data sets, with most concerning the inclusion or exclusion of cases in a fashion related to the dependent variable and the censoring of cases. The first issue, mentioned above, is how to deal with years with ongoing militarized disputes. *EUGene* allows users to code ongoing dispute years as either dispute initiations or non-initiations for purposes of the dependent variable, but also allows users to either drop or include all dyad-years where the countries begin the year with an ongoing dispute (users may want to drop such cases if they believe that a new initiation would be censored by the ongoing dispute).

A second issue concerns the treatment of dyads where a state joins into a dispute that is already in progress. Should joiners be included for analysis in the same way as dispute initiators? The information conditions faced by joiners into disputes is fundamentally different than the conditions facing the initial participants, suggesting that perhaps the same model that explains dispute initiation will not explain dispute "joining." So perhaps we should omit joiners from analysis. *EUGene* also allows users to include or drop such cases by selecting a check box.

A final issue concerns "target vs. initiator" directed dyads. When one state initiates a dispute, it does so against a target state, creating a designated initiator A and target B. But when A initiates vs. B, it is less than clear how to include the directed dyad B vs. A. The problem is that A's initiation may remove B's ability to initiate a dispute against A in that same year, and we then do not know how to code the "initiation" variable for the B vs. A dyad because it is censored. *EUGene* gives users the option to include such target vs. initiator dyads in the data sets it creates, to drop them, or to include them only if there is a subsequent initiation by B vs. A (this would indicate that B in fact did have the opportunity to start a MID against A, and so B's behavior is not censored).

### Merging and Data Conversion

One difficulty with building data sets that combine variables is that input data sets frequently come with different units of analysis and in different formats, requiring conversion at a fundamental level in addition to simply merging. For example, some key IR data sets have the country-year as the unit of analysis (e.g., the Correlates of War national capability data, Gurr Polity data, or data on national risk attitude). Other data sets (or data

constructions) have the dyad as the unit of analysis, such as distance data, the Correlates of War contiguity data set, or data on expected utility. Still other data sets are distributed in a hybrid form, such as the Correlates of War Militarized dispute data set, which is dyadic and annual in its underlying form but comes distributed as three separate files that must be merged together. *EUGene* carries out necessary conversions among the formats, file structures, and differing units of analysis of these data sets as part of the merging process. For users who wish to merge data themselves, *EUGene* can also be used to generate a simple set of case identification data (country codes and years) for various subsets of data.

### Dispute Data Conversion to Dyads

One of the most important data sets for quantitative studies of international relations is the Correlates of War Militarized dispute data set, which is the origin of most dependent variables in recent quantitative international relations studies. The COW project does not distribute these data in a dyadic form. Converting the data to a dyadic format involves checking states for their involvement as originators or joiners (joiners are states who join ongoing disputes, but are not involved initially), identification as initiators or targets (the initiator is the side of the MID that first crosses the militarized threshold, side A in the MID data), and pairing the sides into dyads. Appropriately pairing states as actual dyads rather than simply participants on opposite sides of a multilateral dispute involves checking to be sure that the states are actually involved on opposite sides at the same time (some simple conversions do not perform this check). The procedure to make appropriate pairings involves reading data from 3 files, two of which have the dispute as the unit of analysis (one record is contained per dispute), and one of which has the country-involvement as the unit of analysis (one record is contained per state involved in the dispute, per dispute). Country-dispute-level information in the data set must be matched to the dispute-level data, and multiple country-dispute-level records must be matched to each other to obtain dyadic pairings. *EUGene* does this automatically.<sup>5</sup>

### Dependent Variable Coding

*EUGene* allows users to make critical specifications about how to treat dependent variables created using militarized disputes. Perhaps the most

common dependent variable used in dyadic analyses is the occurrence of a militarized dispute. However, that "occurrence" may be measured in different ways. In a directed-dyad setting, the initiation of a dispute is the appropriate coding of dispute. The COW MID operationalization of initiator in practice is that the initiator is the "first mover," that is, the state who first crosses the MID threshold and makes the first threat or actual use or movement of forces. This gains clarity in terms of the temporal ordering of actions while losing any attempt to get at intent. However, the image of the initiator that first comes to mind when we think about conflict is the predatory state who decides to engage in conflict against a state that wants to remain at peace. This may or may not be the state that moves first. *EUGene* allows users to specify an alternative coding for dispute initiator (which also comes from the MID data set), namely the identity of states as "revisionist" states.

A second issue of dependent variable coding lies in the treatment of disputes that continue for more than one year. Within any single dispute, we must choose how to treat ongoing years. We might separate out ongoing conflict years for separate analysis (and study onset versus continuation), we might treat each subsequent year of an ongoing disputes as an entirely new dispute, or we could drop these observation from our analysis all together. Precise conceptualization of the dependent variable is necessary to determine which of the choices is best for a particular research agenda. We argue (e.g., Bennett and Stam, 2000c) that only the first year of a new MID should be coded as a dispute initiation, and that subsequent years of multi-year disputes should be dropped (such subsequent years are instances of a new dependent variable, namely dispute continuation). Others (Oneal and Russett, 1999) combine continuation and onset and argue that we should code ongoing dispute years as a "1" as well as just the first year. This coding requires our models to do double duty and predict the continuation as well as the initiation of MIDs. Analysis has shown that alternative decisions on this coding leads to different empirical findings (Beck and Tucker, 1997). *EUGene* allows users to either specify a dependent variable with only the first year of a MID as a dispute initiation, or to code all years of a continuing MID as a dispute.

