

EUGene

Expected Utility Generation and Data Management Program

V 3.2

Documentation

September 25, 2007

D. Scott Bennett and Allan C. Stam

D. Scott Bennett
Department of Political Science
The Pennsylvania State University
107 Burrowes Building
University Park, PA 16802-6200
sbennett@psu.edu

Allan C. Stam
Department of Political Science
University of Michigan
stam@umich.edu

<http://www.eugenesoftware.org>

EUGene Copyright 1997-2007 D. Scott Bennett Jr. and Allan C. Stam III

All Rights Reserved

EUGene was developed during work supported by the National Science Foundation under Grants SBR-9601151, SES-9975115, SBR-9975291, and SES-0079120, variously to Allan Stam and Scott Bennett. Significant programming for versions 2.40+ performed by Chris Baker, and 3.1+ by James Lombardi. Additional programming performed by Matthew Rupert and Killian Seper.

Table of Contents

OVERVIEW	1
CONTACTING THE AUTHORS	1
CITATION.....	2
PROGRAM SPECIFICATIONS	2
INSTALLATION PROCEDURE	3
INSTALLATION FROM CD.....	3
INSTALLATION FROM DOWNLOAD.....	3
AFTER INSTALLATION:	3
CREATING SHORTCUTS.....	4
UNINSTALLING EUGENE.....	4
RUNNING EUGENE	4
STARTUP	4
MENU OPTIONS	4
File Menu	4
Browser Menu	5
ReCompute Menu	5
Recomputation Options	6
Create Data Set Menu	7
Unit of Analysis.....	7
Output Choices	8
Tab 1: Files / Format.....	9
Tab 2: Population of Cases.....	9
Tab 3: Sampling (directed and nondirected dyad year only).....	9
Tab 4: Variables	10
Tab 5: Case/Conflict Exclusions	10
Missing Values	11
Create Hypothetical Alliance Dataset	11
Options	12
User Data Menu	13
Prepare Data Set for Submission	13
Transfer User Data Sets to/from Website	13
Trace Menu	13
Help Menu	13
WHILE EUGENE IS RUNNING	14
EXITING EUGENE.....	14
READING DATA INTO OTHER SOFTWARE PROGRAMS	14
VARIABLE CALCULATIONS AND FORMULAS	15
NATIONAL CAPABILITIES / PERCENT SYSTEM CAPABILITIES	15
TAU-B CALCULATION.....	15
S CALCULATION	16
RELEVANT REGION AND REGIONAL IDENTIFICATION	16
UNCERTAINTY	17
DISTANCE.....	17
EXPECTED UTILITY CALCULATION (THE WAR TRAP)	18
RISK ATTITUDE	18

UTILITY (WAR AND REASON).....	20
EQUILIBRIA (WAR AND REASON)	21
SYSTEM CONCENTRATION AND SYSTEM MOVEMENT.....	23
MODIFYING ASSUMPTIONS USED IN VARIABLE CALCULATIONS	23
<i>Modifications to COW Capabilities</i>	23
<i>Distance</i>	23
<i>Risk Attitude Optimization Method</i>	24
<i>Risk Data Source</i>	26
SOFTWARE VERIFICATION.....	27
<i>Differences from Existing Expected Utility Data</i>	27
<i>Software Checks</i>	28
DATA SET CREATION: ADDITIONAL SPECIFICATIONS AND DETAILS.....	30
MISSING VALUES	30
POPULATION OF CASES.....	30
<i>Years to Include in Output</i>	30
All Years.....	30
Specified Years.....	30
<i>Country-Year Selection</i>	31
All States	31
All Major Powers	31
All States within Specified Regions	31
Specific Set of Countries	31
<i>Directed Dyad-Year Selection</i>	31
All Dyads from All Countries.....	31
All Major Power vs. Major Power Dyads.....	31
All Major Power vs. Any State Dyads.....	31
All Contiguous Dyads	31
All Politically Relevant Dyads	31
All Dyads within Selected Regions	32
All Dyads within Maximum Distance	32
Specific Set of Dyads	32
Dyads Read from User File	32
<i>Non-directed Dyad-Year Selection</i>	32
All Dyads from All Countries.....	32
All Major Power vs. Major Power Dyads.....	32
All Major Power vs. Any State Dyads.....	32
All Contiguous Dyads	33
All Politically Relevant Dyads	33
All Dyads within Specified Regions.....	33
All Dyads within Maximum Distance	33
Specific Set of Dyads	33
Dyads Read from User File	33
<i>Directed Dispute Data Selection</i>	34
One Case per Directed Dispute Dyad Initiation	34
One Case per Directed Dispute Dyad Year	34
<i>Non-Directed Dispute Data Selection</i>	35
One Case per Dispute Dyad Onset	35
One Case per Dispute Dyad Year.....	35
VARIABLES.....	35
<i>Variables Available</i>	35
“General” Variable Tab.....	35
“Polity III” Variable Tab	37
“Alliance” Variable Tab.....	37
“Expected Utility” Variable Tab	38
“Conflict Data” Variable Tab.....	40
COW Dyadic MID Data.....	40
Maoz Dyadic MID Data.....	40

MID Variables Always Reported:.....	40
Other Available MID Variables:	41
Mark Joiners.....	41
ICB Crisis Data	42
Peace Years.....	42
Werner Peace Years Option.....	42
Peace Days	43
Initiator / Multiple MID Settings	43
Initiator Coding: Timing:.....	43
Initiator Coding: Identity (Side A vs. revisionist states):.....	44
Multiple MIDs in Year and the “Key MID”:.....	45
“User Data” Variable Tab.....	46
<i>Variable Names and Order in Output File</i>	46
POLITY III MERGING, COUNTRY CODE RECODING, AND NOTES.....	62
CREATING DYADIC MIDS AND MESHING MID DATA SETS.....	63
<i>Variations on the MID data sets</i>	63
<i>Meshing COW MID Data, Maoz’s Dyadic MID Data, and MID 3.0 data</i>	64
Adjusting “Highest Action” level from Maoz dyadic data	64
<i>Converting COW MID Data to Directed Dyads - pre 1993 data</i>	65
<i>Creating dyadic MIDs – 1993+ data (MID 3.0)</i>	65
CREATING DYADIC MID VARIABLES	65
<i>Variables adjusted for dyadic interactions</i>	66
<i>Unadjusted variables (from overall MID/state information)</i>	66
<i>Other notes</i>	67
EXCLUDING OR INCLUDING PROBLEMATIC CASES	67
<i>Ongoing Dispute Year Options</i>	67
Include All Dyads with an Ongoing MID	67
Drop All Dyads with an Ongoing MID	67
Include Ongoing Dispute Dyad Year iff New Dispute	67
Treating Ongoing Dispute Years as Initiations.....	68
<i>Target vs. Initiator Dyads</i>	68
Drop Target vs. Initiator Directed Dyads if no new MID.....	69
Keep Target vs. Initiator Directed Dyads if no new MID	69
<i>Joiners</i>	69
Drop all Joiner Dyads	69
Include all Joiner Dyads	69
Joiner Variables	70
Apparent Anomalies in Joiner Codings	70
COMBINING INCLUDE/EXCLUDE SPECIFICATIONS	71
<i>Dyad-Year Output</i>	72
Basic Output	72
Keep Targets.....	72
Drop Joiners	73
Coding Joiners As Initiators	73
Coding Joiners As Initiators, Dropping Joiners	74
<i>Dispute Dyad Output</i>	74
Basic output	74
Include Joiners.....	74
Treat Joiners as Initiators.....	74
PROGRAM FILES	75
LOG FILE	75
INPUT AND CONFIGURATION FILES.....	75
INTERMEDIATE FILES	80
KNOWN BUGS AND PROBLEMS	81
INTERNAL DETAILS FOR PROGRAMMERS.....	81

LEGAL NOTICE	83
COPYRIGHT	83
CONDITIONS OF USE	84
PROGRAM EXTENSIONS AND MODIFICATIONS	84
DISCLAIMER OF WARRANTY	84
APPENDIX A MODIFIED VALUES IN MODIFIED CAPABILITIES DATA FILE	85
APPENDIX B DEFAULT SPECIFICATIONS FOR PRE-CALCULATED DATA	87
APPENDIX C DATA SOURCES	88
BIBLIOGRAPHY.....	90

Overview

The Expected Utility Generation and Data Management Program (*EUGene*) is designed primarily to generate values for variables pertaining to the so-called Expected Utility Theory of War developed by Bruce Bueno de Mesquita and colleagues (Bueno de Mesquita, 1981, 1985; Bueno de Mesquita and Lalman, 1992). In addition, *EUGene* serves as a data management tool for creating data sets for use in international relations with the country-year, directed-dyad-year, and directed-dispute-dyad-year as the unit of analysis. The dyadic data sets contain information on Militarized Interstate Disputes converted into a directed dyadic format, and include information on a variety of independent variables including expected utility information, tau-b scores, risk attitude values, national capabilities, and distances between states. Data sets are saved in a text format that can be easily read into other programs for statistical analysis.

EUGene is designed to generate expected utility data for all dyads and years. The testing of expected utility theory in Bueno de Mesquita and Lalman (1992) was limited to Europe, primarily because the calculations involved in computing expected utility are complex and time consuming, and a larger data set could not be efficiently generated. *EUGene* is designed to remedy that problem. Earlier software made available to generate expected utility data (the *Tolstoy* program) had some problems and limitations in its design which *EUGene* corrects. *EUGene* calculates expected utility values, but also provides users with options for modifying expected utility calculations and outputting both expected utility and other data for a variety of case subsets and formats. *EUGene* will also predict the dispute outcome expected (game equilibrium) given the International Interaction game developed in Bueno de Mesquita and Lalman (1992), which forms the basis for the game-theoretic version of what has become known as the "Expected Utility Theory of War."

EUGene also makes easier a number of cumbersome tasks associated with building data sets in international relations, especially data sets created with the directed dyad-year as the unit of analysis. We use data from a large number of original data sets in quantitative studies of international relations. Some of those data sets have a unit of analysis of the country-year, such as the Correlates of War national capability data set, or the Gurr Polity data sets. Other data we need to use has the dyad as the unit of analysis, such as data about the physical distance between states, or the Correlates of War contiguity data set. Still other data comes in a hybrid form or with multiple data set structures, such as the Correlates of War militarized interstate dispute data set, which comes as three files, one containing country-dispute level records, and two containing dispute-level records. *EUGene* reads the data from several of the most important other data sets in international relations, merges the data, and will output that data in a uniform format with the directed-dyad-year as the unit of analysis. During this process *EUGene* will carry out necessary conversions between the formats, file structures, and differing units of analysis of these data sets. Because *EUGene* outputs directed-dyad-year data, data sets with different units of analysis ranging from the country-year to the system-year can be accommodated. With those data sets where the unit of analysis is the country-year, *EUGene* also allows merged data to be output with the country-year as the unit of analysis. *EUGene* also allows users to specify subsets of countries and years for output. The set of options provided with *EUGene*, we believe, will significantly simplify the task of building data sets containing information from multiple inputs, allowing analysts to spend less time merging data and more time performing analysis.

EUGene has been used for analysis presented in Bennett and Stam (1997a, 1997b, 1998a, 1998b, 2000b, 2000c, 2000d) and was developed to solve a number of problems that became apparent during the research for Bennett and Stam (1995). Bennett and Stam (2000a) is *EUGene*'s publication of record, containing theoretical discussions of the program's purpose and options.

Contacting the Authors

EUGene's authors, Scott Bennett and Allan Stam, are interested in receiving bug reports, suggestions, and any other feedback about the program. We plan to make program updates available as we make additions and improvements to the software. Please use email to contact us at sbennett@psu.edu or stam@umich.edu.

If you wish to report a bug, please attempt to document as EXACTLY as you can what you were doing when an error occurred. In case of a run-time error, you should record the exact text of the error message that *EUGene* or Windows provided, a description of what you were doing (what menu selections you had made, what options were specified for the current run), and whether or not you can replicate the error. If you suspect an error in the output data or other routines, the more information you provide, the easier it will be for us to examine. If you believe that *EUGene* is dropping or including cases incorrectly, or is coding a dispute variable such as initiation incorrectly, please be certain that you have read the sections “Excluding or Including Problematic Cases” on page 67 and “Combining Include/Exclude Specifications” on page 71. The more information you can provide us in case of errors, the more likely it is that we can quickly locate and correct the source of the problem.

Citation

If you use *EUGene* to generate data subsequently used in a published analysis, we ask that you cite *EUGene*'s publication of record:

Bennett, D. Scott, and Allan Stam. 2000. “*EUGene: A Conceptual Manual.*” *International Interactions* 26:179-204.

EUGene makes use of raw data originally collected by many other scholars. In addition to citing *EUGene*, we ask that you cite the original data sources for your variables as well. If you generate command files to load your created data sets into programs like Stata or SPSS, citations for the various data sets where your variables come from will be included in the command file. In addition, a list of many of these data sources is contained in the section of data sources in an [appendix](#) to this documentation.

Program Specifications

EUGene was written using the Borland Delphi language (v1.0 through 7.0). *EUGene* has been tested on a variety of PC processors starting with the 486 chip, and requires at least 16 MB of memory. More memory will speed up program execution. *EUGene* runs under Microsoft Windows 95 (or higher), NT (version 4.0 or higher), ME, or XP. Any of these systems should perform acceptably when used to output data previously calculated by *EUGene*. However, new calculations are best performed on a fast PC. In particular, the recalculation of risk scores is not recommended except on the fastest systems, as their generation takes months even on a 200 MHz Pentium Pro running Windows NT (which was current during program development). To perform a full installation of the program, you will need approximately 150MB of free disk space; once installation is completed, the final program with all data files will occupy about 90 MB. If you do not plan to use the expected utility data or print equilibrium predictions, you can save space by deleting the largest expected utility data (“EUWarReaTau.dat”; this expected utility data takes up 30 megabytes). If you delete this file and then try to output expected utility, however, the program will crash.

On a 200 MHz Pentium Pro PC running Windows NT 4.0, *EUGene* took approximately the following time for specific calculations:

5 seconds for COW National Capabilities Index calculations;

24 minutes for tau-b calculations;

20 minutes for expected utility calculations (War Trap version);

About 150 days (yes, days) for complete risk attitude calculations with typical genetic algorithm settings (as the number of countries in the system grows, computing risk data takes exponentially longer; for a single year in the mid-1970s such calculations takes 2-3 days, while a computation in 1981 (say) takes 7-8 days);

20 minutes for the expected utility calculations (War and Reason version);

30-45 minutes to output data on all dyads, 1816-1993, outputting ccode, year, capabilities, risk, and expected utility. If you output data while specifying backwards induction to generate expected utility equilibria, output will take approximately ½ hour longer than when using the logical conditions. In addition, adding more variables will slow the total time to output the data set.

The program consists of approximately 51,000 lines / 1.7M of computer code split into 57 units and various Windows forms. The final executable file is about 1.6M. Source code is distributed with the program. Distribution is from the *EUGene* web site, <http://EUGenesoftware.org>, maintained by D. Scott Bennett, The Pennsylvania State University, e-mail . *EUGene* is Copyright 1997-2005 D. Scott Bennett, Jr. and Allan C. Stam III.

Installation Procedure

EUGene can only be installed on Windows 95 (or higher) and Windows NT 4.0 (or higher) systems; this includes Windows 98, Windows 2000, Windows ME, and Windows XP.

Installation from CD

To install *EUGene* from CD, you must run the installation routine from the CD. This will unpack all necessary *EUGENE* files, including the main program executable file, source code, and input data. Most of this space is data files, in particular expected utility data.

1. Insert the *EUGene* CD into the CD drive on your PC.
2. Setup should begin automatically. If it does not (which may happen if the Windows "autorun" is not enabled on your PC), then use the Windows Explorer to locate the CD-Rom drive and double-click the file "SETUP.EXE". OR you may use the "Run" command under the "Start" button to run "SETUP.EXE" from the root CD directory.
3. You will be prompted for installation options, but should normally just accept the defaults. You may install *EUGene* to any directory of your choice; if necessary this directory will be created automatically. Running setup will extract the program and data files, and by default will create a new "*EUGene*" group in Windows under "**Start – Programs**".

Installation from download

To install *EUGene*, you must download a set of files to your PC, and then run an installation routine that will unpack all necessary *EUGENE* files, including the main program executable file, source code, and input data. Most of this space is data files, in particular expected utility data.

1. Create or identify a directory (such as "c:\temp") on your machine where *EUGene*'s installation files can be kept. This can be any directory you want. Once installation is complete, you can delete the initial *EUGene* setup file that you download to this directory.
2. Access the *EUGene* web site at <http://EUGenesoftware.org>.
3. From the menu items listed on the initial screen, select "Download." Decide whether you want the demo or full version of *EUGene*.
4. Download the main setup file "SETUP.EXE" by clicking on the appropriate link in the download page. Download the file to the temporary directory you identified in step 1.
5. In the Windows Explorer, double click on the "SETUP.EXE" file in your temporary directory, OR use the "Run" command under the "Start" button to run "SETUP.EXE" from that directory. You will be prompted for installation options, but should normally just accept the defaults. You may install *EUGene* to any directory of your choice; if necessary this directory will be created automatically. Running setup will extract the program and data files, and create a new group in Windows under "**Start – Programs**".

After Installation:

1. You should read the conditions of use noted in the file "LEGAL.TXT" in the directory where you installed *EUGene*. Note that none of the program files (source code, data, or executable files) may be redistributed.

2. Program documentation is included as an MS-Word document titled "*EUGeneDocumentation.doc*", as a text file titled "*EUGeneDocumentation.txt*", and as a rich text format file in "*EUGeneDocumentation.rtf*". This will be in the "documentation" subdirectory.
3. To run *EUGene*, select the icon labeled "*EUGENE*" in the *EUGENE* program group under the start menu (Start | Programs | *EUGene*), or double click the *EUGENE32.EXE* icon in the Windows Explorer in the "*C:\EUGENE*" directory. If you selected the option during installation, you may also have a *EUGene* icon on your desktop which may also be double-clicked to run the program.

Creating Shortcuts

To run *EUGene*, you will usually select the icon labeled "*EUGENE*" in the *EUGENE* program group under the start menu (Start | Programs | *EUGene*), or double click the *EUGENE32.EXE* icon in the Windows Explorer in the "*C:\Program Files\EUGENE*" directory. However, you may also want to create a shortcut to *EUGene* on your desktop or in another program group. If you do this, you should check the "properties" of the shortcut to ensure that the location in the "start in" directory (under the "shortcut" tab) is set to the directory where you installed *EUGene* (typically *C:\Program Files\EUGENE*).

Uninstalling *EUGene*

To uninstall *EUGene*, double-click "***Add/Remove Programs***" in the Windows Control Panel [select ***Start - Settings - Control Panel***]. You will see an entry for *EUGene*. Select it, and click "***Remove***." *EUGene*'s files will be removed from your system. If you have created data sets using *EUGene*, or modified the input files, you will have to delete those files (and the "*C:\Program Files\EUGENE*" directory, or other directory where you installed *EUGene*) manually.

Running *EUGene*

The following is a summary of what the user sees and the user selections that can be made when the program runs.

Startup

Double-click the *EUGene* icon in the *EUGENE* program group, or double click the *EUGENE32.EXE* icon in the Windows Explorer in the "*C:\Program Files\EUGENE*" directory.

NOTE: When *EUGene* begins, it will first carry out a number of initialization tasks, including reading the list of nations available for analysis and creating various windows onscreen. This may take several seconds to complete.

After initialization is complete, click the button labeled "click to continue" to proceed to the next step.

Menu Options

The main *EUGene* Window and set of Menus will appear. At this point, the user can choose from 4 menu options: File, ReCompute, Create Data Set, Trace and Help.

File Menu

Use this menu to exit the program, and to save and load program settings.

Save Settings: Selecting "Save Settings" will save all of the current specifications you have entered in your *EUGene* run in terms of population of cases selected, output files and format, variables selected, sampling proportions, and MID and case exclusions.

Load Settings: Selecting "Load Settings" will load your previously saved settings.

Exit: This will exit *EUGene*.

Browser Menu

EUGene now has a built in data browser for viewing select datasets. The drag down menu in the top left allows the user to select a dataset to load. Once a file is loaded the user can adjust whether the first column is fixed and search from the current highlighted line down to find a specific value or phrase.

Warning: The browser is in an early phase and has not been fully optimized. Because of the size of some datasets, and the way the DELPHI programming language deals with spreadsheet type grids, the browser uses a large amount of system resources.

ReCompute Menu

Use this menu to force recalculation of variables pertaining to expected utility. Under this menu, the user can select one of five different calculations:

National Capabilities Index (Percent System Capabilities): Select to recalculate each state's proportion of total system capabilities, as defined by Singer, Bremer, and Stuckey (1972).

Tau-b Scores: Select to recalculate Tau-b scores between states' alliance portfolios, to be used as utility scores, as defined by Bueno de Mesquita (1975) and Bueno de Mesquita (1980).

S Scores: Select to recalculate Signorino and Ritter's (1999) S measure of similarity of alliance portfolios.

Expected Utility Calculations (War Trap Method): Select to recalculate expected utility as operationalized by Bueno de Mesquita (1980). This version of these calculations incorporates both the bilateral and multilateral components of expected utility through chapter 3 equation 6 of *The War Trap*, but does *not* incorporate risk attitude as operationalized in *The War Trap*.

Risk Attitude: Select to recalculate risk attitude scores as operationalized in Bueno de Mesquita (1985). Note that running this procedure does NOT require re-running the Expected Utility War Trap calculations previously. The routine recomputes expected utility using the war trap method internally.

All Years: Will recompute risk scores for all years from the first_risk_year to last_risk_year as specified in EUGene.ini. Prompts to either delete existing file, or overwrite records as new records are computed. This latter option will preserve previously calculated risk scores and update/overwrite as it progresses.

Subset of Years: This procedure will recompute risk scores for all states for a particular subset of years. The program prompts the user for start and end year for recomputation. The procedure also prompts the user as to whether the user wants to keep using the existing risk data file and just overwrite records for the specified years, or whether the user wants to save this subset of years to a new risk data file. This option is mainly for programmers; there is a separate (and disabled for most users) procedure to merge various risk files together.

Append New Years: Eventually, this option will allow calculation of only new risk years which can be appended to the existing data set, for instance if new alliance data becomes available. The procedure begins by checking to see what years are currently in data file, and what settings in the EUGene.ini file specify for the last available data year. However, currently "blank" records have been inserted in the risk data file that run up to 1997, and so the procedure would show data to 1997. As a result, this procedure won't actually work, and has been disabled.

Display Single ccode-risk detail. Procedure will prompt user for a single numeric COW country code and a single year, and will compute and save details of the risk calculation into a text file. The user is prompted for this output file as well. This selection will not

overwrite any risk records in the risk data files. *Expected Utility Calculations (War And Reason Version)*: Select to recalculate expected utility for each of the various game outcomes as defined in Bueno de Mesquita and Lalman (1992).

Expected Utility Calculations (War and Reason Method): Select to recalculate expected utility as operationalized by Bueno de Mesquita and Lalman (1992).

NOTE: The only reason to recompute a variable is if you require a different specification than what *EUGene* used initially. *EUGene* is distributed with data continuing the results of initial default calculations using current versions of the input data and what we believe to be the most common specifications for methods, and thus any variables can be obtained as output by choosing from the "Output" menu without performing recalculation. The default specifications used to create the data distributed with *EUGene* are detailed in Appendix C.

If the user requests a re-calculation and "OK" is pushed, *EUGene* performs the calculation requested for all states and years. Some calculations take a long time to execute - see section 1 of this documentation for some of the original run times. Note that since input data sets vary in their time span/coverage, not all subsequent data can be generated for all years since 1816.

Note that calculations for many variables are dependent on and use data generated by previous calculations. For instance, risk attitude depends on capabilities and tau scores. If the user changes a specification and recalculates a variable that is used as input to a subsequent calculation, for instance by choosing a non-default method of calculating distance, *EUGene* will NOT recalculate the subsequent variable automatically. Thus if a user wants new expected utility values using non-default settings, they must go through several sequential steps. For example, if the user wants to use a different distance method for calculating risk scores, they should first recalculate expected utility (*War Trap* version) to reflect that different distance method as well. Calculations are dependent on one another and should be performed in the following order: percent system capabilities; tau-b scores and *S* scores; expected utility (*War Trap* version); risk attitude; expected utility (*War and Reason* version).

Recomputation Options

After choosing an item to recalculate, another window will appear with the allowable modifications that the user can make to the calculation method. Only certain modifications are available for each variable to be recomputed. After any options are specified, clicking "OK" will run the selected recalculation. Clicking "Cancel" will return to the main *EUGene* screen. Clicking "Help" will pop up a brief reminder of what the user is selecting in this screen.