### **Output Format and Use of Other Software**

*EUGene* is not an analysis program, but rather a data management utility. As a result, the merged data created by *EUGene* must be read into and analyzed by other statistical software. *EUGene*'s output files are created in

a uniform format that can be read into any statistical analysis package with ease. The program provides users options to create data that are tab-delimited (tab characters are placed between values in the data file), space delimited, or comma delimited. In addition, as it creates data sets, *EUGene* creates command files to import the data into SPSS, Stata, or LIMDEP. After creating a data set, users can then have the data up and running in the statistics program of their choice in literally a matter of minutes.

### **Data Management Summary**

Overall, by simplifying a number of routine but critical tasks in the creation of data sets for analysis in international relations, and by allowing users to easily make any of a variety of choices critical to the construction of a data set (and in fact by explicitly forcing them to make those choices), *EUGene* allows analysts to proceed more quickly to the stage of testing theoretical arguments. By minimizing the time necessary for the later stages of research, it also allows users to spend more time on the critical stage of *developing* theory as well. The program's flexibility allows it to be used for multiple research agendas, and further expansions will allow it to be used in projects that we cannot anticipate. For instance, one planned addition to the program will allow it to read and incorporate user-supplied data sets (rather than just those we provide), and we plan to make add-on data sets available for download and automatic import into the program. As these data sets become available, the usefulness of the program for a variety of research tasks will increase further.

### **DATA DEVELOPMENT: GENERATING EXPECTED UTILITY AND RISK ATTITUDE DATA**

*EUGene*'s second major goal is to serve as a computational tool for the creation of expected utility data to apply the so-called expected utility theory of war across time and space. Bueno de Mesquita and Lalman's (1992) game-theoretic implementation of an expected utility theory of war is perhaps the best-known application of rational choice theory to the study of international conflict. By building a game which appears representative of many interactions in international relations, and by directly operationalizing key utility and probability concepts, Bueno de Mesquita and Lalman took the game theoretic analysis of international conflict a step beyond theory to rigorous empirical testing. However, the testing of the

theory conducted has been quite limited until recently because only a small set of expected utility data were available. *EUGene* creates the data necessary to better test this key theory.

The critical variables for testing the expected utility theory of war as applied most recently in Bueno de Mesquita and Lalman (1992) are the game-theoretic equilibrium predictions that emerge from the "International Interaction Game" (hereafter, IIG). Making a prediction of the game's equilibrium at any point in time in turn depends on a variety of input data on national capabilities and alliances, which combine to create estimates of states' utility for outcomes. Figure 1 presents the IIG. Bueno de Mesquita and Lalman provide a complete discussion of the structure of the IIG and the choices in it. A very general overview is that the game represents an interaction structure between a potential conflict initiator (state A) and a potential target (state B). The game begins with a choice by state A to issue some type of demand to state B. If A does not make a demand, then B may make a demand; if neither does, then the outcome of the interaction is the status quo. If one state does initiate a crisis by making a demand, then at subsequent decision nodes in the game actors A and B alternately make choices about whether to make further demands (escalating the conflict) or not (resulting in some type of settlement). Each actor makes choices that appear to yield it the greatest expected utility when it looks ahead to the outcome. The key for purposes of discussing the

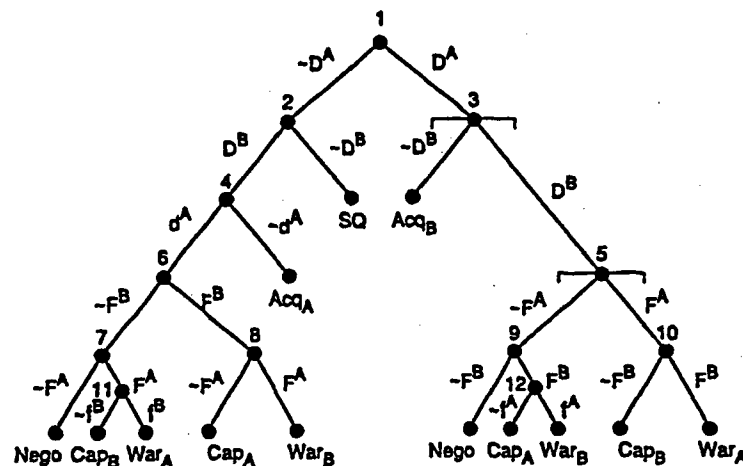


Figure 1 International Interaction Game (Bueno de Mesquita and Lalman, 1992)

empirical application of the game is to note the eight possible game outcomes (eight terminal nodes). Each of these nodes represents a potential final situation in which the members of a dyad might find themselves after they complete their interaction, and each has some analogue in the real world that can be measured. The eight possible outcomes to the game are 1) a status quo outcome (no dispute, designated SQ in the figure); 2) a challenge resolved by negotiation (Nego); 3) a challenge resolved by state A's acquiescence (giving in, designated Acq<sub>A</sub>); 4) a challenge resolved by A's capitulation (an acquiescence to B's demand but under the threat of force, Cap<sub>A</sub>); 5) a challenge resolved by B's acquiescence (Acq<sub>B</sub>); 6) a challenge resolved by B's capitulation (Cap<sub>B</sub>); 7) a war<sup>6</sup> initiated by A (War<sub>A</sub>); or 8) a war initiated by B (War<sub>B</sub>). Each state expects to receive some utility at each terminal node, resulting in a game with 16 relevant utility values.

The states' utilities for these 8 possible outcomes combine in a non-linear manner to yield a prediction of what the game's outcome should be in equilibrium, generated either through backward induction or by focusing on a series of propositions that specify what combinations of utility values for the two players yield each outcome in equilibrium (Bueno de Mesquita and Lalman, 1992, pp. 72-92). Bueno de Mesquita and Lalman develop operationalizations that can yield empirical estimates of states' utilities in any given year, allowing us to estimate the expected equilibrium outcome in any directed dyad-year and compare it to actual events to assess how well the game's prediction correlates with actual behavior.

Unfortunately, it has been impossible until recently to fully test arguments about the fit of the IIG and the related operationalizations from *War and Reason* because the data to predict equilibria have been unavailable. While the IIG is potentially applicable across space and time, Bueno de Mesquita and Lalman only constructed the data necessary to test the theory on a limited subset of 707 dyad-years in Europe (out of over one million possible cases). These data have not been available because computing the expected utility of all states involves several computationally intensive procedures which until recently needed a fast mainframe computer to be performed. Generating the data to test the theory involves measuring the similarities of dyadic alliance portfolios, individual national capabilities, dyadic distances, states' risk propensities, and finally predicting third party choices to join conflicts. Computation of risk propensities in turn involves estimating the hypothetical best and worst security situations in which a state might find itself. This is a time-consuming computation that involves searching over literally billions of possible alliance patterns in order to find the best and worst security scenarios for

each country-year. Until recently, making the required calculations required a fast mainframe computer.

Because of the complexity of operationalizing expected utility, Bueno de Mesquita initially developed a software program for the IBM PC known as *Tolstoy* (Horn, 1990) to generate expected utility data. One of our goals in developing this software was to facilitate the output of data that could be easily transformed into estimates of conflict initiation and escalation probabilities. However, *Tolstoy* could not be used to obtain accurate data for the most recent version of expected utility theory developed in Bueno de Mesquita and Lalman (1992) for a wide set of cases. Largely because of technological limitations existing when it was written, *Tolstoy* implemented only expected utility calculations as presented in Bueno de Mesquita (1985) and was unable to produce estimates of expected utility for all dyads and time periods. This program did not solve the problem that there has simply been no way to obtain these data.

*EUGene* generates expected utility data for any dyad and year for which raw input data is available, and makes the resulting data accessible to researchers. *EUGene* is the first software to generate the data needed to test the IIG following the methods developed in Bueno de Mesquita and Lalman (1992) for all dyads and years, starting from first principles.<sup>7</sup> *EUGene* actually has several intended uses vis-à-vis expected utility data. Most obviously, the availability of complete expected utility and IIG equilibrium predictions allow us to test arguments about the IIG and the expected utility theory of war in a much broader setting. With data on all dyads available, we can also test previously unexamined arguments about expected utility, for example, with data generated by *EUGene*, one could examine differences across regions or look for systematic variation in risk taking propensities. Another purpose is to generate expected utility data that uses the most recent updates to the Correlates of War alliance, national capabilities, and MID data sets ensuring that researchers use the most accurate expected utility estimates available. A final goal is to implement the methods of previous expected utility calculations, which Bueno de Mesquita and others published across a number of books and article, combining the various published calculations into a single, easily accessible package.

*Steps in generating expected utility data.* *EUGene* takes as raw inputs a number of data sets, including national capability data, alliance data, country independence dates, contiguity and geographical location data. The program then computes the COW national capabilities index, Tau-b scores for all dyads, risk attitudes of all states for all regions, and finally states' expected utilities for war, negotiation, the status quo, and the other

possible outcomes of the IIG. In a final step, *EUGene* uses the states' utility scores to predict the outcome expected in equilibrium for each directed-dyad year in the interstate system. Appendix 1 of *War and Reason* contains the detailed formulas for the full operationalizations of the terms needed to estimate utilities and in turn the final IIG equilibrium predictions (with two clarifications made below). The software provides users with some options for modifying these calculations. For instance, users can modify the capability data used as input and regenerate utility predictions for sensitivity analysis, and a future version of the program will allow users to substitute Signorino and Ritter's (1999) "S" score for tau-b as the measure of alliance portfolio similarity used in calculations. The steps carried out by *EUGene* to create final IIG equilibrium predictions are as follows:

- The individual components of national capabilities are assembled into the COW national capabilities index for every country-year (Singer, Bremer and Stuckey, 1972).
- COW alliance data are used to calculate the tau-b score for each directed dyad in each year in the international system. Such scores are directed because the states included in computing tau-b scores depend on the region in which they are expected to interact (Bueno de Mesquita, 1978, 1981, pp. 94–97).
- Geographic location and contiguity data are used to create the distance between each pair of states, with changes over time due to major territorial changes taken into account.<sup>8</sup>
- Distance, tau-b and capability data are combined to create an estimate of the expected utility of war for each directed dyad-year. The initial EU computation follows the methods in *The War Trap*, but does not adjust expected utility for risk attitude using *The War Trap*'s method. That is, the scores generated are the sum of the bilateral and multilateral expected utility components (equation 6, Bueno de Mesquita, 1981, p. 59), but does not take the final step of introducing risk attitude. [These scores are necessary to estimate risk attitude by updated methods in the next step.]
- The initial expected utility values from the prior step are used to produce estimates of states' risk attitudes for all years and with respect to each region in the system following the methods of Bueno de Mesquita (1985). First, alliance data and expected utility data are used to generate actual (realized) national security portfolios for all state-years. Hypothetical alliance data are then used to generate potential maximum and minimum security portfolios for all state-years.