[More information on the details of the calculations is located in section "Variable Calculations", p. 15. Information on the options for recalculation that may be made by the user are located in section "Modifying Assumptions Used in Variable Calculations", p. 23.]

Modifications to COW Capabilities: For computing the capabilities index, the user can choose to either 1) use only the officially released COW data on capabilities, as given by COW; 2) drop the "energy" category from the capability index; 3) include modifications to the COW data as provided by the user in the file "cowmod9-1999.csv". See Appendix B for details on these modifications.

Distance Discounting Options: When computing Expected Utility and Risk Scores, the user can modify how dyadic distance is calculated, and whether or not capabilities are to be discounted by this distance. Note that *EUGene* will compute distance based on the method in its memory, which by default is the distance between capitals, adjusted for contiguity and allowing multiple cities for each country. This default option may be changed for recomputing expected utility values under the recalculation options in the "Recompute" menu item, or for purposes of printing distance by pushing the "Distance Options" button on the variable screen.

Risk Attitude Optimization Method: When computing Risk Scores, the user can specify the search algorithm to be used to locate the best and worst possible security situations for states as steepest descent, genetic algorithm, or a random walk. Selecting "Steepest Descent" will pop up another window requesting the search tolerance (see below) for the descent. Selecting "Genetic Algorithm" will pop up another window requesting user input for the size of the population used in the genetic algorithm search, the probability of mutation per individual alliance per generation, the number of iterations that must be stable before ceasing the genetic routine, the number of top "cloned" pattern to be maintained from one generation to another, and the tolerance of the search (all discussed below). Selecting "random walk" will pop up another window requesting the number of iterations (see below) to be performed.

Risk Data Source: When computing expected utility following the methods in *War and Reason*, the user can choose to use either newly recalculated (and updated) risk attitude scores, or can use the actual risk scores computed by Bueno de Mesquita. Both choices use risk scores following the methods in Bueno de Mesquita (1985).

After the user hits "OK," *EUGene* will begin executing the recalculation as selected by the user. While running, a progress bar will appear on screen showing the progress of the program at doing the selected calculation. If the user wants to interrupt the recalculation, hit "Stop Execution."

IMPORTANT NOTE: Each of these calculations creates an intermediate file containing the results of the selected calculation. If a calculation is interrupted, the intermediate file (see below) containing the results of that calculation will be incomplete, and the program is likely to report an error, either when the program is next run, or at least the next time the program attempts to read that particular intermediate file. *EUGene* does not check for the completeness and integrity of those files. If you must interrupt a running calculation, you will need to replace the corrupted/incomplete intermediate data file. The easiest way to ensure that you do not have a problem when you plan to recalculate is to make a backup of the relevant intermediate file before you start. Then if the calculation is interrupted, you can simply copy the complete initial file back to the *EUGene*\INPUTDAT directory.

Create Data Set Menu

Use this menu to send data to an output file. Output datasets are flat text files, with either a single space, comma, or tab between each variable value. Under this menu, the user can initially select among several units of analysis for output.

[Note that more information on the choices that can be made is located below, in section "Data Set Creation: Additional Specifications and Details", p. 30.]

Unit of Analysis

Country-Year Data: Select this to output data where the unit of analysis is the country-year. That is, each line of data in the output file will contain information about a single state in a year. Within this type of output, the user will be able to select one of several different subsets of states for output including: all states; all great powers; or a subset of states specified by the user from a popup menu.

Directed Dyad-Year Data: Select this to output data where the unit of analysis is the directed dyad-year. That is, each line of data in the output file will contain information about a pair of two states in a year. The output is of directed-dyads, in which Britain vs. Germany in 1939 (for example) is distinct from Germany vs. Britain in 1939. Within this type of output, the user will be able to select one of several different subsets of states for output including: all dyads from all states; all great power dyads; all dyads constructed from a user-specified subset of states; all dyads of contiguous states; all dyads in which at least one member is a great power; or all politically relevant dyads.

Non-Directed Dyad-Year Data: Select this to output data where the unit of analysis is the non-directed dyad-year. Each line of data in the output file will contain information about a pair of two states in a year, but the order of the countries is arbitrary, and only one dyad for any pairs of country is output (non-directed dyadic output outputs Britain vs. Germany in 1939 (for example), but not BOTH Britain vs. Germany and Germany vs. Britain in 1939). In the actual output data, we order the countries so that ccode1 is always less than ccode2, but this is for convenience and does not indicate any particular theoretical order.

Directed Dispute Data: Select this to output data where the unit of analysis is based on the directed dispute dyad. Directed dispute dyads are the dyads in a militarized dispute ordered by initiator (side A of the dispute) vs. target.

Directed dispute-dyad initiation: Each line of data in the output file will contain information about a dispute initiation by one state against another. Every directed dispute initiation (that is, every instance of a state on Side A initiating a dispute vs. a state on Side B) will have a case (subject to program settings for initiators vs. joiners, etc.). If a MID has multiple targets or initiators, then there will be multiple dispute-dyads generated for the MID.

Directed dispute-dyad year: As with directed dispute-dyad initiations, each line of data in the output file will contain information about a dispute initiation by one state against another, but the output will contain multiple lines of data for disputes that continue more than one year. Such cases will contain one observation per year of the dispute (that is, if a dispute starts by December 31 of some year, but continues past January 1, an additional observation will be generated).

Non-Directed Dispute Data: Select this to output data where the unit of analysis is the non-directed dispute dyad. That is, each line of data in the output file will contain information about each dispute interaction involving pairs of states on opposite sides of a dispute. Disputing dyads A vs. B are ordered by country code number (so $A < B$).

Non-directed dispute-dyad onset: Each line of data in the output file will contain information about a disputing pair of states that had a dispute onset. Every dispute interaction (that is, every instance of a state on Side A being in the dispute vs. a state on Side B) may have a case, subject to program settings for including initiators or joiners, etc. If a MID has multiple states on a side, then there will be multiple dispute-dyads generated for the MID.

Non-directed dispute-dyad year: As with nondirected dispute-dyad onsets, each line of data in the output file will contain information about a dispute onset between two states, but the output will contain multiple lines of data for disputes that continue more than one year. Such cases will contain one observation per year of the dispute (that is, if a dispute starts by December 31 of some year, but continues past January 1, an additional observation will be generated).

Output Choices

[Note that more information on the choices that can be made is located below, in section “Data Set Creation: Additional Specifications and Details”, p. 30.]

After choosing a unit of analysis for the output dataset, a paged window will appear requiring the user to define how his or her output will be constructed. This window includes several tabs (some of these tabs will not appear if the user chooses to output country-year data).

Tab 1: Files / Format

Output Destination: specifies where the output should be sent (screen, printer, or file). If the user selects "File" they will be prompted for file name and location. If a selected disk file already exists, the user will be prompted for whether to overwrite it. Files marked as "read-only" in the operating system cannot be selected. If "Screen" is selected a new window will be opened for the output. Note that the "Screen" option only displays about 3280 lines of data. If the user's request results in more than 3280 observations, only the first 3280 lines are shown. If the user wants to view the entirety of a data set larger than 3280 lines, the user can save the data set to a user specified file using the "File" option, then open and view the data through another program (e.g., WordPad or Excel).

Header Information: If checked, the first line of the output data file will be a line of variable names identifying the variables in the file.

Create Command File(s): allows the user to have *EUGene* automatically create command files to make it easier to read data into other software packages. Between 0 and 3 boxes may be checked to create command files for the listed programs. If you ONLY want to create command files, and not output a data set to an output file, then mark the "Create Command File Only" check box.

Variable Separator: specifies whether the numeric variables in the output data will be separated by a tab, space, or comma. *EUGene* output will be in a flat ASCII text file.

Tab 2: Population of Cases

Years to Include in Output: specifies what years should be reported in the output. Clicking "All years" will report all possible years for which any data are available. Clicking "Specified Range" generates data only between the specified first and last year. Note that if data is not available in some year, a missing value (typically either a "-9" or a "-99") will be reported in the output file.

Country-Years Included or Dyad-Years Included: specifies what combination of countries (*all states, great powers only, or a selected subset*) or dyads (*all dyads, all major power dyads, all contiguous dyads, all dyads within a specified distance, all dyads within a specified set of regions, all dyads where at least one state is a major power, all politically relevant dyads, a set of dyads from a user-specified list of countries, or dyads read from a user input file*) will be included in the output.

Self-Referencing Dyads if checked, dyads made up of a particular state matched against itself will be included in the output. For example, checking this box will print information on dyad 200 vs. 200. Normally this is an irrelevant dyad, since analysis does not normally include the possibility of a conflict between a state and itself, and by default the option is unchecked. For many calculations, such as expected utility, final variable values when a state is selected to be against itself are undefined. Thus including these dyads will include a lot of missing values in the output file.

Tab 3: Sampling (directed and nondirected dyad year only)

Instead of outputting all of the cases that meet the criteria you specify in the "cases" tab, you may choose to output a sample. To output all cases as specified, mark the "No sampling" option on this tab. To output a sample of cases, mark the "Stratified Random Sampling" option. Then, select the proportion of cases (from 0 to 1.0) that you want to output. You may make different selections for the proportion to sample from dyad-years without any dispute and dyad-years with a dispute. Because dispute years are so rare relative to non-dispute years, your sampling proportion for dispute-dyad-years should usually be greater than or equal to the proportion for non-dispute-dyad-years. This procedure is an implementation of retrospective sampling as recently discussed by King.

Sampling works as follows. Immediately before any given directed-dyad-year is output, a random number between 0 and 1 is generated. If this random number is less or equal to the specified sampling proportion, then the case will be output. So a non-dispute-dyad-year sampling proportion of 0.95 means that each non-initiation case has a 0.95 chance of being selected. Because this probability is assessed for each directed-dyad-year, the actual number of cases output may differ slightly from run to run.

Use Specific Random Number Seed: *EUGene* uses a random number seed to initialize its random number generator. Normally this seed is itself selected randomly. However, if you wish to be able to replicate the generation of a subset of data, mark the check box and enter a seed value here, and record this value for later reference. If you wish to generate exactly the same set of data again, simply use this same value again.

Tab 4: Variables

Variables to Include in Output: what variables are to be included in the output. Checked variables will be sent to the output location.

Include Header Line Containing Variable Names: When checked, a header line will be put in the output file. Including a header line will add a line as the first line of the output file with a label (name) for each variable. If no header line is included, only numbers will be in the file and the user must record separately the variables that were selected for the output file. The names used in the header line should be self explanatory, but are detailed below. Each name is 8 characters (or less) in length so that they are usable in statistics programs.

Tab 5: Case/Conflict Exclusions

(Note: available for directed and non-directed dyad-year, and directed dispute only)

Exclude Conflict Dyad-Years Based on Which Data Set? The data set identified here will be used to determine which dyad years (if any) are dropped from the created dyad-year data set because of ongoing conflicts, conflict joining, and presence of a dispute initiation in the “reverse” direction. Only one data set may be identified, and all selected options will be applied based on that data set only. The options on the page will not be available for selection until a “master” exclusion data set has been selected.

Note: if Directed Dispute Dyad Data is selected for output, the caption and options on this selection will change. If Directed Dispute Dyad Data is selected, then the options here will be limited to COW MID and Maoz MID data sets, and the choice will determine whether directed COW dispute, or directed Maoz disputes, are the basis for output.

Dyad Years with Ongoing Conflicts: Drop All Dyads with an Ongoing MID, Include Ongoing Dispute Dyad Year iff New Dispute, or Include All Dyads with an Ongoing MID: Under the default option of “Drop All Dyads with an Ongoing MID,” *EUGene* output does not report information for any dyad year in which a militarized dispute that started in a previous year was still ongoing on January 1. That is, those dyad years are dropped from the output. Setting a different option forces output of dyad-years with ongoing disputes. Setting “Include All Dyads with an Ongoing MID” forces the output of directed dyad-years even if there was an ongoing dispute at the start of the year. If the “Include Ongoing Dispute Dyad Year iff New Dispute” option is specified, directed dyad-years will be output either if 1) there is no ongoing dispute at the beginning of the year, or 2) if there is an ongoing dispute but there was also a new dispute initiation by state A vs. state B in the year. See further discussion about ongoing dispute years under the header “Ongoing Dispute Year Options” in discussions concerning case exclusions.

Dyad Years with Ongoing Conflicts: Treat Ongoing Dispute Years as Initiations: Normally *EUGene* codes the output variable “initiation” as a “1” only in the first (initial) year of a MID; subsequent directed dyad-years would be coded as a “0” for initiation. If

you want to analyze a data set in which you consider every ongoing dispute year as a new initiation, then mark this option and the “initiation” variable will be coded as a 1 in the first and all subsequent years of a dispute. To include those ongoing year dyads also included in the output, you should also mark either the “Include Ongoing Dispute Dyad Year iff New Dispute,” or “Include All Dyads with an Ongoing MID” option.

Target vs. Initiator Dyads: When one state initiates a MID vs. another, an initiator and a target are designated. If an initiator A starts a dispute against a target B, then it is problematic to include the directed dyad B vs. A for analysis. The reason is that when A initiates vs. B, it may remove B’s ability to initiate a dispute against A in that same year (the B vs. A dyad is censored). By default, under the option “Drop Target vs. Initiator Directed Dyads if no new MID,” if A initiates a MID vs. B in a given year, and there is no initiation by B vs. A in that same year, the B vs. A directed dyad will NOT be included in the output. With this setting *EUGene* will output directed-dyads B vs. A only where 1) A initiated a MID vs. B but B also initiated a MID vs. A in that year; or 2) neither state initiated a MID vs. the other. The second option available in this section will force the inclusion of the B vs. A dyad (you can think of this as “both directions of all dyads”) even when A initiates vs. B. Marking “Keep Target vs. Initiator Directed Dyads if no new MID” forces output of all cases B vs. A. Those cases will of course have a non-initiation coded on the dispute variable. Note: If non-directed dyads have been selected for output by the user, then the user will not have this option. In non-directed dyads, there is no distinction between the target and initiator in the dyad.

Dyads Involving Joiners: States who join late into an ongoing MID face different conditions that may affect their decisions than did the states who originated/initiated a dispute. A “Joiner” is considered to be any state who enters a MID after the first day. [Note that because of the way the MID data is set up, it is possible for “joiners” to be voluntary joiners who choose to intervene on one side of a MID, or they may actually be targets of the expansion of a MID. “Joiners” really means “latecomers” and should not be read to imply a voluntary choice to become involved.] If the option “Include all Joiner Dyads” is marked, then all directed-dyad years where one or both states are joiners will be included in the output data. If the option “Drop all Joiner Dyads” is marked, then any directed dyad where one or both states are joiners will be omitted in the output data set.

See further discussion about joiner variables and interpretation issues below, under the header “Joiners” in discussions concerning case exclusions.

Missing Values

For most variables, a “-9” in the output indicates that the value is missing. Exceptions include some user variables and the set of polity variables, “dem1” “dem2” “demlg1” “demlg2” “demchg1” “demchg2”, etc. (all polity variables as listed below) for which the values -66 through -99 are missing (-9 is a valid value for these variables; -66 through -88 represent various types of polity interruptions). It is incumbent on the user to ensure that when data is read for subsequent analysis, -9 and -66, -77, -88, and -99 are coded as missing for the appropriate variables. The Stata, SPSS, and LIMDEP command files that *EUGene* creates to read the data contains commands to convert missing values appropriately. Further details of those command files are provided below, in the “Data Set Creation: Additional Specifications and Details” (page 30) section.

Create Hypothetical Alliance Dataset

This command allows the user to output components of internal risk score calculations, specifically hypothetical alliance patterns that would maximize or minimize security. Risk attitude scores are

computed following Bueno de Mesquita's (1985) procedures that measure risk on the basis of where a state's actual choice of alliance partners falls between a potential most-secure alliance pattern and a least-secure alliance pattern. If states choose to form alliances that give them close to the safest available (most-secure) alliance, then they are classified as risk-averse. If they form alliances that leave them vulnerable although they have the potential to do better, then they are classified as risk-acceptant. As part of computing these risk-attitude scores, EUGene evaluated a very large search space of hypothetical alliances to identify the best and worst possible alternative alliance arrangements available to states in each year. This command allows for the output of these hypothetical/potential alliance arrangements.

Options

Scores: Reports hypothetical alliance patterns based on either Tau, or unweighted S. Currently only Tau and S unweighted scores are supported.

Best/Worst Alliance: User can output only the best hypothetical alliance pattern (which maximizes security) , only the worst (minimizes security), or both.

Year Range: Either all years, or a specified subset can be outputted.

Header Information: Because the number of countries involved changes each year, the user will likely want the header for each year. Without the header information identifying what countries are involved will be difficult with a country list format. Note – this option is automatically disabled for Directed Dyad-Year lists as it is unnecessary.

Output To: Select either a file or the screen for output.

Format Type: Country List outputs a list of potential alliances for a country with every other country in the system during the year.

Directed Dyad-Year format outputs the alliance information in the format

ccode1 ccode2 year region besttype worsttype

This specifies that between ccode1 and 2 in year, regarding this region, the best type (maximizing ccode1's security) of alliance for 1 to have with 2 is of type "besttype," which ranges from 1-4 (defense pact to no alliance, per the alliance variable codes). Note that in a Directed Dyad-Year format, ccode1 may have a different best alliance with ccode2 than ccode2 would have with ccode1 to maximize (or minimize) ccode2's security in a given year.

Country list format outputs the information in a format with two rows of output data per country per year per region. Variables appear in the following order.

Year CCode BestWors Region 2 200 210 ... [additional ccodes here]

Each variable/column after the "Region" variable labels a second country, and the entry at the intersection contains the alliance type between the row and column actor in that year. For instance, in a row with year 1816 and ccode 2 (US), where "best/worst is "1" indicating "best alliance" then the entry in the column labeled "200" contains the alliance type between ccodes 2 and 200 in 1816 in the configuration that maximizes US security.

Variable Separator: Allows the user to specify whether variables in the output file are separated by tabs, commas, or spaces.

Variables in output: In various output files, users may see the following variables

- Year: year of hypothetical alliance
- CCode: country code of country for whom hypothetical alliance information is being output.

- BestWors: A “1” indicates that alliance types are being reported in this row that maximize security (“best” security), and a “2” indicates that alliance types are being reported that minimize security (“worst” security).
- Region: States have risk attitude regarding each world region. This number indicates the region with regard to which security is being maximized/minimized in this output entry.

User Data Menu

The commands under the “User Data” menu allow *EUGene* users to prepare and submit data to the *EUGene* project, and to download additional datasets to access through *EUGene*. Complete procedures are described in the auxiliary documentation file, “*EUGene* user dataset procedures.doc”. This file is available through the “Help|User Data Sets” menu item within *EUGene*, or by locating the file in the “docs” subdirectory within the directory where you installed *EUGene*.

Prepare Data Set for Submission

The "Prepare Data Set for Submission" menu guides the user through a step-by-step procedure for submitting a dataset. Following the steps here ensures that user datasets are configured correctly, and that supporting materials are provided. All data must be submitted in a .csv (comma-delimited) format. In addition, a rich-text format documentation file is also required with any data submission. Finally, *EUGene* creates a configuration file which stores key information about the data set and variables. All three of these files must be included in data submission. Following dataset uploading to the *EUGene* server, the programmers request those researchers who do submit data to e-mail the programmers to ensure prompt processing of submitted data files so that they may be downloaded by others. E-mail addresses are available in the main *EUGene* documentation.

Transfer User Data Sets to/from Website

The "Transfer User Data Sets to/from Website" menu item connects the user to the *EUGene* server for direct transfer of datasets. In particular, it is through this menu that new data sets can be downloaded from the *EUGene* website via FTP so that variables from these datasets are available for use in *EUGene*. Once user data sets are downloaded from the *EUGene* server, users should exit and reenter *EUGene* for those data sets to appear as available.

Trace Menu

Under this menu, the user can set a program trace to be either “on” or “off.” If “Trace” is on, an additional window appears onscreen which traces some of the internals of program execution. Whether this is on or off has no effect on the output generated or results of the program, and is mainly to allow curious users to spy on program execution.

Under this menu, you may also reveal the output window if you have closed it.

Help Menu

Use this menu to obtain on-screen help while *EUGene* is running. You may use this menu to open up the various documentation files that accompany *EUGene*, to open some general help resources within *EUGene*, or to access our website. Help within many individual *EUGene* screens appropriate for the context of that screen can be accessed simply by clicking the “Help” buttons on each screen.

While EUGene is Running

While *EUGene* is running to produce output or recalculate variables, a progress bar will appear on screen to show *EUGene*'s progress. The "STOP" button provided on that bar will abort the program run, exiting *EUGene* and leaving the user in Windows. The "PAUSE" button will pause the program, for instance if a user wants to perform some other CPU intensive task without fully stopping *EUGene*.

Exiting EUGene

After all desired calculations are performed and output created, pick File | Exit.

Reading Data into Other Software Programs

After *EUGene* has created data, you will want to read it into other software for analysis. *EUGene* can automatically create command files to read data into SPSS, Stata, or LIMDEP. To create command files, when selecting your output options, under the "Destination/Format Tab", be certain to send your data to an output file, and then check one or more of the boxes to create command files. After your data is generated, look in the destination directory for a file with the same name as the data file you created, but with an extension of ".sps" ".do" or ".lim". A file with one of these extensions will be a command file to read the data into SPSS, Stata, or LIMDEP respectively. For example, if you create a data set and name it "BIGOUT.DAT", and you choose to create an SPSS command file, you will find in the directory where you saved the data a command file with the name "BIGOUT.DAT.SPS". To use these command files, run SPSS, Stata, or LIMDEP and read the output file into the program. Then, following instructions provided with that software, select and run the commands in the file. Each command file contains directions for exactly how to use the command file to read the data into the relevant program. In general, you will run the statistics software, load the command file, and then tell the statistics program to execute the commands that you have loaded.

There is an option on the "variables" output page to include a header line. This option will print a line with a label/variable name for each variable. Note that if you plan to use a command file to read your data set into one of these files, you do NOT need *EUGene* to output a header line on the "variables" output page. If you do output a header line, your data set will contain one case with every variable appearing as missing data.

EUGene also provides an option to generate a command file **only**. You would probably use this command only if you generate a data set and forget to have *EUGene* construct a command file. When this option is selected, *EUGene* will write the statistical package commands to the command file without rewriting every case in the data set.

The command file also contains comment lines that record basic information about the data set *EUGene* has generated, including the number of cases, the selections made by the user (such as the format, time span, sampling, and variables), and the necessary citations based on those selections. Note: when the "command file only" option is set, the command file will NOT contain the number of cases. In the case of LIMDEP (which expects a specified number of cases when reading the file), using this command file will result in an apparent error message stating that the end of file was found. This is normal when LIMDEP is not given a specific number of cases to read in. However, you may wish to inspect your data matrix to ensure that all cases were read properly to the end of the input file.

You can use these files as templates to create command files to read the data into other programs such as SAS. If you want to examine your output data in a word processor, simply use the "File | Open" command in the word processor and select the output file. *EUGene* outputs flat ASCII text data, which can be easily read. Finally, if you want to examine your output data in a spreadsheet (such as Microsoft Excel), simply be sure to specify that you want to create data separated by Tabs, and use the "File | Open" command. The data should be read into separate columns automatically. If you specified a header line on the "variables" output page, you will also have column headers.

Notes on LIMDEP: An exception to the command file naming convention occurs when creating command files to be processed by LIMDEP. LIMDEP sometimes has trouble with Windows 95 style long filenames, that is files more than 8 characters in length or with multiple periods. When you create a limdep command file, the name of the command file will be shortened to use just the first 8 characters of the name of your data file if it is longer. So for example a data set named "LONGFILENAME.DAT" will have an associated LIMDEP command file of "LONGFILE.LIM". LIMDEP also has limitations on the number of variables that may be read into the program, as well as a limit on the number of variables that can be read in using a single command line. To fit LIMDEP requirements, *EUGene* will 1) output a maximum of 400 characters per line when it creates a data set if you have specified a LIMDEP command file, and 2) display an error message if the data set contains more than 200 variables. If you experience an error when running the command file in LIMDEP, you will have to reduce the number of variables in your data set or utilize a different analysis package.