The hypothetical security values are compared to actual security to compute each state's annual risk attitude; the closer a state is to its security-maximizing alliance portfolio, the more risk-averse it is.<sup>9</sup>

- Tau-b scores and risk attitude scores are combined following the methods of *War and Reason* (equations A1.1 to A1.6, Bueno de Mesquita and Lalman, 1992, pp. 293–294) to produce estimates of states' utility for the status quo, their most preferred international political position vis-à-vis the opponent, and their least preferred international political position. These are the estimates of  $U^A(SQ)$ ,  $U^A(\Delta_A)$ , and  $U^A(\Delta_B)$  for each directed-dyad year A-B.
- Tau-b scores, distance data, national capability data and risk attitude scores (for states A, B, and third-party states within the relevant region of conflict) are combined following equations A1.7 to A1.10 of Bueno de Mesquita and Lalman (1992, pp. 294–297) to produce estimates of the probability of success  $P^A$  for a state A in a military conflict against B, taking into account the likely behavior of possible interveners.
- The domestic cost term  $\phi$  is measured (Bueno de Mesquita and Lalman, 1992, p. 297) as  $U^A(SQ)$ . Following the actual method of constructing variables that Bueno de Mesquita and Lalman used (Bueno de Mesquita, personal communication), the other cost terms  $\alpha$ ,  $\tau$ ,  $\gamma$  are set to 1.0. Although these terms are defined theoretically and included as part of the expected utility equations, Bueno de Mesquita and Lalman had no way to measure them empirically. Without this assumption, we could not construct complete measurements.<sup>10</sup>
- The utility measurements  $U^A(SQ)$ ,  $U^A(\Delta_A)$ , and  $U^A(\Delta_B)$ , the probability of success  $P^A$ , and cost terms are combined following Table IIb of Bueno de Mesquita and Lalman (1992, p. 47) to produce estimates of the utility of each of the eight outcomes of the IIG for each directed-dyad year.
- The equilibrium outcome of the IIG played under complete and perfect information conditions is computed for each directed dyad year using the 2 states' utilities for each of the eight IIG outcomes, following the logical conditions in Bueno de Mesquita and Lalman (1992, pp. 72–92).<sup>11</sup> For every directed-dyad year, the 16 final expected utility values corresponding to the utilities of states A and B are used to compute equilibria. Within the set of all dyads, the actual distribution of equilibrium predictions is such that approximately 12% of the cases have a status quo equilibrium prediction; 61% negotiation; 0.07% acquiescence by A; 0.5% acquiescence by B; and 29% a war initiated by A. Within politically relevant dyads, the distribution of equilibria is

status quo 15%; negotiation 65%; acquiescence by A 0.1%; acquiescence by B 1%; war initiated by A 22%.<sup>12</sup>

In addition to final equilibrium predictions, most of the separate components used in computing them are available for output as separate variables from *EUGene*. In particular, tau-b scores, risk scores, and the individual expected utility components may be selected for inclusion in output data sets. The next version of the program will also save information about the best and worst hypothetical alliance portfolios found by our search algorithm. This may allow analysts to examine risk attitude and "off the equilibrium path" security behavior (unrealized alternative alliance arrangements). By making available such data, we hope to generate possible research questions that heretofore could not be examined.

### **Differences from Existing Expected Utility Data**

While we generated significantly more expected utility data than were available previously, we expected our output data and analysis to be potentially rather different from even the same cases in prior work because of updates to a number of elements important to required computations. These differences could lead to different utility values and equilibria, potentially in every single dyad.

- *EUGene* uses COW alliance and national capability data, both of which have been updated since Bueno de Mesquita and Lalman did their calculations.
- We make one modification to the national capabilities data: because data for British energy consumption was often missing between 1816 and 1850, we interpolated data on British energy consumption when it was missing for this period.
- We use the most recent version (v2.1) of the COW MID data to identify disputes and dispute outcomes. The new MID data set contains roughly twice as many disputes as the original MID data set. In addition, outcome codings for many existing disputes are slightly different in the two data sets.
- While *The War Trap* incorporated distance, *War and Reason* did not, since the empirical analysis was purely within Europe, where geographic distances are relatively small. We add distance discounting back to the methods of *War and Reason*, discounting capabilities in each dyad  $ij$  such that a) the capabilities of the challenger  $i$  are

adjusted by the distance to the target  $j$ , and b) the capabilities of third parties  $k$  are adjusted by the shorter of the distance to  $i$  or  $j$ . The states  $k$  are those states that might contribute capabilities to help  $i$  or  $j$ . They include those who are involved in the region of expected conflict for the directed dyad  $ij$ . We followed Bueno de Mesquita's region of conflict definition (1981, p. 97).

- Our method for computing state-to-state distance is slightly different from that used in *The War Trap*. There, Bueno de Mesquita calculated actual port-to-port sea distances and capital-to-capital air distances from various reference sources. In our program, we compute distance by taking the latitude and longitude of international cities and applying the "great circle" navigation distance formula to compute distance (Fitzpatrick and Modlin, 1986). This method has the advantage of consistency and avoids a variety of issues in replicating shortest sea route distances given uncertain air and sea routes. In addition, countries that are contiguous on land we considered 0 miles apart. We use the 1993 version of the COW contiguity data set for this computation.
- The most important change and improvement in our computational methods lie in our computation of risk attitude scores. Bueno de Mesquita (1985) discusses using hypothetical maximum and minimum-security portfolios to compute risk attitude, but does not lay out the underlying computations that go into producing those hypothetical scores. Risk attitude is estimated by examining where a state's actual security arrangement falls between possible alliance arrangements that would maximize or minimize its security. Hypothetical security scores are estimated by finding the combination of alliances for a state that maximize or minimize the function  $\sum_{i < j} E(U_{ji})$ , where  $E(U_{ji})$  is the expected utility of  $j$  for challenging  $i$  as defined in *The War Trap*. That utility is a function of capabilities (which are fixed) and tau-b scores, which change as the alliances of state  $i$  change. Each of  $i$ 's potential alliance configurations requires the recalculation of each utility ( $U_{ji}$ ) and then a re-summing. While it does not take long to re-estimate security scores for any particular alliance configuration, the number of hypothetical alliance configurations (constituting the search space for the risk-attitude calculation problem) is huge. The search space consists of all possible alliance patterns between state  $i$  and the other states involved in  $i$ 's region. "Involved" states are those states that are members of the region, plus those states with major foreign policy commitments in the region as identified by Bueno de Mesquita, 1981, pp. 95-97). In a