Variable Calculations and Formulas

EUGene undertakes a number of calculations leading up to the computation of expected utility values and international interaction game equilibria, and uses a number of variables as follows:

National Capabilities / Percent System Capabilities

EUGene calculates the COW composite national capabilities index as developed by Singer, Bremer and Stuckey (1972). This is an index of a state's proportion of total system capabilities in 6 areas: the country's iron/steel production, the country's urban population, the country's total population, the country's total military expenditures, the country's total military personnel and the country's total amount of energy production. First, the state's proportion of total system capabilities in each area is calculated, and second, the average is taken across all of the areas for that state which data is not missing. This variable can currently be computed from 1816-1993. However, note that the data for 1991 to 1993 are SUBSTANTIALLY less complete in the COW input data files than other years, leading to values for national capabilities that are quite suspect for those years. 1993 in particular is problematic. *EUGene* does not do any sort of missing data imputation, nor have we gathered additional data for this study. When data in one of the six capability areas is missing for a state, that proportion is missing. We take the average for a state across all of the proportions that are not missing. So, if a state has no data on energy for instance, its capabilities are its average across the 5 other areas. In doing national capabilities calculations, users have the option of using a data file that we have modified to try to render these data less problematic (see Appendix B, below).

Tau-b Calculation

The Kendall Tau-b correlation calculation performed by *EUGene* is based on the calculation used in Bueno de Mesquita (1975, 1981). Here, the Kendall tau-b is a rank order correlation for two states' alliance portfolios. The alliances of each state are combined into a 4 x 4 table where alliances are ranked as 1 (defense pact), 2 (neutrality pact), 3 (entente) and 4 (no alliance). Note that all states in the international system are included, even if they do not engage in any alliance behavior (these states contribute to the "no alliance" cells of the table). This is different from some of the earlier work by Bueno de Mesquita, where states were included for purposes of calculation only if they engaged in at least 1 international alliance. Tau-b is calculated from this 4x4 table, and ranges from -1 to +1, representing totally opposite alliance agreements to complete agreement in the alliances formed. The actual algorithm used to calculate tau-b within *EUGene* was obtained from Hays, 1981: 603-604.

When generating Taus, *EUGENE* generates both regional and global taus. The global tau is the tau-b score when all states in the international system are included. The regional tau between states A and B is calculated from the 4 x 4 table with only states in the relevant region of A vs. B are included. Bueno de Mesquita in his calculations uses only the regional taus; that is, tau includes only states in a sub-region. Thus while *EUGene* calculated global taus, they are for information only, as subsequent

calculations such as those for expected utility and risk use the regional conception. Also note that in *War and Reason*, Bueno de Mesquita and Lalman modify taus of 1.0 to be 0.999 (see p. 291) in expected utility calculations. We make the same modification when we compute utility and equilibria values. However, when just tau scores are selected as an individual output variable, we present the calculated values without this modification.

Tau-b can be computed from 1816-2000 using the COW v3.0 alliance data.

S Calculation

The *S* correlation calculation performed by *EUGene* is based on the calculation used by Signorino and Ritter (1999). Like the Kendall tau-b, *S* evaluates the rank order correlation for two states' alliance portfolios. Unlike tau-b, *S* also takes into account both the presence and absence of an alliance in the correlation calculation. For example, the fact that a state has identical alliances with some states as well as no alliances with identical sets of other states is accounted for in the *S* calculation, but not in tau-b. Signorino and Ritter note that while tau-b is an excellent measure of rank order similarity, it has inherent flaws that allow a variety of alliance portfolio tables to have identical tau-b results which impact studies looking beyond rank order similarity.

Like tau-b, *EUGene* calculates *S* on both regional and global levels, with the same country inclusion criteria as with the tau-b calculation (see above). In addition, *S* is calculated in both weighted and unweighted forms, taking into account country capabilities and Signorino and Ritter's discussions about the differences in weighted and unweighted *S* values. For use in expected utility calculations, the regional *s* score is used, and like taus, *S* scores are modified from 1.0 to 0.999 between states. When *s* scores are selected as an individual output variable, we present the calculated values without this modification. Also, like tau-b, all states in the international system are included, even if they do not engage in any alliance behavior (these states contribute to the "no alliance" cells of the table). This is different from some of the earlier work by Bueno de Mesquita, where states were included for purposes of calculation only if they engaged in at least 1 international alliance.

S can be computed from 1816-2000 using the COW v3.0 alliance data.

Relevant Region and Regional Identification

Various EU calculations use information on the states "involved in" a region, and refer to the "relevant region" for dyad A vs. B. The inclusion of states "involved in" a region in tau tables, and adding states "involved in" a region to create security levels for risk are examples of calculations where regional involvement matters. States are defined by COW as belonging to particular regions; all states are considered to be involved in their home region. However, certain states are also involved politically in regions other than their own as well. For example, while the home region of the U.S. is North America, the U.S. is involved in Europe from 1899 on. These regions are fully defined in Bueno de Mesquita (1981:95-98). *EUGene* codes states as being involved in regions other than their home region following Bueno de Mesquita, with dates as follows (present refers to the end of the alliance data):

Americas:

Spain, Portugal, UK, France, Holland/Netherlands 1816-present.
Russia/USSR 1946-present, and throughout the 19th century, 1816-1899.

Europe:

Turkey 1816-present.
US 1898-present.
Japan 1895-1945.
China 1950-present.

Asia:

Britain, France Germany, Holland/Netherlands, Spain, Portugal, US, Russia/USSR 1816-present.

Middle East:

Austria-Hungary 1816-1918.
Italy 1860-1943.
UK, France, Germany, Russia/USSR 1816-present.
Spain 1816-1936.
US 1898-Present.
Greece 1828-present.

Africa:

US 1946-present.
Italy 1816-1943.
Prussia/Germany 1816-1945.
Belgium, Portugal, Russia/USSR, Britain, France 1816-present.
China 1946-present.

These dates may be updated in the future as information becomes available about when states actively became involved in different regions.

The concept of "relevant region" is necessary for almost all calculations about a dyad. Each directed dyad A vs. B is considered to have a single relevant region (see Bueno de Mesquita 1981:97, fn 3, where he develops and uses this idea, although he does not use the explicit designation of "relevant region"). This relevant region determines what subset of countries is included in calculations. For example, if tau is calculated between the U.S. and Britain, it would be plausible to include countries in North and South America, or countries in Europe. The relevant region definition and function resolves any doubt about what region's countries will be included. Generally dyads have as the relevant region the region of the potential target, so in dyad A vs. B, B's region is defined as the relevant region. The exception occurs in dyads where the (potential) initiator and target are from different regions and the initiator is *not* involved in the opponent's region. In this case the relevant region of the dyad is considered to be the initiator's region. So for example, the relevant region for both Britain vs. Thailand and Thailand vs. Britain is Asia, and only countries involved in Asia in a given year would be included in tables for creating Tau-b scores, risk scores, etc.

Uncertainty

Bueno de Mesquita and Lalman (1992:298) operationalize uncertainty as the variance in risk-taking scores. *EUGene* computes this variance across all states in either the home region of an individual state or the relevant region of the dyad when a dyad (A vs. B) is under consideration.

Distance

Distance is calculated between national capitals using the "great circle" distance formula. Following Fitzpatrick and Modlin (1986:xi):

$$\text{Cos}(D) = (\text{Sin}(L1) * \text{Sin}(L2)) + (\text{Cos}(L1) * \text{Cos}(L2) * \text{Cos}(\text{DiffLo}))$$

Where L1 = the latitude of place 1,

L2 = the latitude of place 2,

DiffLo = the difference in longitude between place 1 and 2,

D = the arc distance (in degrees) between places 1 and 2.

Within the formula, northern latitudes must be specified as positive and southern latitudes as negative; eastern longitudes must be specified as positive values and western longitudes as negative.

Given the arc distance D, distance in miles can be computed by multiplying D by the average number of miles per degree, 69.16.

The *EUGene* method for distance calculation is a generalized, uniform version of the state-to-state distance calculation used in *The War Trap*. The formula above calculates the distances between capital cities in most cases, except for the cases of the USSR and US when other cities are included (If

the user selects “Capital to Capital with contiguity and War Trap Location changes” under the recalculation options or variable page “Distance Options” button. Note that if the user selects “Capital to Capital with contiguity” option, states with land borders will be considered to be (and reported as) zero (0) miles apart. The 1993 version of the COW contiguity data set is used in these calculations.

Expected Utility Calculation (The War Trap)

This calculation figures the EU of each dyad constructed following procedures and methods described in *The War Trap*. The EU generated is the sum of the bilateral and multilateral expected utility components, but *without* risk attitude. Thus the EU (*War Trap*) values generated by *EUGene* are *not expected* to exactly equal those reported in *The War Trap*. The formula for the bilateral component of EU in a dyad i vs. j is

$$EU_{\text{bilateral}_{ij}} = P_i * (1 - U_{ij}) + (1 - P_i) * (U_{ij} - 1) \quad (\text{Bueno de Mesquita 1981:47, eqn 1})$$

where $U_{ij} = \tau(i, j)$, and where $P_i = \text{adj_cap}_{ij} / (\text{adj_cap}_{ij} + \text{adj_cap}_{ji})$, where adj_cap_{ij} is i 's capabilities adjusted for distance to j , and adj_cap_{ji} is j 's capabilities adjusted for distance to i (which is a 0 adjustment). The formula for adjusted distance is (Bueno de Mesquita 1981:105):

$$\text{adj_cap}_{ij} = \text{raw_cap}_{ij}^{\log_{10}[\text{miles}/\text{miles per day} + (10 - \epsilon)]}$$

where raw_cap_{ij} is the COW composite capabilities score (Singer, Bremer, and Stuckey 1972), distance is computed using the “great circle” distance formula given above with changes for multiple cities in large countries and contiguous states having distance 0, and miles per day are 250 from 1816-1918, 375 from 191-1945, and 500 from 1946 onward.

The formula for the multilateral component of EU in this dyad is

$$EU_{\text{multilateral}_{ij}} = \sum_k ((P_{ik} + P_{jk} - 1) * (U_{iki} - U_{ikj})) \quad (\text{Bueno de Mesquita 1981:58, eqn 5})$$

Where P_{ik} is i 's perception of its probability of success vs. j , given that state k aids i ; P_{jk} is i 's perception of its probability of losing vs. j , given that state k aids j ; U_{iki} is i 's perception of k 's utility for i [$=\tau(k, i)$]; and U_{ikj} is i 's perception of k 's utility for j [$=\tau(k, j)$]. The sum is over all third states k involved in the relevant region, that is, k involved_in relevant_region(i, j) and $k \diamond i$ and $k \diamond j$.

In this formula, P_{ik} and P_{jk} are estimated as follows ((Bueno de Mesquita 1981:108-109):

$$P_{ik} = (\text{adj_cap}_{ij} + \max(\text{adj_cap}_{kj}, \text{adj_cap}_{ki})) / (\text{adj_cap}_{ij} + \max(\text{adj_cap}_{kj}, \text{adj_cap}_{ki}) + \text{adj_cap}_{ji});$$

$$P_{jk} := (\text{adj_cap}_{ji} + \max(\text{adj_cap}_{ki}, \text{adj_cap}_{kj})) / (\text{adj_cap}_{ji} + \max(\text{adj_cap}_{ki}, \text{adj_cap}_{kj}) + \text{adj_cap}_{ii});$$

where adj_cap_{kj} is k 's capabilities adjusted for distance to j , etc.

The final EU calculation based upon The War Trap formulas combines the EU for the bilateral wars and the EU for multilateral wars. So,

$$EU_{\text{final}_{ij}} = EU_{\text{bilateral}_{ij}} + EU_{\text{multilateral}_{ij}} \quad (\text{Bueno de Mesquita 1981:59, eqn 6}).$$

Because this measure is dependent on τ or S (and hence alliance) data, along with capabilities, it can be computed for years when those two data sets are both available. The user must specify whether τ or S is used in these calculations.

Risk Attitude

This calculation is based on the ‘War Trap Revisited’ (Bueno de Mesquita 1985), where risk scores are based on the sum of other states' utility toward a state. That is:

Define state i's actual security level as $\sum_{j \neq i} E(U_{ji})$, that is, the sum of all other states' expected utilities vs. i. Then define $\sum E(U_{ji})_{\max}$ as the security level associated with the hypothetical alliance pattern that would leave i most vulnerable to defeat (when i is vulnerable, the sum of other states' EU against i are large). Similarly, define $\sum E(U_{ji})_{\min}$ as the security level associated with the hypothetical alliance pattern that would leave i least vulnerable to defeat, i.e. in the best possible security position.

Then, $Risk_i =$

$$(2\sum E(U_{ji}) - \sum E(U_{ji})_{\max} - \sum E(U_{ji})_{\min}) / (\sum E(U_{ji})_{\max} - \sum E(U_{ji})_{\min})$$

$Risk_i = R_i$ in the notation of Bueno de Mesquita (1985). This score ranges from -1 to +1, with -1 indicating a highly risk-averse actor, and a +1 indicating a highly risk-acceptant actor.

Risk scores are region based. Thus, a state may have different risk attitude towards different regions (e.g. US might be more risk averse toward Europe than Africa). Since these scores are region based, the sums above are over the set of states k, where k are the states involved in the region in question.

Note: Bueno de Mesquita (1985) further transforms R_i to r_i , where r_i ranges from 0.5 to 2. In the transformed r_i , a value of 0.5 corresponds to a risk acceptant actor, while a 2 corresponds to a risk averse actor; note that the "polarity" of risk-acceptance and aversion are reversed. The variable r_i is used in calculations of expected utility, as per War and Reason Appendix 1. However, EUGene does not report r_i in its output.

Because this measure is dependent on tau or S (and hence alliance) data along with capabilities, it can be computed for years when both data sets are available. The user must specify whether tau or S is to be used in the calculations.

Programming the algorithm to compute risk attitude scores following these definitions, and then generating actual risk scores, was one of the more challenging programming and computational tasks in creating EUGene. Finding the combination of alliances to maximize or minimize security scores as described above is extremely difficult. An examination of any single potential alliance configuration in which i could find itself is not difficult, as it requires the recalculation of each $E(U_{ji})$ and then a re-summing of security. However, the number of hypothetical alliance configurations (constituting the search space for the computation problem) is huge. The search space consists of all possible alliance patterns between state i and the other states involved in i's region; in a region of 25 countries (typical of Europe) and with 4 types of alliances (defense, neutrality, entente or no alliance), the search space would be 4^{24} potential alliance patterns, or approximately 300 trillion combinations. Because it is impossible to examine exhaustively this space for even one country, let alone the 11,000 country-years that constitute the international system from 1816 to 1993, a search procedure is necessary. Bueno de Mesquita (1985) computed hypothetical risk scores by sampling 30,000 configurations per year (not per country-year or per region-year) from this space and selecting the global minimum and maximum from this search as the hypothetical maximum and minimum for all states. We instead optimized our search by programming a genetic algorithm (Goldberg 1989; Holland 1975) which performs a more systematic search of between 1000 and 2000 configurations per state per region per year (out of roughly 6×10^8 possible configurations). Our search examines roughly two orders of magnitude more hypothetical alliances, searches more "intelligently" with the genetic algorithm, and examines only alliance configurations that each state i could actually find itself in (other states' alliances are held constant). As a result, we identify different (and more accurate) best and worst security configurations, leading to slightly different risk scores from those reported in Bueno de Mesquita 1985.

More specifically, in creating the data distributed with EUGene, we generated risk scores using initial populations of between 20 and 30 alliance configurations, a mutation probability of 0.05 per state per alliance configuration per iteration, and assumed stability after 8 generations without a change in the optimum security value found. We used a single crossover point, and "cloned" the best two

configurations from each generation to include in the subsequent 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 local optimum was reported (i.e. an alliance configuration at which no adjacent configurations were better). When time and multiple computers are available, further work could repeat the optimization with bigger populations and longer stable time to increase further the accuracy of the maximum/minimum identification. *EUGene* can easily accommodate such changes in specification by the user. However, the program took approximately six computer-months on 2 200 MHz Pentium Pro PCs to produce the risk scores distributed with *EUGene* and used in initial publications in this project.

Utility (War and Reason)

Utility calculations are detailed in *War and Reason* (Bueno de Mesquita and Lalman 1992). The main expected utility formulas are given in Table 2.2 (p. 47). Most calculations are detailed in Appendix 1. From the results of the measurements detailed there, the final utilities for various options given in Table 2.2 can be calculated empirically. There are two important notes about those calculations. First, different values are not given to the parameters α , τ , and γ ; Bueno de Mesquita and Lalman do not have empirical methods to distinguish between them. Although it is not clear in the text, for operational purposes each of these parameters is set to a value of 1. Second, there is a slight inconsistency in the measurement of ϕ , the domestic cost term in the utility equations. Theoretically, this cost term is expected to be positive, and more specifically to range between 0 and 1. However, in the operationalization of the utility equations, the utility of the status quo is substituted to represent an increasing cost to a challenge as the status quo improves. As measured, the utility of the status quo ranges from -1 to $+1$. When this utility is negative, the calculations for other expected utility values may yield values that are inconsistent with the basic assumed preference orderings of the international interaction game detailed on p. 47.¹

Bueno de Mesquita and Lalman only analyze Europe, and so all states are in the same region. However, expanding analysis to multiple regions leads to several cases where region becomes important, and where specific assumptions about regional inclusions are not laid out. In the first case, consider the calculation of the initial utilities detailed in Appendix 1 of *War and Reason*. $U^A(\Delta_A)$, $U^A(\Delta_B)$, and $U^A(SQ)$ refer to state A's utility for A's desired outcome, A's utility for B's desired outcome, and A's utility for the status quo, respectively. These utilities could be different for directed dyad A vs. B then directed dyad B vs. A. That is, A's utility for the status quo could be different when we consider A vs. B than when we consider B vs. A. The reason for this is that the relevant region of conflict for A vs. B is sometimes different than the expected region for B vs. A. The risk scores of A (and B) could therefore differ in the calculation for dyad A vs. B and B vs. B, and similarly the Tau values could be different.

For a case where we make assumptions about region inclusions, consider the contributions of third states K to the expected utility of A vs. B. For A vs. B, there is a particular relevant region used for taus and risk scores. For figuring out the tau between K and A, or K and B, there is also a relevant region. However, for those third parties K who must decide whether to support A or B in a conflict, the relevant region is not necessarily the region of A vs. B. Instead, we figure out the relevant region between K and A, or K and B. The logic is that K is making a choice based on its policy preferences vis-a-vis A or B, which are related to the region specified by the relevant_region function. K may not care about the region of conflict, and so uses its own relevant region.

As with calculations for expected utility following *The War Trap* methods, all states K that are involved in the relevant region of A vs. B are included as possible supporters of A and B in the EU-

¹ It is important to note that this is neither a problem of the theory, nor of the operationalizations alone, but with the fit of one element of the operationalization to the theory's assumptions. The fact that phi was allowed to range from -1 to $+1$ in *War and Reason* could actually be viewed as making the operationalization more general than the theory, since the operationalization of ϕ_i to include negative values in effect allows for an externalization boost as a domestic "cost."

War and Reason calculations. Again, though, when their taus with respect to A and B are calculated, the relevant region is considered to be region (K, A) or (K, B).

Normally, the utility values operationalized in Appendix 1 to *War and Reason* are constructed so that the base conditions specified in Table 2.3 are satisfied by construction. However, in a very few cases, due to rounding in the calculations, the value of negotiation actually may appear to equal the value of acquiescence. The cases where this occurs are those where the probability of one side winning are so great that a negotiated settlement is only infinitesimally better than acquiescing.

Also note that in *War and Reason*, Bueno de Mesquita and Lalman modify taus of 1.0 to be 0.999 (see p. 291) in expected utility calculations. We make the same modification when we compute utility and equilibria values.

Because this measure is dependent on tau or *S* (and hence alliance) data, it can be computed from 1816-2000 using the COW v3.0 alliance data. The user must specify whether tau or *S* is to be used in the calculations.

Equilibria (War and Reason)

Any combination of utility values will lead to one and only one expected outcome in equilibrium. It is this equilibrium prediction that makes *War and Reason* game-theoretic rather than decision theoretic, as was *The War Trap* and some other subsequent papers. *Equilibrium predictions should be used as predictors of outcomes rather than straight utility values.*

Equilibria may be generated by either of two methods. First, for any given set of preference orderings (given computed utility values), backwards induction may be used on a set of utility scores to solve the game and predict the outcome. Second, the game may be solved to generate the set of logical conditions required for each equilibrium, and a case (set of utility values) may be evaluated as to where it fits (this second method was used by Bueno de Mesquita and Lalman). *EUGene* allows equilibria to be generated using either method. Normally, these methods would produce the same result. However, because phi ranges from -1 to +1 in the original operationalization employed in Bueno de Mesquita and Lalman (1992), rather than being strictly positive as was assumed in specifying utility equations), some cases empirically appear to violate the basic assumed preference structure of the game. With phi used in this form, then if the logical conditions from *War and Reason* are employed, *different equilibrium predictions may result than if backwards induction is used.* When cases violate the basic preference orderings, then the logical conditions developed assuming that those basic conditions are met are the wrong conditions. *EUGene* provides a solution to this problem, as the user has the option under the “variables” tab of the main dyadic output window to generate equilibria using either the logical conditions given in Bueno de Mesquita and Lalman (1992), or by backwards induction. If the user is concerned that the logical equilibrium conditions are not appropriate to the data, the logical conditions do not have to be used given the option to carry out backwards induction. [We have found in the aggregate empirically that only a few cases are affected by this choice.]

The logical conditions for each equilibrium predicted under complete information conditions are discussed in Bueno de Mesquita and Lalman 1992:71-90. The combinations of preference orderings that must be satisfied to reach each equilibrium are as follows:

War_A: (War started by A): 3 conditions must all be met:

$U_A(\text{War}_A) > U_A(\text{Acq}_A)$ and $U_B(\text{Cap}_A) > U_B(\text{Nego})$ and $U_B(\text{War}_A) > U_B(\text{Acq}_B)$.

Note: a fourth condition listed on p. 72, $U_A(\text{Cap}_A) > U_A(\text{War}_B)$, must be dropped, because there is no way given current operationalization (where $U(\text{War}_A) = U(\text{War}_B)$) to simultaneously satisfy $U_A(\text{War}_A) > U_A(\text{Acq}_A)$ and $U_A(\text{Cap}_A) > U_A(\text{War}_B)$.

Because this condition must be dropped, it is sometimes the case that an outcome will satisfy logical conditions both to be a *War_A* equilibrium and a *Status Quo* equilibrium. Users should think carefully about how to treat those cases where both equilibria are logically possible (for instance, by recoding the overlapping cases as SQ or *War_A* in sensitivity analysis).

War_B: (War started by B): Never expected in equilibrium under complete information conditions.

Acq_A: (Acquiescence by A): Either of 2 main sets of conditions may be satisfied that lead to Acq_A. Either

1) [$U_A(\text{Acq}_A) > U_A(\text{War}_A)$ and $U_A(\text{Acq}_A) > U_A(\text{Cap}_A)$ and $U_B(\text{Cap}_A) > U_B(\text{Nego})$ and $U_B(\text{War}_A) > U_B(\text{Acq}_B)$]

or

2) [$U_A(\text{Cap}_A) > U_A(\text{War}_A)$ and $U_B(\text{Cap}_A) > U_B(\text{Nego})$ and EITHER
2.1) $U_B(\text{War}_A) > U_B(\text{Cap}_B)$ or
2.2) $U_B(\text{Cap}_B) > U_B(\text{War}_A)$ and $U_A(\text{Cap}_A) > U_A(\text{Cap}_B)$].

Acq_B: (Acquiescence by B): 2 main sets of conditions may be satisfied that lead to Acq_B. Either

1) [$U_B(\text{Cap}_B) > U_B(\text{War}_A)$]

and either

1.1) $U_A(\text{Cap}_B) > U_A(\text{Nego})$ and $U_B(\text{Nego}) > U_B(\text{Cap}_A)$ and $U_A(\text{Cap}_A) > U_A(\text{War}_B)$

or

1.2) $U_A(\text{Cap}_B) > U_A(\text{Nego})$ and $U_A(\text{War}_B) > U_A(\text{Cap}_A)$

or

1.3) $U_A(\text{Cap}_B) > U_A(\text{Cap}_A)$ and $U_B(\text{Cap}_A) > U_B(\text{Nego})$ and $U_A(\text{Cap}_A) > U_A(\text{War}_B)$]

or

2) [$U_B(\text{War}_A) > U_B(\text{Cap}_B)$]

and

$U_A(\text{War}_A) > U_A(\text{Cap}_A)$ and $U_B(\text{Cap}_A) > U_B(\text{Nego})$ and $U_A(\text{Cap}_A) > U_A(\text{War}_B)$

and

$U_B(\text{Acq}_B) > U_B(\text{War}_A)$]

Cap_A: (Capitulation by A): Never expected in equilibrium under complete information conditions.

Cap_B: (Capitulation by B): Never expected in equilibrium under complete information conditions.