region of 25 countries, for instance, which is typical of Europe, the search space is then  $4^{24}$  potential alliance patterns for state  $i$ . This is because  $i$  could have any of 4 alliance types (defense, neutrality, entente or no alliance) with the 24 other states involved in the region. It is impossible to exhaustively search this space for one country ( $4^{24}$  is approximately  $2.8 \times 10^{14}$ , or 300 trillion combinations), let alone 11,000 country-years (which is the scope of the international system from 1816 to 1993). Bueno de Mesquita solved this problem by sampling 30,000 configurations per year from this space. With different technology and algorithms available, we optimize using a genetic algorithm technique (Goldberg, 1989; Holland, 1975)<sup>13</sup> which performed a more systematic search of between 1000 and 2000 configurations per state per region per year.<sup>14</sup> Given that we have in most cases identified different (and more accurate) best and worst security configurations, our risk scores differ from those reported in Bueno de Mesquita (1985).

As Bueno de Mesquita and Lalman are careful to point out in their own work (1992, p. 299), of course, these operationalizations are still likely to have measurement error in them that will make the empirical fit between the game equilibrium and actual outcomes imperfect at best.

### Software Verification of Expected Utility Equilibria

Even though we expect substantial differences between *EUGene*'s output and the earlier data, we performed a number of checks to verify that *EUGene* was producing proper expected utility scores.<sup>15</sup> First we correlated expected utility scores constructed following the methods of *The War Trap* with the data printed in Tables A through C of the volume's Appendix. Regression of the new data on the reported expected utilities in cases of wars, interventions and threats gives an  $r$  of 0.78 ( $n = 241$  dyads). This correlation could have been much lower because the data in the appendix of *The War Trap* is adjusted for risk attitude, while the raw utility scores we use at this stage do not yet have risk attitude factored in. We next correlated our new risk scores with those generated through 1970 in Bueno de Mesquita (1985). The number of state-years and the correlation between *EUGene*'s risk measure and the old data are as follows: Europe,  $n = 3493$ ,  $r = 0.73$ ; Asia,  $n = 1437$ ,  $r = 0.51$ ; Middle East,  $n = 1765$ ,  $r = .73$ ; Americas,  $n = 2740$ ,  $r = 0.46$  (the 1985 data for the Americas region contains several illegal risk values greater than 1 as

output by Bueno de Mesquita's *Tolstoy* software [Horn, 1990]). These values are all highly significant; differences here are attributable to the much larger search space we examine in finding hypothetical alliance patterns in our generation of risk scores. Finally, we correlated the new estimates of the initiator's utility for the status quo and acquiescence with those utilities in the *War and Reason* data set. The Pearson's  $r$  between *EUGene*'s data and the *War and Reason* data (707 dyads) is: 0.72 for the initiator's utility for the status quo  $U^A(\text{SQ})$ , 0.76 for the initiator's utility for the status quo  $U^B(\text{SQ})$ , 0.73 for the initiator's utility for victory  $U^A(\Delta_A)$ , 0.69 for the target's utility for victory  $U^B(\Delta_B)$ , 0.91 for the initiator's utility for its own acquiescence  $U^A(\Delta_B)$  and 0.93 for the target's utility for its own acquiescence  $U^B(\Delta_A)$ .

## TECHNICAL INFORMATION

In terms of its appearance and use, *EUGene* runs under Microsoft Windows 95 (or higher) or Windows NT (version 4.0 or higher) on IBM-compatible PCs with at least 16 MB of memory, and has a standard Windows program interface. Users have two main choices when they enter the program. First, users can choose to construct an output data set containing any or all of the variables discussed previously. A tabbed window provides the option for users to set the unit of analysis, population of cases, variables, case exclusions (for joiners or ongoing disputes, for example) relating to their output data. The data emerge as a flat text (ASCII) file that can be read into spreadsheet or statistical analysis programs. Second, they can re-compute expected utility data (or various components of it) after specifying a set of conditions, such as what distance discounting method to use. Normally users will not have to do this, as we have already pre-calculated data following Bueno de Mesquita's methods (the re-calculation options exist for users who wish to create new data sets to examine the sensitivity of the results to various assumptions).

*EUGene* is written in Borland, Inc.'s *Delphi* programming environment, an object-oriented Rapid Application Development package designed to create Windows programs, relying on a Pascal base. The program currently consists of approximately 14,000 lines of computer code split into 40 units and 25 Windows forms. The program is copyrighted, but is available as a free download from <http://www.eugenesoftware.org>. The downloadable program includes the main program executable file (about 950 K), complete (60 page) documentation of the program, *EUGene*'s source code, and the complete expected utility data described in this paper (approximately

50 MB of data). In addition to providing instructions for how to use the program, the program documentation further details the computations involved in making expected utility calculations and outlines key algorithms used in the program. We welcome suggestions for corrections and improvements as part of ongoing program development.