SQ: (Status Quo): 2 conditions must hold: $U_A(\text{SQ}) > U_A(\text{Nego})$ and $U_B(\text{SQ}) > U_B(\text{Nego})$. Then, one of the following 4 conditions must also hold: Either

1) $U_A(\text{Cap}_A) > U_A(\text{War}_B)$ and $U_B(\text{War}_A) > U_B(\text{Cap}_B)$ and $U_B(\text{Nego}) > U_B(\text{Cap}_A)$

or

2) $U_A(\text{Nego}) > U_A(\text{Cap}_B)$ and $U_A(\text{Cap}_A) > U_A(\text{War}_B)$ and $U_B(\text{Cap}_B) > U_B(\text{War}_A)$ and $(U_B(\text{Nego}) > U_B(\text{Cap}_A))$

or

3) $U_A(\text{War}_B) > U_A(\text{Cap}_A)$ $U_B(\text{War}_A) > U_B(\text{Cap}_B)$

or

4) $U_A(\text{Nego}) > U_A(\text{Cap}_B)$ and $U_A(\text{War}_B) > U_A(\text{Cap}_A)$ and $U_B(\text{Cap}_B) > U_B(\text{War}_A)$

Nego: (Negotiation): If the combination satisfies no other conditions, negotiation is expected.

Because this measure is dependent on tau or S (and hence alliance) data as well as capabilities, it is limited by available data sets containing those variables. The user must specify whether tau or S is to be used in the calculations.

System Concentration and System Movement

Concentration and Movement are defined in Singer, Bremer, and Stuckey (1972). CON is in footnote 5 on page 27; MOVE is in footnote 6 on page 28. *EUGene* modifies these calculations in three ways from the original. First, *EUGene* uses a modified version of the CINC score as noted below under “modifications to COW Capabilities,” option 3. Second, while the original scores were calculated over five-year aggregated periods, *EUGene* calculates both scores yearly. *EUGene* also provides the five-year moving average of MOVE. Finally, MOVE is available calculated across the entire system as well as among the major power subsystem.

Modifying Assumptions Used in Variable Calculations

One of the purposes of *EUGene* is to allow users to regenerate expected utility making different assumptions or using different input data than we did. Several modifications are available to use when recalculating the Expected Utility scores using *EUGene*. These changes will modify the data assumptions under which calculations are performed.

Modifications to COW Capabilities

For computing the capabilities index, the user can choose to either 1) use only the officially released COW data on capabilities, as given by COW; 2) drop the "energy" category from the capability index (the energy category is perhaps the most volatile category and contains the most missing data of the 6 capability categories, especially up to about 1850); 3) include modifications to the COW data as provided by the user in the file "cowmod9-1999.csv". If the "include modified data" option is specified, then the input file "cowmod9-1999.csv" must be present in the "INPUTDAT" subdirectory. The capabilities data that are in the modified file will replace (in memory) the original COW capabilities for the appropriate country-years. Details of the file format are discussed below under "INPUT".

Distance

The user has the option of choosing between three separate methods for calculating distance. Since both the EU and Risk calculations use distance methods, the distance chosen by the user will affect both calculations. So, three slightly different risk scores could be obtained based upon the three different distance options. If the user recomputes expected utility and risk, the distance method selected for intermediate EU calculation must match what is selected when risk calculations are done. If it does not, values for expected utility and risk results will be obtained, but their validity will be in doubt because of the inconsistent inputs. The three different distance options are as follows:

No Distance Discounting: No discounting for distance is performed.

Capital to capital: the distance between two states will be calculated as the distance between their two capitals. Capital cities and their latitude/longitude coordinates are listed in the "INPUTDAT\LATLONG.RAW" file.

Capital to capital, with contiguity: the distance between two states will be calculated using the distance between their two capitals, while accounting for contiguity. Normally distance will be the distance between two states' capitals. However, if the two states are contiguous on land, then the distance is zero, regardless of the distance between the capital states. Contiguity data is from COW; see below.

Capital to capital, with contiguity and location: the distance between two states will be calculated using the distance between their two capitals, accounting for contiguity. In addition, however, for states with a very large landmass, extra cities are also used for the distance calculation. Distance is calculated the shortest distance between two listed cities in the countries. Additional cities are listed in the "INPUTDAT\LATLONG.RAW" file.

Distance is calculated from latitude and longitude coordinates using the "Great Circle" formula, as detailed previously.

Risk Attitude Optimization Method

As discussed previously, risk scores are calculated based on a state's actual security level relative to the hypothetical best and worst security level it could have obtained if it had arranged its alliances differently. When computing Risk Scores, the user can specify the search algorithm to be used to locate the best and worst possible security situations for states as steepest descent, genetic algorithm, or a random walk. Finding the hypothetical best and worst security level and associated alliance pattern requires a search over a large space of hypothetical alliances. 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 1980:95-97. In a region of 25 countries, for instance, which is typical of Europe, the search space is then 4^{24} potential alliance patterns for a 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 impractical to exhaustively search this space even for one country in one year (4^{24} is approximately 2.8×10^{14} , or 300 trillion combinations). These algorithms thus are variants of non-exhaustive search methods.

Steepest Descent: Designate the state whose risk score is being calculated the "reference state." This state's alliance pattern with all other states can be represented as a string of *n* integers, where *n* is the number of other countries in the relevant region and each other country has a unique position in this string. Each integer can take a value of 1 to 4 with a 1 representing a defense pact, a 2 a neutrality pact, a 3 an entente, and a 4 representing no alliance between the reference state and the state in that position. This string is a complete description of the alliances the reference state actually has with all other states. The algorithm then proceeds as follows:

1. The steepest descent algorithm evaluates the reference state's security level at the starting pattern.
2. Then, the best (maximum or minimum) security level the reference state could obtain by changing exactly one of its alliances is found. Each alliance pattern that differs by only alliance (this is $3 \times n$ alliance patterns) is designated as an alliance that is "adjacent" to the starting alliance. Each adjacent alliance pattern is analyzed in turn. Depending on whether security is being maximized or minimized, the alliance pattern with the most improved or worsened security level (respectively) is selected.
3. The selected alliance pattern is designated as the new starting alliance. The procedure in step 2 is then repeated.
4. When all adjacent patterns to some starting point have been examined, and the absolute value of the change in security to all of those patterns differs by less than the specified "risk search tolerance," then the procedure ends. The last alliance pattern identified is the hypothetical best or worst security situation for the reference state.

Sub-menu: When *Steepest Descent* is selected, a second window will pop up requesting the search tolerance for the descent.

Risk Search Tolerance: When no adjacent alliance to a starting alliance differs by more than this amount, the algorithm will stop searching for a better hypothetical minimum or maximum security alliance. Risk tolerance can be altered to be any real number. The larger the tolerance, the sooner the risk search will end. However, the larger the tolerance the less accurate the risk calculation is likely to be because the search for minimum and maximum security alliances will be terminated prematurely. The default setting for the risk tolerance level is .01.

Random Walk: This algorithm is similar to steepest descent in that it iteratively checks patterns adjacent to a starting alliance until there is a minimal change in security at all adjacent alliances. However, while the steepest descent algorithm checked each adjacent alliance and selected the pattern with the most increased (or decreased) security, the *random walk* algorithm will check adjacent patterns in a random order until any pattern is found that increases (or decreases) security. This pattern is selected as the starting point for the next iteration. The search terminates only when no adjacent pattern has a security value that is better (or worse) from the starting point.

Sub-menu: There is no guarantee that a particular random path through adjacent alliances will lead to the best or worst security position for a state. Because of this, when *Random Walk* is selected, a second window will pop up another window requesting a number of iterations to be performed.

Number of Iterations: This many random walks from the reference state's actual alliance pattern will be performed before the result is reported as the best or worst. Multiple iterations increase the probability of finding better (maximum or minimum) alliance patterns and associated security positions.

Genetic Algorithm: Steepest descent and random walk are simple but rather limited search algorithms. As a result, they may be unable to find the optimal maximum and minimum security arrangement for a state. It has been suggested that the genetic algorithm is useful in search, especially in complicated landscapes. *EUGene* implements a version of the genetic algorithm, modified for our purpose of finding optimal security arrangements in a large space in a reasonable amount of time. The theory and terminology of genetic algorithms can be researched in sources such as Goldberg (1989) and Holland (1975). In general, the implementation is:

1. Initialize Population.
 2. Evaluate fitness of all in population.
 3. Identify best fitness seen so far
 4. Repeat (one generation)
 - Select from old population by random selection weighted by fitness.
 - Crossover those selected from old population to form new population
 - Mutate the new population according to user specified probability of mutation.
 - Evaluate new fitness.
 - If a better alliance (in terms of security) than previously seen is found, then reset a variable marking the number of generations that have been stable to 0. Otherwise, increment a counter for generations stable.
- Until termination conditions are met.
5. From the best pattern found by the genetic algorithm, a single "random walk" optimization iteration is run, taking a quick look for a local optimum near the pattern identified by the genetic algorithm. If it finds a better optimum, this is returned as the best or worst alliance pattern found. (This last step makes this a hybrid algorithm.)

Details:

Population:

First iteration: Initially, one member of the population is set to the reference state's current (existing) alliance configuration. Including this configuration as a member of the population ensures that the algorithm will return a found pattern at least as good as the actual alliance pattern.

Cloning: If the user so specifies, a subset of the best patterns found in the previous generation is kept ("cloned") into the new population. If at least one pattern is cloned, this ensures that each successive generation will find at least as good a security level as previous generations. However, the

more cloning, the faster the algorithm will converge (perhaps prematurely) and the less search area will be explored.

Termination conditions: The search for better patterns is terminated when two conditions are met. First, the largest change in security between the best alliance pattern in the previous generation and the best found in the current generation must be less than the specified risk search tolerance. Second, a particular number of generations must pass without any additional security improvement.

Mutation: With a user-specified probability, any single alliance value (bit) in each member of the population (except for cloned alliance members, and the actual alliance pattern) could change to a random alliance type.

Crossover: New alliance patterns in each generation are formed by applying single-point crossover to previous alliance patterns. A random crossover point is selected, and the second part of alliance string 2 is concatenated onto the first part of alliance string 1, while the second part of alliance string 1 is concatenated onto the first part of alliance string 2.

Sub-menu: Several options relating to the genetic algorithm must be set by the user. When *Genetic Algorithm* is selected, a pop-up window appears requiring the user to specify several options:

Population Size: Size of population (number of hypothetical alliance patterns) used in each generation of the genetic algorithm search.

Mutation Probability: Probability of mutation, per individual alliance per alliance pattern per generation.

Stable Iterations for Convergence: The number of iterations that must be observed without a change in the security level of the optimal pattern before exiting the genetic routine.

Number of Top Patterns to Maintain: The number of top "cloned" pattern to be maintained in the new population without mutation or crossover from one generation to another. *We recommend keeping this setting at 3 or higher for efficiency in search.* Internally, *EUGene* keeps track from year to year of the best/worst alliance patterns it finds for each country. Externally, *EUGene* has as an optional input file a saved file of the best/worst alliance patterns found during previous generation of risk scores. If you set the number of patterns to maintain at 3 or higher, *EUGene* will keep in memory these starting patterns from prior years/runs, rather than starting with random patterns. Presumably, starting with at least some information on what alliances looked optimal previously will help speed subsequent optimization, and allow the routine to pick up "where it left off" in prior searches. After the first search loop, the prior alliances found will be replaced internally by any better alliances, but initially a setting of 3 or higher is necessary to use these previously found alliance patterns.

Search Tolerance: The tolerance of the search - the number of iterations specified must pass with a change in optimal security level less than this amount before convergence will be declared and the algorithm terminated.

Risk Data Source

When computing expected utility following the methods in *War and Reason*, the user can choose to use either newly recalculated (and updated) risk attitude scores, or can use the actual risk scores computed by Bueno de Mesquita. Both choices use risk scores following the methods in Bueno de Mesquita (1985).

Software Verification

Differences from Existing Expected Utility Data

There are several individually minor differences in the way we have implemented expected utility calculations compared to Bueno de Mesquita and Lalman (1992) and prior publications (Bueno de Mesquita 1981, 1985), and in the data that we use, that could lead to slightly different computed values for elements such as risk scores, the utility of game outcomes, and possibly even equilibria, than previously calculated.

- We use updated COW Alliance and national capability data.
- We make one modification to the national capabilities data: because data for British energy consumption was often missing between 1816 and 1850, namely we interpolated data on British energy consumption when it was missing for this period.
- We use the newest version (v2.1) of the COW MID data to identify disputes and dispute outcomes. The new MID data set contains many new disputes (in fact, about twice as many as in the original MID data). However, many of the original disputes have been recoded as well, with previously coded dispute participants now excluded from some disputes, and with outcome/hostility codings changed. The below table presents a cross-tabulation between the dispute escalation levels in the new and old MID data set for the 1414 state-disputes that were analyzed in *War and Reason* (707 disputes, with two participants in each dispute). While the outcome codings are highly correlated ($r=0.732$), there are important differences. Note in particular the many cases where disputes had been coded (old hostility level > 0) that are now coded as non-MID dyads. These differences in turn lead to differences in the coding of outcome categories (acquiescence, status quo, war, capitulation, and negotiation). In this table, a value of “0” indicates that no militarized dispute occurred, while a value of “1” indicates no militarized response to an initiation by another state. Both imply that the state took no militarized action.

Cross-tabulation of Hostility Levels, Old versus New MIDs

		Hostility Level, New MID Data					Total	
		0	1	2	3	4		5
Hostility Level, War and Reason MID data	0	476	0	0	0	0	0	476
	1	35	86	2	12	12	7	154
	2	19	1	26	7	6	2	61
	3	60	15	4	111	13	6	209
	4	57	28	1	23	190	5	304
	5	35	3	2	0	10	160	210
Total		682	133	35	153	231	180	1414

- We add distance discounting to the methods of *War and Reason*. While *The War Trap* incorporated distance into its calculations, *War and Reason* did not, since the empirical analysis was purely within Europe, where all geographic distances are relatively small. 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 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 , and include states that are involved in the region of expected conflict for the directed dyad ij . We followed Bueno de Mesquita’s definition of the relevant region for conflict (1981:97).
- Our method for computing state-to-state distance has been generalized from that used in *The War Trap*, and made uniform. 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). 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 use the 1993 version of the COW contiguity data set for this computation.

- As detailed in the above section on generating risk scores, we use an improved algorithm to generate risk attitude scores following the risk definitions in Bueno de Mesquita (1985).
- Our dispute cases are somewhat different from those in *War and Reason*, both because of changes in the MID data as described above, but also because we exclude states that join disputes. For comparison to the distribution of MID relationships presented in Bueno de Mesquita and Lalman (1992:285), the next table presents a cross-tabulation of hostility levels between the initiator and target among the dispute dyads within our data. In most categories, our count of dyads with a given hostility combination exceeds that in the previous data set, as expected given our inclusion of all regions. However, in a few cells, we have less cases than did the data previously analyzed. A particular consequence of this change is that we have less “Big War” dyads in the data than did prior analysis.

Crosstabulation of Hostility Levels, Initiators vs. Initial Targets, All Dyads 1816-1984

		Initiator Hostility Level				Total
		2	3	4	5	
Target Hostility Level	1	111	322	599	3	1035
	2	3	23	25	0	51
	3	9	121	96	1	227
	4	4	62	509	10	585
	5	0	2	16	83	101
Total		127	530	1245	97	1999

Software Checks

We performed a number of checks on *EUGene* to verify that it was, in fact producing valid expected utility scores. This task is made difficult by the fact that there are a number of reasons to expect different scores than those originally generated by Bueno De Mesquita and Lalman as we mentioned above. We made comparisons at a number of stages. Our first comparison involved correlating the expected utility we constructed following the methods of *The War Trap* with the expected utility data printed in Tables A through C of the Appendix to that volume. We expected a substantial difference here because the numbers in those tables represent expected utility adjusted for risk attitude and uncertainty, while our calculation computed expected utility only through chapter 3, equation 6 of *The War Trap*, and does not adjust for risk attitude. Even so, the regression of our new data (with no risk adjustments) on the reported expected utilities (with risk adjustments) in cases of wars, interventions and threats gives an *r* of 0.78 and an adjusted R2 of 0.60 (n=241 dyads).

Another set of comparisons focused on our new risk attitude scores. We can make two comparisons. The first is to correlate directly our newly generated risk scores with those generated for Bueno de Mesquita (1985) up to 1970. For all of the reasons discussed above, we expect substantial differences between our risk scores, which were generated using better algorithms and updated data, and the earlier data. In addition, we note that the risk scores output by Tolstoy, which represent the full set of risk data generated as part of Bueno de Mesquita (1985) appear problematic in some cases, as they take values greater than 1. Theoretically, the risk

variable must range from -1 to +1. Bueno de Mesquita (1981:124) suggests that there are some calculation errors in the earlier data. The number of state-years in each region, and the correlation between our measure and the old data set, are as follows: Europe, n=3493, r=0.73; Asia, n=1437, r=0.51; Middle East, n=1765 r=0.73; Americas, n=2740, r=0.46 (the Americas region contains several illegal risk values greater than 1 as output by Tolstoy). These values are all highly significant.

Our second comparison is between risk attitude and victory in international crises. For this verification test we constructed a data set where the unit of analysis was the dyadic-dispute-initiation and examined the relationship between the dispute initiator's risk attitude scores the dispute outcome. From the rational choice perspective, one would expect that risk acceptant conflict initiators would lose the conflicts they initiate more often than risk averse initiators. We also expect to see that initiators win disputes and wars more often than they lose them, since potential initiators choose not to start conflicts that they are likely to lose. However, we should also see more of the losses among risk acceptant states.

For this check, we used v2.1 of the COW MID data set, and identified conflict initiators as those states on side A (the initiating side) who were involved on the first day of the dispute. We identified as their targets all states on side B who were also involved on the first day of the dispute, and then constructed all dyadic pairs of initiators by target. Outcomes were identified from the MID "outcome of dispute" variable, with "Victory for side A" and "Yield by side B" coded as an initiator victory. Cases of "Victory for side B" and "Yield by side A" we coded as initiator defeat. We dropped cases with indeterminate, stalemate, or compromise outcomes. We follow Bueno de Mesquita (1985:169) and examine this relationship for the combined set of war and intervention cases, selecting all cases where at least one side used force (level 4) or higher in the dispute (intervention occurs when one state uses military force in a crisis and the other state does not reciprocate). [While we only report this table, the relationship also holds among all disputes and among disputes that became wars.] We dichotomize the continuous measure of risk attitudes, splitting it at 0. Our analysis includes data from all regions of the globe, and we find 308 cases of war and intervention with clear outcomes to include, almost four times as many cases as analyzed this way in Bueno de Mesquita (1985). The following table presents the results.

Risk Attitude and Conflict Outcome, Interventions and Wars, All Regions

			Outcome		Total
			Initiator Lost or Yielded	Initiator Won or Target Yielded	
Initiator's Risk Attitude	Risk Averse	Count	47	126	173
		% Initiators with Outcome	27.2%	72.8%	100.0%
	Risk Acceptant	Count	51	84	135
		% Initiators with Outcome	37.8%	62.2%	100.0%
Total		Count	98	210	308
		% Initiators with Outcome	31.8%	68.2%	100.0%

As seen in this table, risk-averse conflict initiators do win their conflicts more often than risk-acceptant initiators. The difference in this table is statistically significant at approximately the 0.04 level. This table understates the relationship between risk attitude and victory because it loses significant amounts of information by dichotomizing actors into risk-averse and acceptant. When we conducted a logit analysis of the continuous version of initiator's risk attitude on initiator victory (which takes into account the magnitude rather than just direction of states' risk-attitude) we find that risk attitude is actually statistically significant at <0.001. In this logit, the

coefficient on the constant had a value of 0.74 with standard error of 0.13, while the coefficient on risk acceptance was -1.14 with a standard error of 0.31.

We made a final comparison with Bueno de Mesquita and Lalman's data in which we correlated our newly generated estimates of the initiator's utility for the status quo and acquiescence with those utilities in the *War and Reason* data set. The correlation coefficients (r) between our data and the War and Reason data (707 dyads) are: 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 $U^B(\Delta_A)$. While our new measures have additional information in them, they do still correlate in large part with prior data in Europe.

To test our core rules for generating equilibrium predictions from input preferences, we also applied the complete set of possible preference orderings to the core logical algorithms for equilibria (the rules listed above under "Equilibria") and replicated Bueno de Mesquita and Lalman's Table 2.4, a table showing the distribution of outcomes with complete information and uniformly distributed preferences. We use this test to verify that our rules implementing either backwards induction or the logical rule specification do fit the international interaction game.

Data Set Creation: Additional Specifications and Details

Missing Values

For most variables, a "-9" in the output file the *EUGene* creates indicates that the value is missing. The exception is for the set of polity variables, "dem1" "dem2" "demlg1" "demlg2" "demchg1" "demchg2", etc. (all polity variables as listed below) for which the values -66 through -99 are missing (-9 is a valid value for these variables; -66 through -88 represent various types of polity interruptions). It is incumbent on the user to ensure that when data is read for subsequent analysis, -9 and -66, -77, -88, and -99 are coded as missing for the appropriate variables. Output data will be missing when input data are missing or a value cannot be otherwise calculated. For example, expected utility data will be missing when national capability data are missing or when alliance data are missing (since alliances are necessary for Tau, S, and risk scores). Similarly, lagged values of polity types will be missing when no data is available in the lagged year.

The Stata, SPSS, and LIMDEP command files that *EUGene* creates to read the data contains commands to convert missing values appropriately.

If there is no MID or ICB crisis in a given year, these dyads will have values of "0" in the output for most conflict variables, NOT missing values.

Population of Cases

Years to Include in Output

Note that specifying a year under the output options does not affect the calculations that the *EUGene* program performs. It only affects the output that *EUGene* provides.

All Years

Information will be printed for all years possible in the analysis, that is, any year for which at least some data is available. At the time of this writing, this is 1816 to 1993.

Specified Years

Allows you to enter a subset of years. Only years from the first year to last year (inclusive) will be printed.

Country-Year Selection

All States

The output will include every country that is a state in a given year.

All Major Powers

The output will include every country that is listed as a great power according to Singer and Small specifications in a given year.

All States within Specified Regions

A window will appear requiring the user to specify a set of regions for output. The output will include all states within the specified region(s).

Specific Set of Countries

A window will appear allowing the user to define a subset of states that are to be included in the output. Initially, a list of countries will appear in the left pane of the window, while the right pane will be blank. From this window the user may select a subset of countries to be included in the output by highlighting a country name and pressing the ">" button to move the highlighted country to the list of selected countries. The "<" button can be pressed to move a country out of the selected list. The ">>" and "<<" buttons will move all countries. Note that *EUGene* calculations are unaffected by this selection, e.g. calculations involving all of the states involved in a region will still include all regional states.

Directed Dyad-Year Selection

As an example, *EUGene* will create output records both for the US vs. USSR dyad in 1946 and the USSR vs. US dyad in 1946. This is appropriate given *EUGene*'s focus on dispute initiation, which is directed, rather than just dispute onset, which is not.

All Dyads from All Countries

The output will include all dyad-years in which both members were states as defined by Singer and Small.

All Major Power vs. Major Power Dyads

The output will include all dyad-years in which both members were Major Powers as defined by Singer and Small.

All Major Power vs. Any State Dyads

The output will include all dyads where at least one of the states is classified as a great power according to the Singer and Small specifications.

All Contiguous Dyads

A window will appear requiring the user to specify a level of direct contiguity following the COW five classifications of contiguity (contiguous on land, or separated by up to 12, 13-24, 25-150, or 151-400 miles of water). The output will include all dyads where the two states are contiguous by that classification rule.

All Politically Relevant Dyads

The output will include all of the dyads that are classified as "politically relevant" following the general rules given by Maoz and Russett (1993). Politically relevant dyads are defined as all contiguous dyads plus all dyads in which at least one of the states is a Great Power. You may check or change the degree of direct contiguity required (e.g. contiguous on land, separated by up to 12 miles of water, etc.) with the "Change Contiguity" button next to the

politically relevant button option. Note that setting contiguity for the population of cases will also affect contiguity for outputting the politically relevant variable in the “Variables” tab.

All Dyads within Selected Regions

A window will appear requiring the user to specify a set of regions for output. The output will include all dyads where both members are within one of the specified region(s).

All Dyads within Maximum Distance

The output will include all dyad-years in which the intercapital distance between both members is less than the specified value in miles. The user may specify the limiting distance by entering it in the text box to the right of the “Within Maximum Distance” button.

Specific Set of Dyads

A window will appear allowing the user to define a subset of states that are to be included in the output. All dyad-years involving these states and where both members are states (following Singer and Small) in a given year will be printed. That is, if A, B, and C are selected, dyads A-B, A-C, and B-C will be printed for all years where both members are states.

Dyads Read from User File

The output will include all of the directed dyads and years that are in a user specified text file. A sample input file is provided in the directory “inputdat” under the name “userdyad.txt”. To specify the file name, press the “Change Input File” button. Note that if you also specify a sub-range of years, then only those years that are BOTH listed in the user dyad file AND contained within the sub-range will be output. Also note that dyads entered in this file are treated as **directed**, so that if you enter ccode 2 vs. ccode 365, you will get only the directed of 2 vs. 365. If you want the reverse direction as well, enter 365 vs. 2 also.

Note for user chosen dyads: What is in the file (userdyad.txt by default) are treated as directed dyads. If you want both directions, put both directions. If you have requested non-directed dyad output, the ccode1 and ccode2 will appear in the order you put them in the userdyad.txt file. Also, if you happen to list 220 vs. 200 and 200 vs. 220 (e.g.), then both of these will appear in the output data set, although the variables output will be the correct non-directed output. Similarly, if you list 220 vs. 200, 200 vs. 220, and 200 vs. 220, you will get multiple output lines in the output file with the same information. Users are responsible for making sure the dyads in the data set work!

Non-directed Dyad-Year Selection

As an example, *EUGene* will create output records for the US vs. USSR dyad in 1946, representing the dyad of countries but not the direction of the dyads in terms of who first crossed the MID threshold. This is appropriate when dispute onset is under study rather than dispute initiation.

All Dyads from All Countries

The output will include all dyad-years in which both members were states as defined by Singer and Small.

All Major Power vs. Major Power Dyads

The output will include all dyad-years in which both members were Major Powers as defined by Singer and Small.

All Major Power vs. Any State Dyads

The output will include all dyads where at least one of the states is classified as a great power according to the Singer and Small specifications.

All Contiguous Dyads

The output will include all dyads where the two states are contiguous. You may check or change the degree of direct contiguity required (e.g. contiguous on land, separated by up to 12 miles of water, etc.) with the “Change Contiguity” button next to the politically relevant button option. The level of contiguity follows the COW five classifications of contiguity (contiguous on land, or separated by up to 12, 13-24, 25-150, or 151-400 miles of water). The output will include all dyads where the two states are contiguous by that classification rule.

All Politically Relevant Dyads

The output will include all of the dyads that are classified as “politically relevant” following the general rules given by Maoz and Russett (1993). Politically relevant dyads are defined as all contiguous dyads plus all dyads in which at least one of the states is a Great Power. You may check or change the degree of contiguity required (e.g. contiguous on land, separated by up to 12 miles of water, etc.) with the “Change Contiguity” button next to the politically relevant button option. Note that setting contiguity for the population of cases will also affect contiguity for outputting the politically relevant variable in the “Variables” tab.

All Dyads within Specified Regions

A window will appear requiring the user to specify a set of regions for output. The output will include all dyads where both members are within one of the specified regions.

All Dyads within Maximum Distance

The output will include all dyad-years in which the intercapital distance between both members is less than a specified value in miles. The user may specify the limiting distance by entering it in the text box to the right of the “Within Maximum Distance” button.

Specific Set of Dyads

A window will appear allowing the user to define a subset of states that are to be included in the output. All dyad-years involving these states and where both members are states (following Singer and Small) in a given year will be printed. That is, if A, B, and C are selected, dyads A-B, A-C, and B-C will be printed for all years where both members are states.

Dyads Read from User File

The output will include all of the directed dyads and years that are in a user specified text file. A sample input file is provided in the directory “inputdat” under the name “userdyad.txt”. To specify the file name, press the “Change Input File” button. Note that if you also specify a sub-range of years, then only those years that are BOTH listed in the user dyad file AND contained within the sub-range will be output. Also note that dyads entered in this file are treated as **non-directed**.

Note for user chosen dyads: What is in the file (userdyad.txt by default) are treated as directed dyads. If you want both directions, put both directions. If you have requested non-directed dyad output, the ccode1 and ccode2 will appear in the order you put them in the userdyad.txt file. Also, if you happen to list 220 vs. 200 and 200 vs. 220 (e.g.), then both of these will appear in the output data set, although the variables output will be the correct non-directed output. Similarly, if you list 220 vs. 200, 200 vs. 220, and 200 vs. 220, you will get multiple output lines in the output file with the same information. Users are responsible for making sure the dyads in the data set work!

Directed Dispute Data Selection

This unit of analysis reports information about cases of dyadic dispute initiation (not all disputes, but dyadic *initiations* only; the resulting dyads are hence directed). As an example, if the US initiates a dispute in 1946 vs. the USSR, then *EUGene* will create an output record for the US vs. USSR dyad in 1946. If the USSR also initiates a different dispute vs. the US in 1946, then another record will be output for USSR vs. US in 1946. If the USSR did not initiate a dispute against the US in 1946, then that directed-dyad-year would not be output, as there was no dispute initiation in it.

Users may select either the COW computed version of MID data or the Maoz version of the MID data to define these cases. You may select one version of the MID data to define the base data set/conflict exclusions, and also request variables from the other version of the MID data under the “Conflict Variables” tabs. However, cases will still be output based on the base data set selection, and variables about disputes from the 2nd data set will only be output if there is a directed dispute initiation within the second set as well as the first.

One Case per Directed Dispute Dyad Initiation

Output will contain one line of data per dyadic-dispute initiation.

Two key options can be set on the Variables | Conflict Data | Initiator/Multiple MID Settings tab. If the first option (“Code Only Originators (Involved Day 1) as Initiators”) is selected, then only originators (states involved on day 1) are coded as initiators. Under this setting (and specifying that states on Side A are initiators), output will contain one line for each originator state on Side A vs. each originator state on Side B. Only originators are considered and included in any dyad, as originator or target. If revisionist states are considered initiators, then there will be a line of data between each revisionist originator and each originator on the opposite side of the dispute.

Under the second option (“Code Initiating-Side Joiners as Initiators, and Target-Side Joiners as Targets”), all states who are joiners on the initiating side will be considered to be dispute initiators, and all joiners on the target side will be considered to be targets of initiation. Thus, a dispute-initiation-dyad will be created between each originator and joiner on Side A vs. each state (originator or joiner) on the opposite side (or, if revisionist states are selected as initiators rather than states on Side A, then revisionist states will be considered as initiating). This option creates all dyads of states on side A vs. all on side B (subject to overlapping dates of involvement). Note that there is *not* an option to consider the joiners on side A as initiators, but to not create dyads against joiner targets. This is because there is not an adequate way in the MID data (through 1992) to distinguish between “voluntary joiners” who start conflicts and “involuntary joiners” who are targeted. It would not be valid to draw a distinction between joiners on side A and side B in terms of going first in the dispute, as we cannot guarantee the order of actions involving joiners. If revisionists are specified as initiators, then there will be a line of data between each revisionist-side originator or joiner vs. each state on the opposite side of the dispute.

EUGene used to include a third option that specified that all originators and all joiners were to be considered to be initiators. This setting basically destroyed all meaning to the “initiate” variable, as it no longer distinguishes the identity of joiners vs. originators or the side that they are on as initiators or targets, and has been removed.

Note that if you specify originators and joiners to be considered as initiators under any options, you must also specify to include joiners in the output on the “Base Data Set/Conflict Exclusions” tab, or they will be dropped.

One Case per Directed Dispute Dyad Year

Output will contain one line of data for each dyad-year of each directed-dyadic-dispute initiation. That is, there will be one line for the initiation year, and one additional line for each year that the dispute continues, for all directed dyadic dispute initiations.

Non-Directed Dispute Data Selection

This unit of analysis reports information about cases of dyadic dispute onsets. Output is non-directed; the initiator of a dispute is not considered in this unit of analysis, but only the start of a dispute. As an example, if the US and USSR have a dispute in 1946, regardless of who started it (who was on side A, or who was revisionist), then *EUGene* will create an output record for the US vs. USSR dyad in 1946. If there is a second dispute between the US and USSR in 1946, then a second dispute record will be output for that second dispute in 1946. The order of the states in the dyad are arbitrary (there is no inherent direction in the dispute), so the state with the smaller country ccode is listed first (as ccode1).

Users may select either the COW computed version of MID data or the Maoz version of the MID data to define these cases. You may select one version of the MID data to define the base data set/conflict exclusions, and also request variables from the other version of the MID data under the “Conflict Variables” tabs. However, cases will still be output based on the base data set selection, and variables about disputes from the 2nd data set will only be output if there is a directed dispute initiation within the second set as well as the first.

One Case per Dispute Dyad Onset

Output will contain one line of data per dyadic-dispute onset.

Two key options can be set on the Variables | Conflict Data | Originator/Joiner Settings tab. If the first option (“Create Dyads only for Originators”) is selected, then only onsets between originators (states involved on day 1) will be included in the output. Output will thus one line for each originator state on Side A vs. each originator state on Side B, with the dyads ordered in ccode order.

The second option (“Create Dyads for Originators and Joiners”) specifies that dyads involving both originators and targets in a dispute will be listed in output. Under this option, a dispute--dyad will be created between each originator and joiner on Side A vs. each state (originator or joiner) on the opposite side (subject to overlapping dates of involvement). This option creates all dyads of states on side A vs. all on side B.

Note that if you specify originators and joiners to be included, then you must also specify to include joiners in the output on the “Base Data Set/Conflict Exclusions” tab, or they will be dropped.

One Case per Dispute Dyad Year

Output will contain one line of data for each dyad-year of each dyadic-dispute. That is, there will be one line for the initial onset year, and one additional line for each year that the dispute continues if it lasts past January 31 of the given year.

Variables

Some variables must be selected or an error will be reported. Some variables can only be printed if dyadic output is selected.

Variables Available

“General” Variable Tab

CCode: Output will list the COW country code for the individual country (if country-year output is selected) or for the two members of the dyad (if dyad-year output is selected) for whom the rest of the data correspond to. This variable is required in the output.

Abbreviations: Output will list the 3 letter COW country abbreviation for the individual country (if country-year output is selected) or for the two members of the dyad (if dyad-year output is selected) for whom the rest of the data correspond to.

Year: Output will list the year that the rest of the data corresponds to. This variable is required in the output.

Capabilities: Output will include the values of the COW Composite Indicator of National Capabilities (CINC) index, and/or index's component variables. The six capability country-year variables are listed in the following order: Military Personnel (milper), Military Expenditures (milex), Energy Production (energy), Iron/Steel Production (irst), Nominal Urban Population (upop), and Nominal Total Population (tpop).

Major Power Status: Output will include a "1" if the country is a major power in the given year, or a "0" if it is not.

Dyad Duration (Available for Dyad-Year and Dispute-Dyad Output Only): Output will include an integer representing the number of years that both members of the dyad have been states continuously since 1816. For example, Britain and France receive a "0" in 1816, a "1" in 1817, etc. Canada becomes a state in 1920, and so the US-Canada dyad would have missing values to 1920, a 0 in 1920, a 1 in 1921, etc. This counter resets when a state drops out of the state system by COW criteria. So the France-West Germany duration variable starts at 0 in 1955. [Note: this was modified in v1.19. Before v1.19, the duration counter counted the number of years from the first time the two states became states, and did not reset even when states left the system.

Relevant Region (Available for Dyad-Year and Dispute-Dyad Output Only): Output will include an integer marking the region that is relevant for the computation of expected utility. Regional memberships are defined as given by the COW Interstate System Members list. Integers correspond to regions as follows: Europe=1; Middle East=2; Africa=3; Asia=4; North and South America=5.

Home Region: Output will include an integer marking the region that the country (if country-year output is selected) or countries (if dyad-year output is selected) is in. Regional memberships are defined as given by the COW Interstate System Members list. Integers correspond to regions as follows: Europe=1; Middle East=2; Africa=3; Asia=4; North and South America=5.

Politically Relevant (Available for Dyad-Year and Dispute-Dyad Output Only): Output will include a dummy variable marking the cases of politically relevant dyads. Politically relevant dyads are those where at least one state is a major power, or the states are contiguous. The user may change the degree of contiguity required with the "Change Contiguity" button under the variable option. Note that setting contiguity for variable output will also affect contiguity for outputting politically relevant dyads only, if you select politically relevant dyads for your output population. A '1' marks a politically relevant dyad, while a '0' marks a non-politically relevant dyad.

Direct Contiguity Level (Available for Dyad-Year and Dispute-Dyad Output Only): Output will include a variable marking the level of contiguity between the two states. Levels range from 1 to 6. The first five levels follow the COW specifications for the types of contiguity: 1) land contiguity; 2) contiguous for up to 12 miles of water; 3) contiguous for 13-24 miles of water; 4) contiguous for 25-150 miles of water; 5) contiguous for 151-400 miles of water. Level 6 indicates that the states are not contiguous (or are contiguous over more than 400 miles of water).

Colonial Contiguity Level (Available for Dyad-Year and Dispute-Dyad Output Only): Output will include 3 variables to mark the closest level of contiguity between the two states while taking into account colonial possessions. Levels range from 1 to 6. The

relationship is assessed between states A and B looking at several relationships: A with colonies of B, B with colonies of A, and colonies of A with colonies of B. The first five levels follow the COW specifications for the types of contiguity between 2 of these entities: 1) land contiguity; 2) contiguous for up to 12 miles of water; 3) contiguous for 13-24 miles of water; 4) contiguous for 25-150 miles of water; 5) contiguous for 151-400 miles of water. Level 6 indicates that the states are not contiguous through any colonial relationship (or are contiguous through colonies over more than 400 miles of water). Three variables are output. The first is the level of colonial contiguity. The 2nd and 3rd give state/entity codes for the relevant colonies of state A and B, if any. If the contiguity level emerges from a relationship between 2 colonies, then both the 2nd and 3rd variables will have a state/entity code. If the case has contiguity between a state and a colony of the other, only one of these entries will have a code, while the other will be 0. For instance, output of "1 2000 2001" indicates that states A and B had level 1 contiguity between their colonies 2000 and 2001. Output of "3 0 2001" would indicate level 3 contiguity between state A and entity 2001, which is a colony of B. Note that only the closest colonially-contiguous relationship is reported. The system membership numbers of colonies can be found in the input data file "States and Entities 1492.dat" in the "INPUTDAT" subdirectory under *EUGene*.

Distance Between States (Available for Dyad-Year and Dispute-Dyad Output Only):

This outputs the distance between states in the dyad as calculated by the program. Note: *EUGene* will compute distance based on the method in its memory, which by default is the distance between capitals, adjusted for contiguity and allowing multiple cities for each country following *The War Trap*. This default option may be changed either under the recalculation options in the "Recompute" menu item, or by pushing the "Distance Options" button on the variable screen.

System Characteristics: Output will include the selected variables as computed by *EUGene*. These variables include the number of states in the system (as defined by COW specifications), the number of major powers in the system, yearly system concentration (CON), and system movement (MOVE). System movement is available computed across the entire system and the great power subsystem, and calculated both yearly and as a five-year moving average. CON and MOVE scores are defined in Singer, Bremer, and Stuckey (1972).

ISO Data: Outputs a selection of ISO identifier variables for the given country. Uses the ISO_3166-1 dataset. The numeric ISO Code, 2 and 3 letter abbreviations, short name, and full name (all in English) are available for output.

"Polity III" Variable Tab

Polity III Data: Output will include selected variables from the Jagers and Gurr (1995) Polity III data set. These variables are democ, autoc, xrreg, xrcomp, xropen, mono, xconst, parreg, parcomp, and cent. In addition, the derived variable "dem" used by Russett and others is available (Dem = Democ - Autoc). In addition, lagged versions of the democ, autoc, and dem variables are available, along with democratization computed as demchg = dem - lag(dem). Select a specific subset of Polity III variables by pressing the "Variable Selection" button.

NOTE: We hope to have Polity IV data available as a user data set in the Fall of 2003.

"Alliance" Variable Tab

Alliance Type (Available for Dyad-Year and Dispute-Dyad Output Only): Output will contain the COW alliance type. As of version 3 of *EUGene*, the "Dyadic Alliance File" uses the COW v3.0 alliance data released in December, 2002. Options to use other versions of the data have been removed. The codes for various types of alliance are 1=defense pact, 2=neutrality, 3=entente, 4=no agreement.

Alliance Portfolio - unweighted (Available for Dyad-Year and Dispute-Dyad Output Only): Output variables are the cells of a 4x4 matrix representing the alliance portfolio between the two states, as seen initially in Bueno de Mesquita 1978 and 1980. Two separate matrices are output, one with membership including all states in the relevant region of conflict (corresponding to regional tau), and the second with membership including all states in the system (corresponding to global tau). Cell values are counts of the number of states falling into each combination of alliance types, e.g. variable “rportu11” refers to unweighted portfolio (based on states in relevant region) [1,1], and counts the number of states with whom both ccode1 and ccode2 share defense pacts. Variable “rportu12” then refers to regional portfolio[1,2] (row 1, column 2), and so on. Row and column 2 refers to neutrality pacts, 3 refers to entente agreements, and 4 refers to states where there is no alliance. So “rportu24” is portfolio[2,4], the count of states where ccode1 has neutrality agreements and ccode2 has no alliance. Variables “gportu11” through “gportu44” refer to the unweighted portfolio using all states around the globe. A variable “rportn” (and correspondingly “gportn”) is also output, representing the total number of states counted in the unweighted alliance portfolio table.

Alliance Portfolio - weighted (Available for Dyad-Year and Dispute-Dyad Output Only): Output variables are the cells of two 4x4 matrices representing the alliance portfolio between the two states, but with modifications for the weighted version of S as developed by Signorino and Ritter (1999). Here, cell values represent the sum of COW CINC scores for the states falling into each combination of alliance types, e.g. variable “gportw11” refers to the sum of capabilities falling into weighted portfolio cell [1,1] in the global alliance portfolio version. A variable “dmax” is also output, representing the total sum of all capabilities in the weighted alliance portfolio table.

Tau-b Scores (Available for Dyad-Year and Dispute-Dyad Output Only): Output will include the tau-b between the two states. Both global tau-values (calculated using every state in the system) and regional tau-values (calculated using only states in the relevant region for the dyad) will be reported. In directed dyads, the output will always be from CCode1 to CCode2. In non-directed dyads, *EUGene* provides the output in both directions.

S Scores (Available for Dyad-Year and Dispute-Dyad Output Only): (See Signorino and Ritter 1999) Output will include the unweighted or weighted *S* between the two states. Weighted *S* scores are weighted using system capabilities. Both global *S*-values (calculated using every state in the system) and regional *S*-values (calculated using only states in the relevant region for the dyad) will be reported. In directed dyads, the output will always be from CCode1 to CCode2. In non-directed dyads, *EUGene* provides the output in both directions.

Tau or S with System Leader: Output will include the tau or *S* between the state in question (country-year unit) or states in question (dyad-year unit) with the system leader, which is Britain up to 1945, and the US from 1946 forward.

Options: User may set whether the tau or *S* computed with the system leader is based on the alliances of states only involved in the relevant region of the ccode vs. the system leader dyad (regional option), or is based on all states in the international system (global option). For *S*, the user may specify whether to use the weighted or unweighted *S* with the system leader.

“Expected Utility” Variable Tab

Expected Utility - War Trap (Available for Dyad-Year and Dispute-Dyad Output Only): Output will include the expected utility of country 1 vs. country 2, based on the operational rules specified in *The War Trap* as discussed in this documentation. The user must specify whether tau or *S* measurements is to be used. In directed dyads, the output

will always be from CCode1 to CCode2. In non-directed dyads, *EUGene* provides the output in both directions.

Risk Attitude – War Trap Revisited: Output will include the values from the risk attitude calculations performed by Bueno de Mesquita (1985), available in both tau and S based calculations.

Risk Attitude – War and Reason / EUGene: Output will include the values R_i from the risk attitude calculations performed by *EUGene*, following the methods in Bueno de Mesquita (1985) and Bueno de Mesquita and Lalman (1992). If the “Country-Year” unit of analysis is selected for output, the output will include risk scores for a given state in all regions. If the “Dyad-Year” unit of analysis is selected, risk scores will be reported for the two states in the dyad for the region that is relevant for their conflict and so to the calculation of expected utility.

Risk Details: This outputs the values from the intermediate components of the risk attitude calculation from *EUGene*. Specifically, this outputs actual, maximum hypothetical, and minimum hypothetical security values. If the “Dyad-Year” unit of analysis is selected, these values will be reported for the two states in the dyad for the relevant region of their conflict and expected utility calculations. If the “Country-Year” unit of analysis is selected for output, the output will include detailed intermediate information on the risk scores for the given state in all regions. Values are from -1 to $+1$, with -1 indicating a highly risk-averse actor, and a $+1$ indicating a highly risk-acceptant actor.

Uncertainty: Output will include regional uncertainty as defined by Bueno de Mesquita and Lalman (1992), available in both Tau-b and S versions.

Utility - War and Reason (Available for Dyad-Year and Dispute-Dyad Output Only): Output will include various values related to the expected utility of country 1 vs. country 2, based upon the measures developed in War and Reason. A number of values are output:

From Bueno de Mesquita and Lalman (1992:293-294): $U^A(\Delta_A)$, $U^A(\Delta_B)$, $U^A(SQ)$, $U^B(\Delta_A)$, $U^B(\Delta_B)$, $U^B(SQ)$, $Stakes_A$, $Stakes_B$.

From Bueno de Mesquita and Lalman (1992:297): P^A , P^B .

From Bueno de Mesquita and Lalman (1992:47): $U^i(SQ)$, $U^j(SQ)$, $U^i(Acq_i)$, $U^i(Acq_j)$, $U^j(Acq_i)$, $U^j(Acq_j)$, $U^i(Nego)$, $U^j(Nego)$, $U^i(Cap_i)$, $U^i(Cap_j)$, $U^j(Cap_i)$, $U^j(Cap_j)$, $U^i(War_i)$, $U^i(War_j)$, $U^j(War_i)$, and $U^j(War_j)$

State i corresponds to state A which is the first state in the dyad A vs. B, while state j corresponds to state B. In directed dyads, the output will always be from CCode1 to CCode2. In non-directed dyads, *EUGene* provides the output in both directions.

The user must specify whether tau or S is to be used.

Equilibria - War and Reason (Available for Dyad-Year and Dispute-Dyad Output Only): Output will include dummy variables marking what equilibrium in the international interaction game is expected given the various utility values in the dyad. In directed dyads, the output will always be from CCode1 to CCode2. In non-directed dyads, *EUGene* provides the output in both directions. The user must specify whether tau or S is to be used. Please note the differences between the particular variables included in the equilibria output, especially in non-directed dyadic output. For example, “eqTcpa12” indicates capitulation by A (or, ccode1) in equilibrium and “eqTcpb12” indicates capitulation by B (ccode2) in equilibrium when the directed dyad ccode1 vs. ccode2 is examined in the non-directed dyadic output. However, “eqTcpa21” represents capitulation by A (ccode2 this time) in equilibrium, as the directed dyad ccode2 vs. ccode1 is also examined in the non-directed dyad output. The first number will always

be Bueno de Mesquita and Lalman's state A, and the second their state B, in the non-directed dyadic output.

Options: User may set whether to generate the equilibrium for a dyad-year by either 1) using the logical conditions given in *War and Reason*, or by using backwards induction using the computed utility values for each dyad-year. For more details about this choice, see the discussion under section "Equilibria (War and Reason)" on page 21.

"Conflict Data" Variable Tab

Under this tab, you may select any (or all) of a variety of conflict-related variables to be included in the output data set. Starting in version 3.0, you need not select just one conflict set for your output.

If you wish to make exclude or include dyad-years based on the options on the "Exclusions" and "Dispute Initiator" output tabs, you will need to select some conflict data for output.

COW Dyadic MID Data

(Note: Available for Dyad-Year and Dispute-Dyad Output Only): Output will include a *EUGene*-created dyadic version of the COW MID dispute data, combined with COW v3.0 dyadic MID data. From 1816-1992, this data consists of a conversion within *EUGene* of the original non-dyadic COW MID data into a dyadic form (additional details of converting the COW MID data into dyadic form is given in the next section, beginning on page 62). From 1992 forward, this data consists of the COW v3.0 dyadic MID data. This part of the series uses incident-level data to compute actual confrontational dyads. See additional discussion in the section "Creating Dyadic MIDs and Meshing MID Data Sets."

Maoz Dyadic MID Data

(Note: Available for Dyad-Year and Dispute-Dyad Output Only): Output will include Maoz's version of the dyadic MID dispute data from 1816-1992 coupled with the COW v3.0 dyadic MID data from 1992 forward. Maoz's v1.1 dispute data examined multi-party disputes to determine what states are actually in dyadic confrontation. From 1816-1992, Maoz's data set data is used to create the requested variables. From 1992-2001, the series consists of the COW v3.0 dyadic MID data. This part of the series uses incident-level data to compute actual confrontational dyads. In cases of overlap, codings from the COW v3.0 data set supercede Maoz's codings. See additional discussion in the section "Creating Dyadic MIDs and Meshing MID Data Sets."