## CONCLUSION

By creating a user-friendly tool to assist in creating data sets of fundamental importance to quantitative IR, we intend to provide a valuable resource to help the community of scholars interested in understanding international politics. The availability of complete data on expected utility running from 1816–1984 also enables broad testing of one of the most important theories of international politics, testing pursued in Bennett and Stam (2000a, 2000b, 2000c). *EUGene* will not solve all data set construction problems, of course. In particular, the fundamental temporal level of aggregation in *EUGene* is the year. *EUGene* implicitly directs researchers towards event history types of analysis. While this does tilt us away from some types of data aggregations, it makes simple the construction of data needed to test dynamic multiple state models to take into account possible selection bias (Reed, 2000). If analysts require output or have input data at a finer level, *EUGene* is not the most appropriate tool to use.<sup>16</sup>

Others may suggest that by making available tools that make quantitative data easily accessible, we are bypassing a filtering process that ensures that only skilled users will employ these data. This may open the door for poor analysis, since users may create data sets that they do not fully understand, or where they are not forced to gain an appreciation for the nuances of the data they use. For example, scholars who have wrestled first-hand with converting and using the Correlates of War MID data have had to become familiar with the critical issues involved in using it dyadically. We believe that this danger is present even without the availability of tools like *EUGene*, given the ease of downloading data sets for replication and the fact that when using data sets such as the MID data, even experienced users may miss some nuances. We believe that by providing extensive documentation, online help, and by confronting users with choices that they *must* make in the course of constructing a data set (for example choosing how to code initiators and joiners), we are actually making it easier for users to make informed choices about the data sets they construct. In addition, we believe that *EUGene* can be a useful tool

even for scholars well-versed in the data sets we employ, again simply by reducing the time it takes to construct data sets.

Further development of *EUGene* is continuing. Future versions of the program will include non-directed dyads as an output option, additional variables, automated addition of user-supplied data sets, improved methods of tracking versions of input data, a mechanism to track the user selections and specifications made when creating particular output data sets, and internal calculation of new variables that derive from variables already in the program. As new data become available (e.g., new COW alliance data), we plan to update our available data sets, and to recompute key variables (i.e. expected utility variables) before releasing new versions. We also welcome suggestions about future improvements in the program. With continued feedback and improvements, we hope that scholars will be able to pursue the quantitative study of international relations even more vigorously than they have to date.

## ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation under grants SBR-9601151, SES-9975115, and SBR-9975291. We would like to thank Bruce Bueno de Mesquita for his invaluable assistance in generating expected utility data.

## NOTES

1. Bueno de Mesquita's tests in War and Reason were limited to 707 observations with a limited number of control variables. The software described below generates over 1,000,000 observations on over 60 variables of interest to international relations scholars and students.
2. A dyad-year is a pair of states in a given year, e.g. US vs. USSR 1984. A directed-dyad-year data set separates each dyad into two directions which can designate action directed by one side *against* another. Such a data set would include 1) US vs. USSR 1984, and 2) USSR vs. US 1984 for analysis.
3. If scholars wish to use data sets with a sub-annual level of temporal aggregation, e.g. the quarter, *EUGene* is not an appropriate tool. But if scholars wish to create data sets with a longer aggregation time period (e.g., the decade), *EUGene* will create an annual data set which can be read into other statistical programs and then aggregated by the user (for example, by using the "aggregate" command in SPSS).
4. Version 2.0, which is in development, will add non-directed dyad-years to this set.
5. Zeev Maoz has developed a dyadic version of the MID data (Maoz, 1999) that is designed to correct problems involved in converting the original MID data in dyadic format. While

Maoz begins by converting the original MID data sets into a dyadic format, he then conducts a variety of additional checks intended to verify that actual disputations interaction occurred between the members of the dyad. Maoz's data will be incorporated in a future version of *EUGene* as an option.

6. Technically, the game predicts up to the mutual use of force. Bueno de Mesquita and Lalman label this outcome "war," but this does not imply that other conditions for war (such as sustained fighting and a high number of casualties) are met. Theoretically, there is no separation between the mutual use of force and a larger war in the game.
7. As *EUGene* was developed, we communicated frequently with Bueno de Mesquita who provided us with feedback, clarifications, and specified a number of details of auxiliary assumptions that were made during the development of the data used in *War and Reason*. The complete source code and details of our software implementation are available for download and inspection with the program.
8. *War and Reason* did not incorporate distance discounting into its calculations because the empirical analysis was solely within Europe, where all geographic distances are relatively small. Since we are analyzing dyads worldwide, we reintroduce the distance discounting methods of *The War Trap*. We generalize and make uniform the method for computing state-to-state distance by taking the latitude and longitude of international cities and applying the "great circle" navigation distance formula to compute distance (Fitzpatrick and Modlin, 1996). For most dyads, we used the national capitals as the ends of the curve to compute distance. However, in the case of the US and USSR, we used multiple cities as described in *The War Trap*. In addition, we considered countries that are contiguous on land to be 0 miles apart. We used the 1993 version of the COW contiguity data set for this computation. We discount capabilities in each dyad  $ij$  so that 1) the capabilities of the challenger  $i$  are adjusted by the distance to the target  $j$ , and 2) the capabilities of potentially involved third parties  $k$  are adjusted by the shorter of the distance to  $i$  or  $j$ .
9. Identifying hypothetical alliance patterns that minimize and maximize security involves a search over literally millions of possible alliances per state per year. In a region of 25 states with any of 4 alliance types possible with each other state, any one state has the possibility of allying in  $4^{24}$  overall alliance patterns. It is impossible to exhaustively evaluate this space. While Bueno de Mesquita sampled ~30,000 configurations per year, *EUGene* optimizes its search by implementing a genetic algorithm (Holland, 1975; Goldberg, 1989) and examining roughly 100 times more combinations than did Bueno de Mesquita (1985). Even with these techniques, it took approximately 12 computer-months on 200MHz Pentium-Pro PCs to produce the risk scores used here. Users can take advantage of the data we generated without having to repeat this procedure.
10. An additional implication of setting  $\alpha = \tau$  (and so not measuring the magnitude of a first strike advantage) is that the utility of  $War_A$  and  $War_B$  is always equal. The fact that  $\alpha$ ,  $\tau$ ,  $\gamma$  are set to 1.0 allows us to make unique equilibrium predictions for each dyad (with the exception mentioned in note 11), enabling us to avoid most of the problems discussed by Signorino (1999) of dealing with unknown cost parameters.
11. Because the utility of  $War_A$  and  $War_B$  is always equal (because  $\alpha = \tau$ , above), there remains one indeterminate situation in solving the game. Bueno de Mesquita and Lalman's Domestic Proposition 3.1, the "Basic War Theorem," specifies that logically a war will be started by state A if and only if the conditions  $U^A(War_A) > U^A(Acq_A)$ ,  $U^A(Cap_A) > U^A(War_B)$ ,  $U^B(Cap_A) > U^B(Nego)$ , and  $U^B(War_A) > U^B(Acq_B)$  hold. However, with  $U(War_A) = U(War_B)$ , the first two conditions are incompatible, as they together imply  $U^A(Cap_A) > U^A(War) > U^A(Acq_A)$ , which violates the basic theoretical condition of the



game that  $U^A(\text{Acq}_A) > U^A(\text{Cap}_A)$ . Because of this, we omit the condition  $U^A(\text{Cap}_A) > U^A(\text{War}_B)$  in order to generate the war equilibrium (see Bueno de Mesquita and Lalman, 1992, p. 76n). However, with this condition omitted, it is possible for a directed dyad year to satisfy the conditions for both the WarA and Status Quo equilibria. In analyses we have undertaken elsewhere (Bennett and Stam, 2000b), the results do not change significantly if we treat these cases as having a status quo equilibrium prediction instead of a WarA equilibrium.

12. It is clear from these results and the actual frequency of conflictual interactions in the international system that the status quo is under predicted by the IIG and associated operationalizations, while negotiation and war are over-predicted. With this clue, future work could focus on the logical conditions and associated operationalizations of these equilibria in particular. For instance, it may be that measures of the utility of the status quo relative to other outcomes are underestimated in the operationalizations used by Bueno de Mesquita and Lalman, or that measures of the possible gains from conflict are overestimated by differences in foreign policy portfolios (tau-b scores), despite discounting capabilities for distance. It could also be that the assumption that the war cost terms  $\alpha$ ,  $\tau$ ,  $\gamma$  equal 1 (made for empirical estimation purposes) is problematic, and instead the costs of war loom larger than the model assumes.
13. Genetic algorithms rely on analogies and insights from genetics and evolution to conduct effective search over very large spaces. The algorithm works by creating a "population" of individuals who represent particular points in the search space. For instance, one "individual" is an alliance pattern with the US, Britain, and France in a defense pact and all other states with no alliance; a different individual has the US and USSR in a neutrality pact while US, Britain and France have a defense pact and all others have entente agreements, etc. Each individual can be represented by a "chromosome" represented by a string of bits indicating the alliance membership of each state in the system. Each individual (alliance pattern) in this population is assessed for "fitness," which in this case is a measure of the security provided to a state within that pattern. Through an iterated procedure, this population of individuals (points in the search space) go through successive generations, in which poor fitness (low security) alliance patterns are discarded, while high fitness patterns are kept. Then, new points in the search space are added to keep the population size the same by recombining parts of multiple good alliance patterns into new patterns. For instance, a new alliance working from the above population (assume both of the above patterns are relatively good at providing security) might maintain the US, Britain, and France in a defense pact, but combine that part of the alliance pattern with other states having ententes with one another rather than no alliance. By taking alliance patterns that are relatively good (provide more security) and combining and modifying elements of those patterns, the search proceeds by improving on positive performance. The genetic analogy comes in through the idea of a set of search points in a population, and by the way in which different alliance patterns combine to move the search forward, much as a new individual takes some of its parentage from the mother and some from the father.
14. The data distributed with the program were generated using population sizes between 20 and 30, mutation probability of 0.05 per state per alliance configuration per iteration, and was assumed to stabilize after 8 generations without a change in the maximum (or minimum) security value found. A single crossover point was used to produce successive generations, and the best two configurations found in each generation were "cloned" and included in the next generation. Variations on these parameters, in particular increases in population size, only occasionally improved the optima identified. After this best point

was identified by the genetic algorithm, a random walk was followed to ensure that a true optimum was reported (i.e. no adjacent configurations were better). Thus our optimization method was in the end a hybrid optimization method in which the genetic algorithm found the largest hills, and in the last iteration, a random search was used to follow that hill to its peak.

15. Our communications with Bueno de Mesquita while developing *EUGene* clarified key elements of the operationalization, such as that  $\alpha$ ,  $\tau$ , and  $\gamma$  in the expected utility equations were set equal to 1 because they cannot be estimated. Any programming errors, of course, remain ours.
16. *EUGene* is also not, for example, the best tool to analyze data formatted as just a list-of-events formats (such as a list of wars or crises). Such data on events must be transformed to fit into a data set including non-events, which *EUGene* does by focusing on the dyad-year or country-year as the unit of analysis. Nor is it optimal for managing data such as relational data bases such as Sherfacs, although if such data are transformed to a dyadic format they could easily be merged into *EUGene*.