MID Variables Always Reported:

EUGene will always output several items, including whether or not there is a new or ongoing conflict, the MID numbers of these conflicts (if any), and the hostility level of the states in these MIDs:

1 and 2) Ongoing MIDs: Whether there was a MID ongoing involving ccode1 and ccode2 on January 1 of the year. The MID number is output along with a 0/1 coding for ongoing. Note that in the case of dyadic MIDs created because a state exited and reentered the MID, each re-entry is a new MID, and is not considered ongoing.

3 and 4) Initiation/Onset, and number of MID initiated or begun in year: Whether state A initiated a MID vs. state B in this year (directed-dyad output) or whether there was a MID onset between A and B in this year (nondirected-dyad output). Whether or not A is considered to have initiated a dispute depends on user settings for a) marking subsequent years as initiations, b) marking either side A or revisionists as initiators, and c) marking initiators as only originators or including joiners. Note that initiation marks specifically whether A initiated vs. state B. This variable is directed. The "onset" variable marks the

nondirected start of a MID. The MID number is output along with a 0/1 coding for ongoing.

5 and 6) Hostility levels in key MID (defined below): the highest hostility level reached by state A vs. state B in this year in the key MID, and the highest hostility level reached by B vs. A in the key MID in this year. “Relevant” is specified to determine the hostility level when there are multiple disputes involving A and B in a given year.

7) The COW MID number of the key MID (if any) between A and B in this year.

Other Available MID Variables:

A series of other MID variables are available for the user to select into the *EUGene* generated data set. Most of these variables are available from both the COW-based and Maoz-based dispute sets.

- 1) Name of the MID identified as the key MID in a year;
- 2) A dummy indicator of whether a MID is reciprocated;
- 3) A count of the number of states involved in each side of a MID;
- 4) Fatality levels for the overall dispute, and each country;
- 5) Highest activity and hostility levels for the overall dispute, and for each country;
- 6) Variables containing the MID outcome and settlement codes;
- 7) A count of the number of MIDs in a dyad-year, both total and new MIDs;
- 8) Peace Years values computed from each MID (and ICB) data set;
- 9) Start day, start month, start year, end day, end month, and end year for each country in the MID;
- 10) Dummy variables indicating whether CCode1 or CCode2 are MID originators.
- 11) MID side variables, indicating whether a state is on the initiator or target side of the MID;
- 12) Dummy variables indicating whether a state is revisionist;
- 13) Revisionist type variables;
- 14) Markers for joiner states (states becoming involved after day 1 of a MID);
- 15) The role of each state in the MID (e.g. joiner on target side,

Mark Joiners

Checking “Mark joiners” will output several variables concerning joining. In directed dyad output, variables *cwjomidi* and *cwjomidt* mark whether, if this is a joining dyad (where at least one state is a joiner in the key MID), state A was on the initiating side or target side against B in either the key dispute. That is, this will mark dyads where *cocode1* was on the initiating side against *cocode2*, but one of the two states was not an originator (was not involved on day 1), or where A was on the target side against B but again where one state was not an originator. Variables *cwjoanyi* and *cwjoanyt* mark whether state A was on the initiating (or target) side against state B in the given year and where at least one was a joiner, in any MID in the given year. In nondirected output, *cwjoany* marks whether EITHER A or B is a joiner in ANY MID in the given year, while *cwjomid* marks whether EITHER A or B is a joiner in the key identified MID in the given year. Similar variables (e.g. *mzjomidi*) are output for the Maoz dyadic version of the MID data set.

In marking joiners, rules for coding initiators as Side A or Revisionist apply; a Joiner will be coded only if it is a state on Side A (if Side A is marked to be the initiator) or the

revisionist side (if revisionists are specified as the initiator). See also the detailed discussion of issues in coding joining below, in section “Joiner Variables.”

Also note that if states are to be marked as joiners, then they are marked as such in subsequent years of the same dispute whether or not the user has specified wanting subsequent years coded as an initiation. That is, the variable marking joining is unaffected by the setting on coding subsequent initiation.

ICB Crisis Data

Variables from the Dyadic ICB Crisis data set may be selected under this tab. Variables are listed below, in the complete list of variables output by *EUGene*, and are explained in the ICB dyadic crisis data set documentation (the documentation is located in the “inputdat” subdirectory under the main directory where *EUGene* was installed, or the entire data set and documentation are available for download from ICB website at <http://www.icbnet.org/Data/>).

Because ICB data does not contain an indicator of how serious ICB crises are, there is no option to select the most serious ICB crisis in a year. For ICB data, information about the first crisis of the year is always reported. More precisely, if there is a new crisis in a year, information about the first new crisis of a year is output. If there are no new crises but there is an ongoing crisis, then information about the ongoing crisis is output. Finally, if there is neither a new nor ongoing crisis in a dyad-year, then the output consists of 0s.

Joiners into ICB crises are determined as any states whose dyadic trigger dates are later than the first dyadic trigger date. Individual trigger dates are not used, but just the dyadic trigger dates. See the ICB documentation for fuller description/definition.

Peace Years

(Available for Dyad-Year and Dispute-Dyad Output Only): Outputs a variable counting the number of years since the end of the last dispute (of any type – whether states were originators, joiners, on Side A or revisionist) in the dyad to use in creating Beck, Katz, and Tucker’s (1997) spline variables to account for serial autocorrelation. This variable normally takes a value of 0 for all dyads in 1816, and for the first year a country is considered a state (but see the option to select the Werner peace years adjustment below). It increments by one for each year a dyad goes without a dyadic dispute. In the first year that a MID occurs, Peace Years takes the regular incremented value. But in the next year, either the year after the MID if the MID lasts one year, or in the 2nd year of the MID if the MID lasts longer than a year, it resets to 0. As long as the MID continues, the variable takes the value 0, as it does in the first year after the MID ends. In the 2nd year after a MID’s conclusion, the peace years counting again increments, starting at 1. So, if a MID lasts from 1820 to 1825, the value of Peace Years would be 0 from 1821 through 1826, and a “1” in 1827 (a full year passed without a MID). Users should recognize that the peace years variable resets to 0 after a war, MID, or change in system membership for a state. For instance, peace years with Germany resets to 0 in 1990, because the reunified Germany is considered a distinct state from either East or West Germany.

Separate versions of the variable are computed for the COW-based, Maoz-based, and ICB based conflict data when those data are selected. That is, if the user requests the variable under the Maoz dyadic dispute data set section (for example), then the occurrence of disputes in the Maoz data set is used to reset the peace years count.

Werner Peace Years Option

Each “Peace Years” check box has an options button next to it. This button brings up a choice to include (or not) the adjustment developed by Suzanne Werner (Werner 2000) to account for the timing of pre-1816 disputes. Werner looked before 1816 to see when the actual prior dispute involving the two states in a dyad occurred, if such a dispute had

occurred. If no dispute occurred before 1816, then 1648 is taken as the base year of the year of previous dispute, and so 1816 will have a value of 168. If the option is chosen, Werner's value will be added to the assumed 1816 value of 0 (and in turn to subsequent values as appropriate). Note that Werner initially computed values as real numbers, reflecting the number of days to prior conflict. However, EUGene's calculation for peace years is of an integer number of years. Werner's real value is rounded inside EUGene, since EUGene's values are not computed to the same level of precision.

Peace Days

In addition to Peace Years, EUGene will compute the number of days between the last day of the previous dispute in the dyad until January 1 of the year in question. [Since the temporal unit of analysis in EUGene is annual, it would be ex post and inappropriate to do a calculation of the number of days up until the start of a dispute in that year.] Peace Days supports the incorporation of Werner's adjustment to Peace Years consisting of counting years (or in this case days) to the last dispute before 1816.

Initiator / Multiple MID Settings

The settings under this subtab allow the user to set several global parameters for dealing with conflict data across several data sets. The settings here will affect all of the conflict data sets as relevant (i.e. the Maoz dyadic MIDs, COW MIDs, and ICB data). There is not a way to separate and have one data set code initiation or multiple conflicts in one way, while having variables from a different conflict data set coded with different assumptions.

Initiator Coding: Timing:

Setting this option specifies whether initiators are considered to be only states that were involved in a militarized dispute on its first day (dispute originators), or whether states that join into an ongoing dispute are also considered initiators (dispute joiners). This will affect the coding of the initiation variable in a directed setting, which indicates that ccode1 initiated a MID against ccode2.

Background: Individual states have been considered as dispute initiators by scholars in one of two ways. EUGene allows coding of initiation by either originators, or including dyads with joiners.

1) The conventional way to mark dispute initiation is to consider only states involved in a militarized dispute on its first day, that is, dispute originators, as involved in a dyadic initiation as either the initiator or the target. This is available under the first program option below.

2) Method 1 leaves out states that join a militarized dispute on the 2nd (or any subsequent) day of a militarized dispute. These states are not initiators in the sense of "states who started the dispute," and faced a different choice problem than states making a decision to initiate a dispute when no dispute was ongoing. In addition, some of these latecomers may have voluntarily chosen to join the MID, or may be targets of MID expansion by states already involved. If the user wants "joinings" to be marked as "initiations" then one of the available options to include joiners should be marked.

Important Data Note: it is not possible in the COW MID data to tell whether "joiners" 1) joined into a conflict voluntarily (and so truly initiated their entry into the conflict) or 2) were targeted by some state at a subsequent stage of a dispute. Thus "joiner" is really "latecomer," as no information about whether their involvement is by choice is available.

Program options – Directed Dyad: EUGene provides 3 ways to treat originator vs. joiner states when marking initiation in a directed dyad:

1. “Code Only Originators (Involved Day 1) as Initiators”: Marking this first option marks initiation in a dyad only if both states in the dyad are dispute originators (involved day 1).
2. “Include Joiners on Initiating Side as Initiators, and Joiners on Target side as Targets”: Marks initiation in both originator and joiner dyads, assuming all on initiating side are initiators, and all on target side are targets. Will code dispute initiations in directed dyads where both states are originators, where the target state in the directed dyad was a joiner on the target side, and where the initiating state in the directed dyad was a joiner on the initiating side. After a dispute begins between two originators A and B (A initiating), all joining states on the side of state A are marked as initiators against state B and all states on B’s side. All joining states on the side of state B are also considered to be targets of state A and all states on A’s side. For example, if A starts a dispute vs. B, C joins on A’s side, and D joins on B’s side, initiation would be coded in directed dyads A-B, A-D, C-B, and C-D. This option codes the dyad as if we are certain that all states on side A took all of the initiating actions against the states on side B.
3. “Code Originators and Joiners (regardless of side) as Initiators”: Marking this option treats all joiners as initiators regardless of side. It marks an initiation in originator dyads according to other rules (so only the Side A originators are considered initiators, for instance), and marks directed dispute initiations in both directions in all joiner dyads. *Important:* with this option, joiners on both the initiating and target side of the dispute will appear as initiators. That is, if A vs. B is the originating directed dyad, and C joins on B’s side, then “initiation” will be reported for directed dyads A-B, A-C, and C-A. This option makes no assumption about which joiners undertook initiating actions, but codes them all as taking the actions. It is typically the case that before conducting analysis, the user will want to either examine these cases to determine who were voluntary joiners and who were involuntary, or perhaps to simply drop joiners on the target side of the MID.

Program options – NonDirected Dyad: For nondirected dyads, *EUGene* provides 2 ways to treat originator vs. joiner states when marking dispute onset:

1. “Mark dispute/crisis onset ONLY for originators (involved day 1)”: Marking this first option marks a dispute onset in a dyad only if both states in the dyad are dispute originators (involved day 1).
2. “Mark dispute/crisis onset for originators and joiners”: Marking this option marks a dispute onset in a dyad for opposed states in the dispute if both are originators, if one is an originator and one a joiner, or if both are joiners.

Variable note: the identity of joining states can be marked by selecting “Identify Joiners” under the *Variables* page. Selecting “Identify Joiners” will code variables marked “1” for joining states, and “0” for true initiators or states not involved in a dispute.

Statistical note: including joiners as initiators has the effect of increasing the effects of multi-state and multi-year disputes in which many states made later decisions to participate on the results of statistical analysis. It also conflates decisions to join in a dispute with decisions to initiate disputes. We do not believe that these decisions are equivalent.

Initiator Coding: Identity (Side A vs. revisionist states):

Setting this option determines whether 1) initiators are considered to be those states who first crossed the MID use/threat of force threshold (the COW MID definition of initiator, as operationalized in the COW MID data as being on side A), or 2) initiators are instead

marked by coding “revisionist” states as initiators. The first of these possibilities is selected by choosing “Side A” as initiators (side A is designated by COW as the state which crossed the force threshold first). The second is selected by marking the “Use Revisionist State(s)” option.

Background: Scholars have used the MID data in two ways to identify dispute initiators. 1) The MID data codes dispute participants as “Side A” or “Side B.” Side A is the side which crossed the MID threshold (use or threat of force) first, and so are considered dispute initiators if the first option is selected. 2) Sometimes the state which first uses or threatens force is not the state that scholars would view as the “aggressor” or “cause” of a MID. In one extreme example, Poland is coded in the MID data as the initiator of the MID that developed into WWI. This case appears odd because we believe that Germany provoked Poland, and that Germany wanted to change the status quo fundamentally, while Poland did not. The MID data provides a coding of “revisionist” states in a MID which may fit this conception of initiation. Choosing the 2nd option, “Code Revisionist States as Initiators” will mark an initiation between a revisionist state and the states on the other side in a MID.

Case note: in some instances, states on both sides of a dispute are coded as revisionist in the MID data. In these cases, a given dyad-year may display dispute initiations by each side against the other, even though there is really a single MID. Note that the MID data set was developed to focus on decisions to use or threaten militarized force, which is reflected in the Side A / Side B codings. The choice to use the revisionist coding to mark initiation should be made with this in mind.

Option note: If non-directed dyads have been selected for output by the user, then the user will not have the option to choose between these types of initiator. In non-directed dyads, the output is only that a MID occurred, with no identification of the initiator vs. target.

Multiple MIDs in Year and the “Key MID”:

Some variables (initiation, onset, joining) are reported if any new MID is started or joined in a given year. However, most variable values correspond to a single MID in any dyad year. If multiple MIDs occur in a year (new or ongoing), then one must be identified as the MID from which to take critical variable values. Starting in EUGene version 3.0, the proper dispute (referred to as the “key MID” for the dyad in the year) from which to determine hostility levels is determined by the user setting on the “Multiple MIDs in Year” setting under the “Initiator/Multiple MID Settings tab” in the conflict variables section.² This setting will determine whether information is taken from the first MID of a year, or from the MID with the highest hostility level (and so, presumably the most serious dispute).

Each dyad-year thus identifies the number of a “key MID” about which other data are reported. This new procedure should assure more internal consistency in the reported MID variables than in prior versions, where different variables could actually come from distinct MIDs. Note that a few output variables explicitly look at other MIDs besides the key MID, particularly the “ongoing” variable (which reports if any MID was ongoing at the start of a dyad-year) and the “joinany” variables (which reports if the states in a dyad joined any dispute in the dyad-year in question). Also note that there could be a keymidnum reported, and hostility levels reported, even if there is NO initiation and NO reported ongoing dispute. In this case, the hostility levels and MID probably correspond to a dispute that B initiated against A in that year (following the rules for identifying the

² Version 2 of EUGene used a series of rules to determine which of multiple disputes were used in any dyad-year. These rules gave precedence to the first MID in a dyad-year, even if a later dispute was more important or serious (e.g. a war beginning late in a year). The new options add flexibility and clarity.

key MID below). Alternatively, it could be a dispute that A joined in against B, if joiners are not set to be coded as initiators.

Rules: When the “Use most serious MID” option is selected, if there is only one MID with the highest hostility level in the dyad-year, then that MID is identified as the key MID. If there are two MIDs with equally high hostility levels, then a series of rules are followed to identify the appropriate MID as the “key MID” for the dyad in the year. Of MIDs with equally high hostility levels, first, a MID which is newly begun in the year is selected over all MIDs that are not new. Second, if there is not a new MID origination, then the first new event in this year is taken including joining an ongoing dispute. Third, if there is no new event/dispute but there is an ongoing dispute, then hostility values are taken from an ongoing MID in which both states were originators. Fourth, hostility values are taken from an ongoing MID in which one or both states were joiners. In nondirected dyad output, the dispute is taken for the MID with highest hostility level for either country 1, or country 2, whether those disputes involve origination, joining, or are ongoing.

When the “Use First MID” option is selected, in directed-dyad output *EUGene* first looks for a new MID with ccode1 on sideA (if user has specified side A as initiators) or ccode1 as a revisionist state (if user has specified to use revisionist states as initiators). In non-directed dyad output, *EUGene* looks for the first new dispute onset. If no such new MID exists, then *EUGene* identifies the first new MID of any type (including MIDs where one or both states joined the MID). If no new MID exists, then *EUGene* identifies the first MID which is ongoing in a given year.

Note that because ICB data does not contain an indicator of how serious ICB crises are, there is no option to select the most serious ICB crisis in a year. For ICB data, information about the first crisis of the year is always reported. More precisely, if there is a new crisis in a year, information about the first new crisis of a year is output. If there are no new crises but there is an ongoing crisis, then information about the ongoing crisis is output. Finally, if there is neither a new nor ongoing crisis in a dyad-year, then the output consists of 0s.

“User Data” Variable Tab

If you have downloaded additional datasets from the *EUGene* website (see main menu item “User Data”), then variables from those datasets will appear under the “User Data” subtab under “Variables.” Each subtab corresponds to a single auxiliary dataset.

User supplied variables that you select will always appear last in the variable output order from *EUGene*.

If you download user data sets from the *EUGene* server via the procedures under the “User Data” menu, you must exit and reenter *EUGene* for those data sets to appear on the “User Data” variable tab.

Variable Names and Order in Output File

The following are the order of the variables sent to the output file, depending on the level of analysis selected. If the user does not select particular variables, those variables will not appear in the *EUGene* generated data. The variables selected will appear in the order as they appear in the columns below. The data output is created so that one case (either country-year, dyad-year, or dispute year) appears per line in the output file.

Monadic	Directed Dyadic	Non-directed Dyadic	Description
ccode			Correlates of War (COW) CCode number for state

	cocode1	cocode1	COW CCode number for state 1
	cocode2	cocode2	COW CCode number for state 2
year	year	year	Year of observation
abbrev			COW abbreviation
	abbrev1	abbrev1	COW abbreviation for state 1
	abbrev2	abbrev2	COW abbreviation for state 2
ISOcode			International Standards Organization (ISO) numeric country code
	ISOcode1	ISOcode1	ISO numeric country code, state 1
	ISOcode2	ISOcode2	ISO numeric country code, state 2
ISO2let			ISO 2 letter country abbreviation
	ISOcode1	ISOcode1	ISO 2 letter country abbreviation, state 1
	ISOcode2	ISOcode2	ISO 2 letter country abbreviation, state 2
ISO3let			ISO 3 letter country abbreviation, state 2
	ISOcode1	ISOcode1	ISO 3 letter country abbreviation, state 1
	ISOcode2	ISOcode2	ISO 3 letter country abbreviation, state 2
ISOShNm			ISO short country name (English)
	ISOcode1	ISOcode1	ISO short country name (English), state 1
	ISOcode2	ISOcode2	ISO short country name (English), state 2
ISOLgNm			ISO long country name (English)
	ISOcode1	ISOcode1	ISO long country name (English), state 1
	ISOcode2	ISOcode2	ISO long country name (English), state 2
cap			National Capabilities Index for CCode
	cap_1	cap_1	National Capabilities Index for CCode1
	cap_2	cap_2	National Capabilities Index for CCode2
milper			COW military personnel for CCode
milex			COW military expenditure for CCode
energy			COW energy production for CCode
irst			COW iron/steel production for CCode
upop			COW nominal urban population for CCode
tpop			COW nominal total population for CCode
	milper_1	milper_1	COW military personnel for CCode1
	milex_1	milex_1	COW military expenditure for CCode1
	energy_1	energy_1	COW energy production for CCode1
	irst_1	irst_1	COW iron/steel production for CCode1
	upop_1	upop_1	COW nominal urban population for CCode1
	tpop_1	tpop_1	COW nominal total population for CCode1
	milper_2	milper_2	COW military personnel for CCode2
	milex_2	milex_2	COW military expenditure for CCode2
	energy_2	energy_2	COW energy production for CCode2
	irst_2	irst_2	COW iron/steel production for CCode2
	upop_2	upop_2	COW nominal urban population for CCode2
	tpop_2	tpop_2	COW nominal total population for CCode2
majpow			Major power status of CCode
	majpow1	majpow1	Major power status of CCode1
	majpow2	majpow2	Major power status of CCode2
	rlregion		Region relevant to Directed Dyad CCode1 vs. CCode2
		rlreg12	Region relevant to Directed Dyad CCode1 vs. CCode2
		rlreg21	Region relevant to Directed Dyad CCode2 vs. CCode1
region1	region1	region1	Home region of CCode1 (Europe=1, Middle East=2, Africa=3, Asia=4, Americas=5)
	region2	region2	Home region of CCode2
	pol_rel	pol_rel	Politically relevant dyad (0=no, 1=yes)
	tau_regi		Tau, CCode1 vs. CCode2, computed using countries in relevant region
		tau_re1	Tau, CCode1 vs. CCode2, using countries in CCode1's home region

		tau_re2	Tau, CCode1 vs. CCode2, using countries in CCode2's home region
	tau_glob	tau_glob	Tau, CCode1 vs. CCode2, using all countries
	s_un_reg		s score, unweighted, calculated with countries in the relevant region in CCode1 vs. CCode2 directed dyad
		s_un_re1	s score, unweighted, regionally based using countries in CCode1's home region
		s_un_re2	s score, unweighted, regionally based using countries in CCode2's home region
	s_un_glo	s_un_glo	s score, unweighted, global computation using all states
	s_wt_reg		s score, weighted, calculated with countries in the relevant region in CCode1 vs. CCode2 directed dyad
		s_wt_re1	s score, weighted, regionally based using countries in CCode1's home region
		s_we_re2	s score, weighted, regionally based using countries in CCode2's home region
	s_wt_glo	s_wt_glo	s score, weighted, global computation using all states
	rportu11	rportu11	Alliance portfolio between ccode1 and ccode2, unweighted version (number of states in each cell), regional version (counting states in the relevant region of conflict between ccode1 and ccode2), row 1 column 1 (defense pact – defense pact).
	rportu12	rportu12	Alliance portfolio between ccode1 and ccode2, unweighted, regional, row 1 column 2 (defense pact – neutrality/non-aggression pact).
	rportu13	rportu13	Alliance portfolio between ccode1 and ccode2, unweighted, regional, row 1 column 2 (defense pact – entente).
	rportu14	rportu14	Alliance portfolio between ccode1 and ccode2, unweighted, regional, row 1 column 4 (defense pact – no alliance).
	rportu21 through rportu44	rportu21 through rportu44	Alliance portfolio between ccode1 and ccode2, unweighted, regional, row 2 column 1 through row 4, column 4.
	rportn	rportn	Total number of states counted in the unweighted regional alliance portfolio table.
	gportu11	gportu11	Alliance portfolio between ccode1 and ccode2, unweighted version (number of states in each cell), global version (counting all states in the system), row 1 column 1 (defense pact – defense pact).
	gportu12	gportu12	Alliance portfolio between ccode1 and ccode2, unweighted, global (all states), row 1 column 2 (defense pact – neutrality/non-aggression pact).
	gportu13	gportu13	Alliance portfolio between ccode1 and ccode2, unweighted, global, row 1 column 2 (defense pact – entente).
	gportu14	gportu14	Alliance portfolio between ccode1 and ccode2, unweighted, global, row 1 column 4 (defense pact – no alliance).
	gportu21 through gportu44	gportu21 through gportu44	Alliance portfolio between ccode1 and ccode2, unweighted, global, row 2 column 1 through row 4, column 4.
	gportn	gportn	Total number of states counted in the unweighted, global, alliance portfolio table.
	rportw11	rportw11	Alliance portfolio between ccode1 and ccode2, regional, weighted version (sum of capabilities in each cell), row 1 column 1 (defense pact – defense pact).
	rportw12	rportw12	Alliance portfolio between ccode1 and ccode2, regional, weighted, row 1 column 2 (defense pact – neutrality/non-aggression pact).
	rportw13	rportw13	Alliance portfolio between ccode1 and ccode2, regional, weighted, row 1 column 2 (defense pact – entente).
	rportw14	rportw14	Alliance portfolio between ccode1 and ccode2, regional, weighted, row 1 column 4 (defense pact – no alliance).

	rportw21 through rportw44	rportw21 through rportw44	Alliance portfolio between ccode1 and ccode2, regional, weighted, row 2 column 1 through row 4, column 4.
	rpordmax	rpordmax	Total sum of all capabilities in the weighted regional alliance portfolio table.
	gportw11	gportw11	Alliance portfolio between ccode1 and ccode2, global, weighted version (sum of capabilities in each cell), row 1 column 1 (defense pact – defense pact).
	gportw12	gportw12	Alliance portfolio between ccode1 and ccode2, global, weighted, row 1 column 2 (defense pact – neutrality/non-aggression pact).
	gportw13	gportw13	Alliance portfolio between ccode1 and ccode2, global, weighted, row 1 column 2 (defense pact – entente).
	gportw14	gportw14	Alliance portfolio between ccode1 and ccode2, global, weighted, row 1 column 4 (defense pact – no alliance).
	gportw21 through gportw44	gportw21 through gportw44	Alliance portfolio between ccode1 and ccode2, global, weighted, row 2 column 1 through row 4, column 4.
	gpordmax	gpordmax	Total sum of all capabilities in the weighted global alliance portfolio table.
tau_lead			Tau between CCode and system leader. If regional setting is specified, this uses countries in the relevant region for the directed dyad CCode vs. leader. If global setting is specified, this uses all countries in Tau calculations.
	tau_ld_1		Tau between CCode1 and system leader using countries in the relevant region.
	tau_ld_2		Tau between CCode2 and system leader using countries in the relevant region.
		tau_ld_1	Tau between CCode1 and system leader. If regional setting specified, the calculation is based on countries from the home region of CCode1.
		tau_ld_2	Tau between CCode2 and system leader. If regional setting specified, the calculation is based on countries from the home region of CCode2.
s_lead			S between CCode and system leader. If regional setting is specified, this uses countries in the relevant region for the directed dyad CCode vs. leader. If global setting is specified, this uses all countries in S calculations.
	s_ld_1		S between CCode1 and system leader using countries in the relevant region.
	s_ld_2		S between CCode2 and system leader using countries in the relevant region.
		s_ld_1	S between CCode1 and system leader. If regional setting specified, the calculation is based on countries from the home region of CCode1.
		s_ld_2	S between CCode2 and system leader. If regional setting specified, the calculation is based on countries from the home region of CCode2.
	alliance	alliance	Alliance agreement type (1=defense pact, 2=neutrality, 3=entente, 4=no agreement). This is from the variable “sstype” in the COW dyadic alliance data set.
democ			Jagers and Gurr (1995) Polity III Democracy Score for CCode
	democ1	democ1	Polity III Democracy Score for CCode1
	democ2	democ2	Polity III Democracy Score for CCode2
autoc			Polity III Autocracy Score for CCode
	autoc1	autoc1	Polity III Autocracy Score for CCode1
	autoc2	autoc2	Polity III Autocracy Score for CCode2
xrreg			Polity III Executive Recruitment Regulation Score, CCode
	xrreg1	xrreg1	Polity III Executive Recruitment Regulation Score, CCode1
	xrreg2	xrreg2	Polity III Executive Recruitment Regulation Score, CCode2

xrcomp			Polity III Executive Recruitment Competition Score, CCode
	xrcomp1	xrcomp1	Polity III Executive Recruitment Competition Score, CCode1
	xrcomp2	xrcomp2	Polity III Executive Recruitment Competition Score, CCode2
xropen			Polity III Executive Recruitment Openness Score, CCode
	xropen1	xropen1	Polity III Executive Recruitment Openness Score, CCode1
	xropen2	xropen2	Polity III Executive Recruitment Openness Score, CCode2
mono			Polity III Monocratism Score for CCode
	mono1	mono1	Polity III Monocratism Score for CCode1
	mono2	mono2	Polity III Monocratism Score for CCode2
xconst			Polity III Executive Constraints Score for CCode
	xconst1	xconst1	Polity III Executive Constraints Score for CCode1
	xconst2	xconst2	Polity III Executive Constraints Score for CCode2
parreg			Polity III Participation Regulation Score for CCode
	parreg1	parreg1	Polity III Participation Regulation Score for CCode1
	parreg2	parreg2	Polity III Participation Regulation Score for CCode2
parcomp			Polity III Participation Competitiveness Score for CCode
	parcomp1	parcomp1	Polity III Participation Competitiveness Score for CCode1
	parcomp2	parcomp2	Polity III Participation Competitiveness Score for CCode2
cent			Polity III Centralization Score for CCode
	cent1	cent1	Polity III Centralization Score for CCode1
	cent2	cent2	Polity III Centralization Score for CCode2
dem			Democracy Score for CCode (democ – autoc)
	dem1	dem1	Democracy Score for CCode1 (democ – autoc)
	dem2	dem2	Democracy Score for CCode2 (democ – autoc)
democlg			“democ”, lagged one year, for CCode
autoclg			“autoc”, lagged one year, for CCode
demlg			“dem”, lagged one year, for CCode
	democlg1	democlg1	“democ”, lagged one year, for CCode1
	democlg2	democlg2	“democ”, lagged one year, for CCode2
	autoclg1	autoclg1	“autoc”, lagged one year, for CCode1
	autoclg2	autoclg2	“autoc”, lagged one year, for CCode2
	demlg1	demlg1	“dem”, lagged one year, for CCode1
	demlg2	demlg2	“dem”, lagged one year, for CCode2
demchg			Democratization Score for CCode (democ – democlg)
	demchg1	demchg1	Democratization Score for CCode1 (democ1 – democlg1)
	demchg2	demchg2	Democratization Score for CCode2 (democ2 – democlg2)
	contig	contig	Direct contiguity level (1-6, 1=land contiguity, 2=contiguous across up to 12 miles of water, 3=contiguous across 13-24 miles of water, 4=contiguous across 25-150 miles of water, 5=contiguous across 151-400 miles of water, 6=not contiguous)
	colcont	colcont	Colonial contiguity level (1-6, 1=land contiguity, 2=contiguous across up to 12 miles of water, 3=contiguous across 13-24 miles of water, 4=contiguous across 25-150 miles of water, 5=contiguous across 151-400 miles of water, 6=not contiguous), taking into account the closest colonial relationship.
	colony1	colony1	The state/entity number of CCode1’s colony taken into account in computing the colonial contiguity level. If the level is between CCode1 and a colony of CCode2, this is 0.
	colony2	colony2	The state/entity number of CCode2’s colony taken into account in computing the colonial contiguity level. If the level is between CCode2 and a colony of CCode1, this is 0.
	distance	distance	Distance, from CCode1 to CCode2, in miles
	dyaddur	dyaddur	Dyadic duration
numstate	numstate	numstate	Number of states in the international system, as defined by COW
numGPs	numGPs	numGPs	Number of great powers in the international system
syscon	syscon	syscon	System concentration (CON), entire system
sysmove1	sysmove1	sysmove1	System movement (MOVE), yearly computation, entire system
sysmove5	sysmove5	sysmove5	System movement (MOVE), 5-year moving average, entire

			system
sysmvGP1	sysmvGP1	sysmvGP1	System movement (MOVE), yearly computation, great powers
sysmvGP5	sysmvGP5	sysmvGP5	System movement (MOVE), 5-year moving average, great powers
	euwtT1v2	euwtT1v2	Expected Utility, CCode1 vs. CCode2, by War Trap methods (Tau)
		euwtT2v1	Expected Utility, CCode2 vs. CCode1, by War Trap methods (Tau)
	euwtS1v2	euwtS1v2	Expected Utility, CCode1 vs. CCode2, by War Trap methods (s)
		euwtS2v1	Expected Utility, CCode2 vs. CCode1, by War Trap methods (s)
riskTeur			Risk Score for CCode in Europe (based on Tau-b)
riskTmid			Risk Score for CCode in Middle East (Tau)
riskTafr			Risk Score for CCode in Africa (Tau)
riskTasi			Risk Score for CCode in Asia (Tau)
riskTame			Risk Score for CCode in Americas (Tau)
riskTglo			Risk Score for CCode globally (Tau)
	riskT1		Risk Score for CCode1 in relevant region (Tau)
	riskT2		Risk Score for CCode2 in relevant region (Tau)
		riskT11	Risk Score for CCode1 in CCode1's home region (Tau)
		riskT12	Risk Score for CCode1 in CCode2's home region (Tau)
		riskT21	Risk Score for CCode2 in CCode1's home region (Tau)
		riskT22	Risk Score for CCode2 in CCode2's home region (Tau)
riskSeur			Risk Score for CCode in Europe (based on s)
riskSmid			Risk Score for CCode in Middle East (s)
riskSafr			Risk Score for CCode in Africa (s)
riskSasi			Risk Score for CCode in Asia (s)
riskSame			Risk Score for CCode in Americas (s)
riskSglo			Risk Score for CCode globally (s)
	riskS1		Risk Score for CCode1 in relevant region (s)
	riskS2		Risk Score for CCode2 in relevant region (s)
		riskS11	Risk Score for CCode1 in CCode1's home region (s)
		riskS12	Risk Score for CCode1 in CCode2's home region (s)
		riskS21	Risk Score for CCode2 in CCode1's home region (s)
		riskS22	Risk Score for CCode2 in CCode2's home region (s)
seurTact			Actual security level for CCode in Europe (for Tau risk calculation)
seurTmax			Maximum hypothetical security level for CCode in Europe (Tau)
seurTmin			Minimum hypothetical security level for CCode in Europe (Tau)
smidTact			Actual security level for CCode in Middle East (Tau)
smidTmax			Maximum hypothetical security level for CCode in Middle East (Tau)
smidTmin			Minimum hypothetical security level for CCode in Middle East (Tau)
safrTact			Actual security level for CCode in Africa (Tau)
safrTmax			Maximum hypothetical security level for CCode in Africa (Tau)
safrTmin			Minimum hypothetical security level for CCode in Africa (Tau)
sasiTact			Actual security level for CCode in Asia (Tau)
sasiTmax			Maximum hypothetical security level for CCode in Asia (Tau)
sasiTmin			Minimum hypothetical security level for CCode in Asia (Tau)
sameTact			Actual security level for CCode in Americas (Tau)
sameTmax			Maximum hypothetical security level for CCode in Americas (Tau)
sameTmin			Minimum hypothetical security level for CCode in Americas (Tau)
sgloTact			Actual security level for CCode globally (Tau)
sgloTmax			Maximum hypothetical security level for CCode globally (Tau)
sgloTmin			Minimum hypothetical security level for CCode globally (Tau)

	s1Tact		Actual security level found for CCode1 (for Tau risk calculation)
	s1Tmax		Maximum hypothetical security level found for CCode1 (Tau)
	s1Tmin		Minimum hypothetical security level found for CCode1 (Tau)
	s2Tact		Actual security level found for CCode2 (Tau)
	s2Tmax		Maximum hypothetical security level found for CCode2 (Tau)
	s2Tmin		Minimum hypothetical security level found for CCode2 (Tau)
		s1Tact1	Actual security level for CCode1 in CCode1's home region (for Tau risk calculation)
		s1Tmax1	Maximum hypothetical security level for CCode1 in own home region (Tau)
		s1Tmin1	Minimum hypothetical security level for CCode1 in CCode1's home region (Tau)
		s1Tact2	Actual security level for CCode1 in CCode2's home region (Tau)
		s1Tmax2	Maximum hypothetical security level for CCode1 in CCode2's home region (Tau)
		s1Tmin2	Minimum hypothetical security level for CCode1 in CCode2's home region (Tau)
		s2Tact1	Actual security level for CCode2 in CCode1's home region (Tau)
		s2Tmax1	Maximum hypothetical security level for CCode2 in CCode1's home region (Tau)
		s2Tmin1	Minimum hypothetical security level for CCode2 in CCode1's home region (Tau)
		s2Tact2	Actual security level for CCode2 in CCode2's home region (Tau)
		s2Tmax2	Maximum hypothetical security level for CCode2 in CCode2's home region (Tau)
		s2Tmin2	Minimum hypothetical security level for CCode2 in CCode2's home region (Tau)
seurSact			Actual security level for CCode in Europe (for s risk calculation)
seurSmax			Maximum hypothetical security level for CCode in Europe (s)
seurSmin			Minimum hypothetical security level for CCode in Europe (s)
smidSact			Actual security level for CCode in Middle East (s)
smidSmax			Maximum hypothetical security level for CCode in Middle East (s)
smidSmin			Minimum hypothetical security level for CCode in Middle East (s)
safrSact			Actual security level for CCode in Africa (s)
safrSmax			Maximum hypothetical security level for CCode in Africa (s)
safrSmin			Minimum hypothetical security level for CCode in Africa (s)
sasiSact			Actual security level for CCode in Asia (s)
sasiSmax			Maximum hypothetical security level for CCode in Asia (s)
sasiSmin			Minimum hypothetical security level for CCode in Asia (s)
sameSact			Actual security level for CCode in Americas (s)
sameSmax			Maximum hypothetical security level for CCode in Americas (s)
sameSmin			Minimum hypothetical security level for CCode in Americas (s)
sgloSact			Actual security level for CCode globally (s)
sgloSmax			Maximum hypothetical security level for CCode globally (s)
sgloSmin			Minimum hypothetical security level for CCode globally (s)
	s1Sact		Actual security level found for CCode1 (for s risk calculation)
	s1Smax		Maximum hypothetical security level found for CCode1 (s)
	s1Smin		Minimum hypothetical security level found for CCode1 (s)
	s2Sact		Actual security level found for CCode2 (s)
	s2Smax		Maximum hypothetical security level found for CCode2 (s)
	s2Smin		Minimum hypothetical security level found for CCode2 (s)
		s1Sact1	Actual security level for CCode1 in CCode1's home region (for

			s risk calculation)
		s1Smax1	Maximum hypothetical security level for CCode1 in CCode1's home region (s)
		s1Smin1	Minimum hypothetical security level for CCode1 in CCode1's home region (s)
		s1Sact2	Actual security level for CCode1 in CCode2's home region (s)
		s1Smax2	Maximum hypothetical security level for CCode1 in CCode2's home region (s)
		s1Smin2	Minimum hypothetical security level for CCode1 in CCode2's home region (s)
		s2Sact1	Actual security level for CCode2 in CCode1's home region (s)
		s2Smax1	Maximum hypothetical security level for CCode2 in CCode1's home region (s)
		s2Smin1	Minimum hypothetical security level for CCode2 in CCode1's home region (s)
		s2Sact2	Actual security level for CCode2 in CCode2's home region (s)
		s2Smax2	Maximum hypothetical security level for CCode2 in CCode2's home region (s)
		s2Smin2	Minimum hypothetical security level for CCode2 in CCode2's home region (s)
uncertT	uncertT		Regional uncertainty (based on Tau-b)
		uncertT1	Home region uncertainty, CCode1 (Tau)
		uncertT2	Home region uncertainty, CCode2 (Tau)
uncertS	uncertS		Regional uncertainty (based on s)
		uncertS1	Home region uncertainty, CCode1 (s)
		uncertS2	Home region uncertainty, CCode2 (s)
	wrTu1v1	wrTu1v1	War and Reason, $U^A(\Delta_A)$, using Tau
	wrTu1v2	wrTu1v2	War and Reason, $U^A(\Delta_B)$, using Tau
	wrTu1vsq	wrTu1vsq	War and Reason, $U^A(SQ)$, using Tau
	wrTp1win	wrTp1win	War and Reason, P^A , using Tau
	wrTstk1	wrTstk1	War and Reason, $Stakes_A$, using Tau
	wrTu1sq	wrTu1sq	War and Reason, $U^j(SQ)$, using Tau
	wrTu1ac1	wrTu1ac1	War and Reason, $U^j(Acq_i)$, using Tau
	wrTu1ac2	wrTu1ac2	War and Reason, $U^j(Acq_i)$, using Tau
	wrTu1neg	wrTu1neg	War and Reason, $U^j(Nego)$, using Tau
	wrTu1cp1	wrTu1cp1	War and Reason, $U^j(Cap_i)$, using Tau
	wrTu1cp2	wrTu1cp2	War and Reason, $U^j(Cap_i)$, using Tau
	wrTu1wr1	wrTu1wr1	War and Reason, $U^j(War_i)$, using Tau
	wrTu1wr2	wrTu1wr2	War and Reason, $U^j(War_i)$, using Tau
	wrTu2v2	wrTu2v2	War and Reason, $U^B(\Delta_B)$, using Tau
	wrTu2v1	wrTu2v1	War and Reason, $U^B(\Delta_A)$, using Tau
	wrTu2vsq	wrTu2vsq	War and Reason, $U^B(SQ)$, using Tau
	wrTp2win	wrTp2win	War and Reason, P^B , using Tau
	wrTstk2	wrTstk2	War and Reason, $Stakes_B$, using Tau
	wrTu2sq	wrTu2sq	War and Reason, $U^j(SQ)$, using Tau
	wrTu2ac2	wrTu2ac2	War and Reason, $U^j(Acq_i)$, using Tau
	wrTu2ac1	wrTu2ac1	War and Reason, $U^j(Acq_i)$, using Tau
	wrTu2neg	wrTu2neg	War and Reason, $U^j(Nego)$, using Tau
	wrTu2cp2	wrTu2cp2	War and Reason, $U^j(Cap_i)$, using Tau
	wrTu2cp1	wrTu2cp1	War and Reason, $U^j(Cap_i)$, using Tau
	wrTu2wr2	wrTu2wr2	War and Reason, $U^j(War_i)$, using Tau
	wrTu2wr1	wrTu2wr1	War and Reason, $U^j(War_i)$, using Tau
	wrSulv1	wrSulv1	War and Reason, $U^A(\Delta_A)$, using s (unweighted)
	wrSulv2	wrSulv2	War and Reason, $U^A(\Delta_B)$, using s (unweighted)
	wrSulvsq	wrSulvsq	War and Reason, $U^A(SQ)$, using s (unweighted)
	wrSplwin	wrSplwin	War and Reason, P^A , using s (unweighted)
	wrSstk1	wrSstk1	War and Reason, $Stakes_A$, using s (unweighted)
	wrSulsq	wrSulsq	War and Reason, $U^j(SQ)$, using s (unweighted)

	wrSu1ac1	wrSu1ac1	War and Reason, $U^1(\text{Acq}_i)$, using s (unweighted)
	wrSu1ac2	wrSu1ac2	War and Reason, $U^1(\text{Acq}_j)$, using s (unweighted)
	wrSu1neg	wrSu1neg	War and Reason, $U^1(\text{Nego})$, using s (unweighted)
	wrSu1cp1	wrSu1cp1	War and Reason, $U^1(\text{Cap}_i)$, using s (unweighted)
	wrSu1cp2	wrSu1cp2	War and Reason, $U^1(\text{Cap}_j)$, using s (unweighted)
	wrSu1wr1	wrSu1wr1	War and Reason, $U^1(\text{War}_i)$, using s (unweighted)
	wrSu1wr2	wrSu1wr2	War and Reason, $U^1(\text{War}_j)$, using s (unweighted)
	wrSu2v2	wrSu2v2	War and Reason, $U^B(\Delta_B)$, using s (unweighted)
	wrSu2v1	wrSu2v1	War and Reason, $U^B(\Delta_A)$, using s (unweighted)
	wrSu2vsq	wrSu2vsq	War and Reason, $U^B(\text{SQ})$, using s (unweighted)
	wrSp2win	wrSp2win	War and Reason, P^B , using s (unweighted)
	wrSstk2	wrSstk2	War and Reason, Stakes_B , using s (unweighted)
	wrSu2sq	wrSu2sq	War and Reason, $U^j(\text{SQ})$, using s (unweighted)
	wrSu2ac2	wrSu2ac2	War and Reason, $U^j(\text{Acq}_i)$, using s (unweighted)
	wrSu2ac1	wrSu2ac1	War and Reason, $U^j(\text{Acq}_j)$, using s (unweighted)
	wrSu2neg	wrSu2neg	War and Reason, $U^j(\text{Nego})$, using s (unweighted)
	wrSu2cp2	wrSu2cp2	War and Reason, $U^j(\text{Cap}_i)$, using s (unweighted)
	wrSu2cp1	wrSu2cp1	War and Reason, $U^j(\text{Cap}_j)$, using s (unweighted)
	wrSu2wr2	wrSu2wr2	War and Reason, $U^j(\text{War}_i)$, using s (unweighted)
	wrSu2wr1	wrSu2wr1	War and Reason, $U^j(\text{War}_j)$, using s (unweighted)
	wrWu1v1	wrWu1v1	War and Reason, $U^A(\Delta_A)$, using s (weighted)
	wrWu1v2	wrWu1v2	War and Reason, $U^A(\Delta_B)$, using s (weighted)
	wrWu1vsq	wrWu1vsq	War and Reason, $U^A(\text{SQ})$, using s (weighted)
	wrWp1win	wrWp1win	War and Reason, P^A , using s (weighted)
	wrWstk1	wrWstk1	War and Reason, Stakes_A , using s (weighted)
	wrWu1sq	wrWu1sq	War and Reason, $U^i(\text{SQ})$, using s (weighted)
	wrWu1ac1	wrWu1ac1	War and Reason, $U^i(\text{Acq}_i)$, using s (weighted)
	wrWu1ac2	wrWu1ac2	War and Reason, $U^i(\text{Acq}_j)$, using s (weighted)
	wrWu1neg	wrWu1neg	War and Reason, $U^i(\text{Nego})$, using s (weighted)
	wrWu1cp1	wrWu1cp1	War and Reason, $U^i(\text{Cap}_i)$, using s (weighted)
	wrWu1cp2	wrWu1cp2	War and Reason, $U^i(\text{Cap}_j)$, using s (weighted)
	wrWu1wr1	wrWu1wr1	War and Reason, $U^i(\text{War}_i)$, using s (weighted)
	wrWu1wr2	wrWu1wr2	War and Reason, $U^i(\text{War}_j)$, using s (weighted)
	wrWu2v2	wrWu2v2	War and Reason, $U^B(\Delta_B)$, using s (weighted)
	wrWu2v1	wrWu2v1	War and Reason, $U^B(\Delta_A)$, using s (weighted)
	wrWu2vsq	wrWu2vsq	War and Reason, $U^B(\text{SQ})$, using s (weighted)
	wrWp2win	wrWp2win	War and Reason, P^B , using s (weighted)
	wrWstk2	wrWstk2	War and Reason, Stakes_B , using s (weighted)
	wrWu2sq	wrWu2sq	War and Reason, $U^j(\text{SQ})$, using s (weighted)
	wrWu2ac2	wrWu2ac2	War and Reason, $U^j(\text{Acq}_i)$, using s (weighted)
	wrWu2ac1	wrWu2ac1	War and Reason, $U^j(\text{Acq}_j)$, using s (weighted)
	wrWu2neg	wrWu2neg	War and Reason, $U^j(\text{Nego})$, using s (weighted)
	wrWu2cp2	wrWu2cp2	War and Reason, $U^j(\text{Cap}_i)$, using s (weighted)
	wrWu2cp1	wrWu2cp1	War and Reason, $U^j(\text{Cap}_j)$, using s (weighted)
	wrWu2wr2	wrWu2wr2	War and Reason, $U^j(\text{War}_i)$, using s (weighted)
	wrWu2wr1	wrWu2wr1	War and Reason, $U^j(\text{War}_j)$, using s (weighted)
	eqTsq		Status quo is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau
	eqTnego		Negotiation is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau
	eqTacqa		Acquiescence by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau
	eqTacqb		Acquiescence by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau
	eqTcapa		Capitulation by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau
	eqTcapb		Capitulation by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau

	eqTwara		War initiated by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau
	eqTwarb		War initiated by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau
		eqTsq12	Status quo is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau
		eqTneg12	Negotiation is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau
		eqTaca12	Acquiescence by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau
		eqTacb12	Acquiescence by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau
		eqTcpa12	Capitulation by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau
		eqTcpb12	Capitulation by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau
		eqTwra12	War initiated by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau
		eqTwrb12	War initiated by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using Tau
		eqTsq21	Status quo is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using Tau
		eqTneg21	Negotiation is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using Tau
		eqTaca21	Acquiescence by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using Tau
		eqTacb21	Acquiescence by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using Tau
		eqTcpa21	Capitulation by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using Tau
		eqTcpb21	Capitulation by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using Tau
		eqTwra21	War initiated by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using Tau
		eqTwrb21	War initiated by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using Tau
	eqSsq		Status quo is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)
	eqSnego		Negotiation is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)
	eqSacqa		Acquiescence by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)
	eqSacqb		Acquiescence by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)
	eqScapa		Capitulation by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)
	eqScapb		Capitulation by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)
	eqSwara		War initiated by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)
	eqSwarb		War initiated by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)
		eqSsq12	Status quo is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)