## REFERENCES

- Beck, Nathaniel and Richard Tucker (1997). "Conflict in Space and Time." Weatherhead Center for International Affairs Working Paper 97-8, Harvard University, Cambridge, MA.
- Beck, Nathaniel, Jonathan Katz and Richard Tucker (1997). "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science*, Vol. 42 (October), pp. 1260-1288.
- Bennett, D. Scott and Allan C. Stam (2000c). "Research Design and Estimator Choices in the Analysis of Interstate Dyads: When Decisions Matter." *Journal of Conflict Resolution*. (forthcoming)
- Bennett, D. Scott and Allan C. Stam (2000a). "A Cross-Validation of Bueno de Mesquita and Lalman's International Interaction Game." *British Journal of Political Science* (forthcoming)
- Bennett, D. Scott and Allan C. Stam (2000b). "A Universal Test of an Expected Utility Theory of War." *International Studies Quarterly*, Vol. 44, pp. 451-480.
- Bremer, Stuart A. (1992). "Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War, 1816-1965," *Journal of Conflict Resolution*, Vol. 36, pp. 309-341.
- Bueno de Mesquita, Bruce and David Lalman (1992). *War and Reason: Domestic and International Imperatives*. New Haven: Yale University Press.
- Bueno de Mesquita, Bruce (1975). "Measuring Systemic Polarity." *Journal of Conflict Resolution*, Vol. 19, pp. 187-216.
- (1978). "Systemic Polarization and the Occurrence and Duration of War." *Journal of Conflict Resolution*, Vol. 22, pp. 241-267.
- (1981). *The War Trap*. New Haven: Yale University Press.
- (1985). "The War Trap Revisited." *American Political Science Review*, Vol. 7, pp. 156-177.
- Fitzpatrick, Gary L. and Marilyn J. Modlin (1996). *Direct-Line Distances (United States Edition)*. Metuchen, NJ: The Scarecrow Press, Inc.
- Goldberg, David E. (1989). *Genetic Algorithms in Search, Optimization, & Machine Learning*. New York: Addison-Wesley Publishing Company, Inc.

- Holland, John H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.
- Horn, Michael (1990). *Tolstoy*. Computer program. University of Rochester.
- Huth, Paul K., Scott Bennett D. and Christopher Gelpi (1992). "System Uncertainty, Risk Propensity, and International Conflict," *Journal of Conflict Resolution*, Vol. 36, pp. 478–517.
- Huth, Paul, Christopher Gelpi and Scott Bennett, D. (1993). "The Escalation of Great Power Militarized Disputes: Testing Rational Theory and Structural Realism." *American Political Science Review*, Vol. 87, pp. 609–623.
- Jagers, Keith and Ted Robert Gurr (1995). "Tracking Democracy's Third Wave with the Polity III Data." *Journal of Peace Research*, Vol. 32 (November), pp. 469–482.
- Jones, Daniel M., Stuart A. Bremer and J. David Singer (1996). "Militarized Interstate Disputes, 1816–1992: Rationale, Coding Rules and Empirical Patterns." *Conflict Management and Peace Science*, Vol. 15, pp. 162–213.
- Lemke, Douglas and William Reed (1998). "Power is not Satisfaction." *Journal of Conflict Resolution*, Vol. 42, pp. 511–516.
- Maoz, Zeev and Bruce Russett, M. (1993). "Normative and Structural Causes of Democratic Peace, 1946–1986," *American Political Science Review*, Vol. 87, pp. 624–638.
- Maoz, Zeev (1999). Dyadic MID data set, version 1.0. <http://spirit.tau.ac.il/~zeevmaoz/>.
- Oneal, R., John and Bruce Russett (1999). "Assessing the Liberal Peace Debate with Alternative Specifications: Trade Still Reduces Conflict." *Journal of Peace Research*.
- Oneal, R., John and Bruce Russett, M. (1997). "The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950–1985." *International Studies Quarterly*, Vol. 41 (June), pp. 267–294.
- Ray, James Lee and Yijia Wang (1998). "Integrating Levels of Analysis in World Politics: Increased Utility or Exercises in Futility?" Paper presented at the Peace Science Society (International) annual meeting, Rutgers University, October 16–18.
- Reed, William (2000). "A Unified Statistical Model of Conflict Onset and Escalation." *American Journal of Political Science*, Vol. 44 (January), pp. 84–93.
- Russett, Bruce M., David Singer, J. and Melvin Small (1968). "A Standardized List of National Political Entities in the Twentieth Century." *American Political Science Review*, Vol. 62, No. 2 (September), pp. 932–51.
- Signorino, Curtis S. and Jeffrey Ritter, M. (1999). "Tau-b or Not Tau-b: Measuring the Similarity of Foreign Policy Positions," *International Studies Quarterly*, Vol. 43 (March), pp. 115–144.
- Singer, J. David, Stuart Bremer and John Stuckey (1972). "Capability Distribution, Uncertainty, and War, 1820–1965." In Bruce Russett, (ed.), *Peace, War and Numbers*. Beverly Hills: Sage.
- Small, Melvin and David Singer, J. (1969). "Formal Alliances, 1815–1965: An Extension of the Basic Data." *Journal of Peace Research*, Vol. 6, pp. 257–282.
- Small, Melvin and David Singer, J. (1982). *Resort to Arms: International and Civil Wars, 1816–1980*. Beverly Hills: Sage Publications.