		eqSneg12	Negotiation is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)
		eqSaca12	Acquiescence by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)
		eqSacb12	Acquiescence by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)
		eqScpa12	Capitulation by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)
		eqScpb12	Capitulation by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)
		eqSwra12	War initiated by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)
		eqSwrb12	War initiated by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (unweighted)
		eqSsq21	Status quo is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (unweighted)
		eqSneg21	Negotiation is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (unweighted)
		eqSaca21	Acquiescence by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (unweighted)
		eqSacb21	Acquiescence by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (unweighted)
		eqScpa21	Capitulation by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (unweighted)
		eqScpb21	Capitulation by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (unweighted)
		eqSwra21	War initiated by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (unweighted)
		eqSwrb21	War initiated by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (unweighted)
	eqWsq		Status quo is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)
	eqWnego		Negotiation is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)
	eqWacqa		Acquiescence by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)
	eqWacqb		Acquiescence by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)
	eqWcapa		Capitulation by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)
	eqWcapb		Capitulation by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)
	eqWwara		War initiated by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)
	eqWwarb		War initiated by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)
		eqWsq12	Status quo is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)

		eqWneg12	Negotiation is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)
		eqWaca12	Acquiescence by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)
		eqWacb12	Acquiescence by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)
		eqWcpa12	Capitulation by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)
		eqWcpb12	Capitulation by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)
		eqWwra12	War initiated by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)
		eqWwrb12	War initiated by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode1 vs. CCode2, using s (weighted)
		eqWsq21	Status quo is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (weighted)
		eqWneg21	Negotiation is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (weighted)
		eqWaca21	Acquiescence by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (weighted)
		eqWacb21	Acquiescence by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (weighted)
		eqWcpa21	Capitulation by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (weighted)
		eqWcpb21	Capitulation by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (weighted)
		eqWwra21	War initiated by CCode2 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (weighted)
		eqWwrb21	War initiated by CCode1 is expected in equilibrium (0=no, 1=yes) in directed dyad CCode2 vs. CCode1, using s (weighted)
	cwongo	cwongo	A MID was ongoing at the beginning of this year, based on "dyadization" of the COW MID data set to 1992, and dyadic MID data from v3.0 from 1992+ (0=no, 1=yes)
	cwongonm	cwongonm	Number of the MID that was ongoing at the beginning of the year, if any (0 if no MID was ongoing)
	cwinit		MID Initiation: CCode1 initiated a MID vs. CCode2 in this year (0=no, 1=yes). Depending on settings for ongoing disputes and joiners, this may be a new MID or the 2 nd (3 rd , etc.) year of an ongoing MID and may include only dispute originators or MID joiners.
	cwinitnm		Number of the new MID that was <u>initiated</u> in this year, if any (0 if no MID was initiated)
		cwmid	MID Onset: A MID occurred between CCode1 and CCode2 in this year (0=no, 1=yes). Depending on settings for ongoing disputes and joiners, this may be a new MID or the 2 nd (3 rd , etc.) year of an ongoing MID, and may include only dispute originators or MID joiners.
		cwmidnm	Number of the new MID that <u>began</u> in this year, if any; marks dispute <u>ONSET</u> (0 if no MID was ongoing)
	cdynm	cdynm	Dyadic MID number (new number for each dyad-episode within each MID)
	cwhost1	cwhost1	Relevant hostility level reached by CCode1 in a MID vs. CCode2 in this year (0=No hostility [no MID], 1=No militarized action, 2=Threat to use force, 3=Display of force, 4=Use of Force, 5=War)
	cwhost2	cwhost2	Relevant hostility level reached by CCode2 in a MID vs. CCode1 in this year (0 to 5, as in hostlev1)
	cwhostd	cwhostd	Overall dispute hostility level (0 to 5, as in hostlev1)
	cwkeynum	cwkeynum	Number of the MID for which all other MID variables are

			reported in dyad-year data. Depending on user settings, this could be the number of the first MID that occurs in year, or the most serious MID, or the number of an ongoing MID. Referred to as the “key MID.”
	cworig1	cworig1	CCode1 is originator of the key MID, i.e. is involved on first day (0=no, 1=yes)
	cworig2	cworig2	CCode2 is originator of the key MID (0=no, 1=yes)
	cwjoanyi		In this directed dyad, in <u>any</u> MID in this year, one of the states was a latecomer, and ccode1 was on the <u>initiating</u> side of the MID (0=no, 1=yes).
	cwjoanyt		In this directed dyad, in <u>any</u> MID in this year, one of the states was a latecomer, and ccode1 was on the <u>target</u> side of the MID
	cwjomidi		In this directed dyad, in <u>the MID number reported</u> under variable “cwkeynum” in this year, one of the states was a latecomer, and ccode1 was on the <u>initiating</u> side of the MID (0=no, 1=yes)
	cwjomidt		In this directed dyad, in <u>the MID number reported</u> under variable “cwkeynum” in this year, one of the states was a latecomer, and ccode1 was on the <u>target</u> side of the MID (0=no, 1=yes)
		cwjoany	In this dyad, in <u>any</u> MID in this year, one of the states was a latecomer
		cwjomid	In this dyad, in <u>the MID number reported</u> under variable “cwkeynum” in this year, one of the states was a latecomer
	cwmidnme	cwmidnme	Name of the MID identified in “cwkeynum,” if one has been designated by COW.
	cwstmo1	cwstmo1	MID start month for CCode1
	cwstday1	cwstday1	MID start day for CCode1
	cwstyr1	cwstyr1	MID start year for CCode1
	cwstmo2	cwstmo2	MID start month for CCode2
	cwstday2	cwstday2	MID start day for CCode2
	cwstyr2	cwstyr2	MID start year for CCode2
	cwendmo1	cwendmo1	MID end month for CCode1
	cwenddy1	cwenddy1	MID end day for CCode1
	cwendyr1	cwendyr1	MID end year for CCode1
	cwendmo2	cwendmo2	MID end month for CCode2
	cwenddy2	cwenddy2	MID end day for CCode2
	cwendyr2	cwendyr2	MID end year for CCode2
	cwsida1	cwsida1	CCode1 is on side A side of cwkeynum (0=no, 1=yes)
	cwsida2	cwsida2	CCode2 is on side A of cwkeynum (0=no, 1=yes)
	cwrevis1	cwrevis1	CCode1 is a revisionist state in cwkeynum (0=no, 1=yes)
	cwrevis2	cwrevis2	Code2 is a revisionist state in cwkeynum (0=no, 1=yes)
	cwrevt11	cwrevt11	Revision Type 1 for CCode/state 1 (0=not applicable, 1=territory, 2=policy, 3=regime/government, 4=other)
	cwrevt21	cwrevt21	Revision Type 2, State 1, when present in COW v3.0 data (always missing before 1992).
	cwrevt12	cwrevt12	Revision Type 1, State 2
	cwrevt22	cwrevt22	Revision Type 2, State 2, when present in COW v3.0 data.
	cfatal1	cfatal1	Fatality level for CCode1 (0=none, 1=1-25 deaths, 2=26-100 deaths, 3=101-250 deaths, 4=251-500 deaths, 5=501-999 deaths, 6=>999 deaths, -9=missing)
	cfatex1	cfatex1	Exact estimate of fatalities for CCode1, if estimated in MID v3.0.
	cfatal2	cfatal2	Fatality level for CCode2 (0 to 6, as in fatal1)
	cfatex2	cfatex2	Exact estimate of fatalities for CCode2, if estimated in MID v3.0.
	cwhiact1	cwhiact1	Highest action by CCode1 in dispute (see MID B codebook)
	cwhiact2	cwhiact2	Highest action by CCode2 in dispute (see MID B codebook)
	cwhiactd	cwhiactd	Highest action in dispute (see MID A codebook)

	cwoutcm	cwoutcm	Overall dispute outcome (see MID A codebook)
	cwsettle	cwsettle	Overall dispute settlement (1=negotiated, 2=imposed, 3=none, 4=unclear, -9=missing)
	cfatald	cfatald	Overall dispute fatality level (0=none, 1=1-25 deaths, 2=26-100 deaths, 3=101-250 deaths, 4=251-500 deaths, 5=501-999 deaths, 6=>999 deaths, -9=missing)
	cwrecip	cwrecip	Reciprocated dispute (0=no, 1=yes)
	cwnumst1	cwnumst1	Number of states on CCode1's side
	cwnumst2	cwnumst2	Number of states on CCode2's side
	cwnmmdnw	cwnmmdnw	Number of new dyadic MIDs that begin in year with CCode1 and CCode2 on opposite sides. Note that this is dyadic; if creating dyadic MIDs did not find the 2 states to be in dispute, then the MID will not count.
	cwnmmdal	cwnmmdal	Total number of dyadic MIDs that begin or are ongoing in year with CCode1 and CCode2 on opposite sides
	cpceyrs	cpceyrs	Peace years (years without dispute), based on occurrence of disputes in COW-based dyadic set
	cpcedys	cpcedys	Peace days (days without dispute), based on occurrence of disputes in COW-based dyadic set, counted from end of prior MID to January 1 of the observation dyad-year
	cowrolea	cowrolea	CCode1's role in the dispute (computed only for MIDs covered in MID 3.0 data) per Maoz coding scheme (1:Primary Initiator; 2: Joiner on initiator side; 3: Primary target; 4: Joiner on target side).
	cowroleb	cowroleb	CCode2's role in the dispute (computed only for MIDs covered in MID 3.0 data) per Maoz coding scheme.
	mzongo	mzongo	A MID was ongoing at the beginning of this year, based on Maoz dyadic data set, plus MID 3.0 dyadic data (0=no, 1=yes)
	mzongonm	mzongonm	Number of the MID that was ongoing at the beginning of the year, if any (0 if no MID was ongoing)
	mzinit		MID Initiation: CCode1 initiated a MID vs. CCode2 in this year (0=no, 1=yes). Depending on settings for ongoing disputes and joiners, this may be a new MID or the 2 nd (3 rd , etc.) year of an ongoing MID and may include only dispute originators or MID joiners.
	mzinitnm		Number of the new MID that was initiated in this year, if any (0 if no MID was ongoing)
		mzmid	MID Onset: A MID occurred between CCode1 and CCode2 in this year (0=no, 1=yes). Depending on settings for ongoing disputes and joiners, this may be a new MID or the 2 nd (3 rd , etc.) year of an ongoing MID, and may include only dispute originators or MID joiners.
		mzmidnm	Number of the new MID that began in this year, if any (0 if no MID was ongoing)
	mzhost1	mzhost1	Relevant hostility level reached by CCode1 in a MID vs. CCode2 in this year (0=No hostility, 1=No militarized action, 2=Threat to use force, 3=Display of force, 4=Use of Force, 5=War)
	mzhost2	mzhost2	Relevant hostility level reached by CCode2 in a MID vs. CCode1 in this year (0 to 5, as in hostlev1)
	mzhostd	mzhostd	Overall dispute hostility level (0 to 5, as in hostlev1)
	mzkeynum	mzkeynum	Number of the MID for which all other MID variables are reported. Depending on user settings, this could be the number of the first MID that occurs in year, or the most serious MID, or the number of an ongoing MID. Referred to as the "key MID."
	mzorig1	mzorig1	CCode1 is originator of the key MID, i.e. is involved on first day (0=no, 1=yes)
	mzorig2	mzorig2	CCode2 is originator of the key MID (0=no, 1=yes)
	mzjoanyi		In this directed dyad, in <u>any</u> MID in this year, one of the states was a latecomer, and ccode1 was on the <u>initiating</u> side of the

			MID (0=no, 1=yes).
	mzjoanyt		In this directed dyad, in <u>any</u> MID in this year, one of the states was a latecomer, and ccode1 was on the <u>target</u> side of the MID
	mzjomidi		In this directed dyad, in <u>the MID number reported</u> under variable “mzkeynum” in this year, one of the states was a latecomer, and ccode1 was on the <u>initiating</u> side of the MID (0=no, 1=yes)
	mzjomidt		In this directed dyad, in <u>the MID number reported</u> under variable “mzkeynum” in this year, one of the states was a latecomer, and ccode1 was on the <u>target</u> side of the MID (0=no, 1=yes)
		mzjoany	In this dyad, in <u>any</u> MID in this year, one of the states was a latecomer
		mzjomid	In this dyad, in <u>the MID number reported</u> under variable “mzkeynum” in this year, one of the states was a latecomer
	mzmidnme	mzmidnme	Name of the MID identified in “mzkeynum,” if one has been designated by COW.
	mzstmo1	mzstmo1	MID start month for CCode1
	mzstday1	mzstday1	MID start day for CCode1
	mzstyr1	mzstyr1	MID start year for CCode1
	mzstmo2	mzstmo2	MID start month for CCode2
	mzstday2	mzstday2	MID start day for CCode2
	mzstyr2	mzstyr2	MID start year for CCode2
	mzendmo1	mzendmo1	MID end month for CCode1
	mzenddy1	mzenddy1	MID end day for CCode1
	mzendyr1	mzendyr1	MID end year for CCode1
	mzendmo2	mzendmo2	MID end month for CCode2
	mzenddy2	mzenddy2	MID end day for CCode2
	mzendyr2	mzendyr2	MID end year for CCode2
	mzsideA1	mzsideA1	CCode1 is on initiating side of mzkeynum (0=no, 1=yes)
	mzsideA2	mzsideA2	CCode2 is on initiating side of mzkeynum (0=no, 1=yes)
	mzrevis1	mzrevis1	CCode1 is a revisionist state in mzkeynum (0=no, 1=yes)
	mzrevis2	mzrevis2	Code2 is a revisionist state in mzkeynum (0=no, 1=yes)
	mzrevt11	mzrevt11	Revision Type 1 for CCode/state 1 (0=not applicable, 1=territory, 2=policy, 3=regime/government, 4=other)
	mzrevt21	mzrevt21	Revision Type 2, State 1, when present in COW v3.0 data.
	mzrevt12	mzrevt12	Revision Type 1, State 2
	mzrevt22	mzrevt22	Revision Type 2, State 2, when present in COW v3.0 data.
	mzfatal1	mzfatal1	Fatality level for CCode1 (0=none, 1=1-25 deaths, 2=26-100 deaths, 3=101-250 deaths, 4=251-500 deaths, 5=501-999 deaths, 6=>999 deaths, -9=missing)
	mzfatex1	mzfatex1	Exact estimate of fatalities for CCode1, if estimated in MID v3.0.
	mzfatal2	mzfatal2	Fatality level for CCode2 (0 to 6, as in fatal1)
	mzfatex2	mzfatex2	Exact estimate of fatalities for CCode2, if estimated in MID v3.0.
	mzhiact1	mzhiact1	Highest action by CCode1 in dispute (see MID B codebook) (adjusted to match MID 3.0 codings)
	mzhiact2	mzhiact2	Highest action by CCode2 in dispute (see MID B codebook) (adjusted to match MID 3.0 codings)
	mzhiactd	mzhiactd	Highest action in dispute (see MID A codebook) (adjusted to match MID 3.0 codings)
	mzoutcm	mzoutcm	Overall dispute outcome (see MID A codebook) (adjusted to match MID 3.0 codings)
	mzsettle	mzsettle	Overall dispute settlement (1=negotiated, 2=imposed, 3=none, 4=unclear, -9=missing)
	mzfatald	mzfatald	Overall dispute fatality level (0=none, 1=1-25 deaths, 2=26-100 deaths, 3=101-250 deaths, 4=251-500 deaths, 5=501-999 deaths, 6=>999 deaths, -9=missing)

	mzrecip	mzrecip	Reciprocated dispute (0=no, 1=yes)
	mznumst1	mznumst1	Number of states on CCode1's side
	mznumst2	mznumst2	Number of states on CCode2's side
	mzcowwar	mzcowwar	COW war dyad? (0=no, 1=yes) (Maoz data) Reflects actual warring dyads in multilateral disputes.
	mzduridx	mzduridx	Current dyad-year dispute duration index (counts 1, 2, 3, etc., for 1 st , 2 nd , 3 rd , etc., year) (Maoz data)
	mzdurday	mzdurday	Number of dispute days during the year. (Maoz data)
	mznmmdnw	mznmmdnw	Number of new MIDs that begin in year with CCode1 and CCode2 on opposite sides, based on start of new disputes in Maoz-based dyadic set plus MID 3.0 data.
	mznmmdal	mznmmdal	Total number of MIDs that begin or are ongoing in year with CCode1 and CCode2 on opposite sides, based on start of new disputes in Maoz-based dyadic set plus MID 3.0 data.
	mzpceyrs	mzpceyrs	Peace years (years without dispute), based on occurrence of disputes in Maoz-based dyadic set plus MID 3.0 data.
	mzrolea	mzrolea	CCode1's role in the dispute, from Maoz data and computed for MIDs covered in MID 3.0 data, per Maoz coding scheme (1:Primary Initiator; 2: Joiner on initiator side; 3: Primary target; 4: Joiner on target side).
	mzroleb	mzroleb	CCode2's role in the dispute, Maoz data and computed for MIDs covered in MID 3.0 data, per Maoz coding scheme.
	icbongo	icbongo	There is an ongoing ICB crisis between CCode1 and CCode2 in year (0=no, 1=yes).
	icbongnm	icbongnm	ICB crisis-level ID number corresponding to the ongoing crisis in year.
	crisis	crisis	There is a new ICB crisis between CCode1 and CCode2 in year (0=no, 1=yes).
	crisno	crisno	ICB crisis-level ID number corresponding to the new crisis in year.
	crisname	crisname	Crisis name.
	crdynam	crdynam	Unique crisis dyad number.
	oneside	oneside	Indicates whether the crisis dyad is one-sided or not. See dyadic ICB documentation for details.
	trgyrdy	trgyrdy	Dyadic trigger year. For two-sided crises, the dyadic trigger date is set to the later of the two trigger dates for the two dyad members. For one-sided dyads, dyadic trigger and termination dates are based solely on the trigger and termination dates for the sole crisis actor. See dyadic ICB documentation for details.
	trgmody	trgmody	Dyadic trigger month.
	trgdady	trgdady	Dyadic trigger day.
	trmyrdy	trmyrdy	Dyadic termination year. The dyadic termination date is set to the earlier of the two termination dates for the actors. See dyadic ICB documentation for details.
	trmmody	trmmody	Dyadic termination month.
	trmdady	trmdady	Dyadic termination day.
	durdays	durdays	Dyad duration, in days.
	duryear	duryear	Number of calendar years spanned by dyad.
	yrtriga	yrtriga	Year of crisis trigger for state A. When state A is a non-crisis actor, YRTRIGA is set to -99. See dyadic ICB documentation for additional details.
	motriga	motriga	Month of crisis trigger for state A. When state A is a non-crisis actor, MOTRIGA is set to -99. See dyadic ICB documentation for additional details.
	datriga	datriga	Day of crisis trigger for state A. When state A is a non-crisis actor, DATRIGA is set to -99. See dyadic ICB documentation for additional details.
	yrtrigb	yrtrigb	Year of crisis trigger for state B.
	motrigb	motrigb	Month of crisis trigger for state B.

	datrigb	datrigb	Day of crisis trigger for state B.
	yrterma	yrterma	Year of crisis termination for state A. When state A is a non-crisis actor, YRTERMA is set to -99. See dyadic ICB documentation for additional details.
	moterma	moterma	Month of crisis termination for state A. When state A is a non-crisis actor, MOTERMA is set to -99. See dyadic ICB documentation for additional details.
	daterma	daterma	Day of crisis termination for state A. When state A is a non-crisis actor, DATERMA is set to -99. See dyadic ICB documentation for additional details.
	yrtermb	yrtermb	Year of crisis termination for state B.
	motermb	motermb	Month of crisis termination for state B.
	datermb	datermb	Day of crisis termination for state B.
	actnuma	actnuma	ICB actor-level sequence number (variable CRACNO in the actor-level ICB data set) for state A. This ID number corresponds to the actor's case number in the actor-level ICB data set. If state A is not considered a crisis-actor by ICB, then ACTNUMA equals 9999.
	actnumb	actnumb	ICB actor-level sequence number (variable CRACNO in the actor-level ICB data set) for state B.
	cowmema	cowmema	Indicates the interstate system membership status of state A according to the Correlates of War Project's criteria. See dyadic ICB documentation for details.
	cowmemb	cowmemb	Indicates the interstate system membership status of state B according to the Correlates of War Project's criteria.
	gwmema	gwmema	Indicates the interstate system membership status of state A according to Gleditsch and Ward's (1999) criteria. See dyadic ICB documentation for details.
	gwmemb	gwmemb	Indicates the interstate system membership status of state B according to Gleditsch and Ward's (1999) criteria
	iwca	iwca	The ICB actor-level coding for whether the crisis for state A was an intra-war crisis. See dyadic ICB documentation for details.
	iwcb	iwcb	The ICB actor-level coding for whether the crisis for state B was an intra-war crisis.
	icbjoina	icbjoina	CCode1 is a joiner (was not involved on the trigger day of the dyadic crisis) into the crisis identified under crisno.
	icbjoinb	icbjoinb	CCode2 is a joiner (was not involved on the trigger day of the dyadic crisis) into the crisis identified under crisno.
	icbnmnew	icbnmnew	Number of new ICB crises that begin in year with CCode1 and CCode2 on opposite sides.
	icbnmall	icbnmall	Number of new and continuing ICB crises in year with CCode1 and CCode2 on opposite sides.
	icpceyrs	icpceyrs	Peace years (years without crisis), based on occurrence of ICB crises.
	User Variables		Any user data set variables selected

Polity III Merging, Country Code Recoding, and Notes

The Gurr et al. Polity III data set does not perfectly match up with the COW data set codings on variables in a couple of instances. As a result, some minor modifications must be made to the Polity III data before it can be used in *EUGene*. These changes are as follows:

1. COW codes one state of Austria-Hungary (country code 300) until 1919, but Polity III codes Austria and Hungary as separate states both before and after 1919, with country codes 305 and 310 respectively. An inspection of the Polity III data reveals that Austria and Hungary have identical values on all polity III variables for before 1919. *EUGene* recodes

ccode 305 and 310 from the Polity III data into ccode 300 before 1919, and before 1919 only outputs data on Austria-Hungary.

2. COW and Polity III split Germany into two states after 1945, from a single ccode of 255 into ccodes 260 and 265. However, after the reunification of Germany in 1990, COW reverts to the unified Germany code of 255, while Polity III continues to code Germany as ccode 260. The Polity III coding of 260 is changed back to 255 after German reunification. So from 1990 onward in *EUGene* output, ccode 255 again begins to be reported for Germany, and only ccode 255 is reported.

3. COW codes pre-WWI Serbia and Yugoslavia as the same country code, 345. Polity III codes Croatia as ccode 344 up to 1915, and then Yugoslavia from 1921 forward. No 344 is coded after 1915 (until 1991); no 345 is coded before 1921. We use the country code 345 for both of these up to 1915, so only ccode 345 is coded from 1800 to 1990.

4. COW codes pre-unification Italy and Italy 1861+ as the same country code, 325. Polity III codes Sardinia as ccode 324 up to 1860, and then Italy from 1861 forward. No 324 is coded after 1861; no 325 is coded before 1861. We use the country code 325 for both of these, pre and post 1861.

Note also that Polity III data using 4 different values for missing data points (country-years). A -99 represents data that is truly missing, while -66 through -88 represent various types of polity interruptions. When *EUGene* calculates computed variables such as “dem” that require two valid inputs, all values from -66 through -99 are treated as missing, and so the computed variable will be missing if any necessary subcomponent has a -66 to -99 value. However, *EUGene* will output the actual values -66, -77, and -88 to the output file. In addition, in the SPSS, STATA, and LIMDEP command files that *EUGene* creates, all values from -66 through -99 are coded as missing values for the software. If the user wishes to make use of -66, -77, and -88 cases in some other fashion, he or she must customize the input command file accordingly.

Creating Dyadic MIDs and Meshing MID Data Sets

Typically the information contained in the dispute variables will be used to create a dependent variable for analysis. Typically variables for the initiation or onset of disputes along with the hostility level variables will be used in this regard. However, it is important that users understand how these dependent/dispute variables are coded, as MIDs are not coded for every dyad involved in a militarized dispute. For any given overall MID, there may be many more dyads of states involved in conflictual interactions. Getting from MIDs to dyadic MIDs requires careful data conversion in many cases.

Variations on the MID data sets

Prior to version 3.0, which begins in 1993 (with a few disputes backdated to 1992), Correlates of War militarized interstate dispute data has not been provided in a dyadic form. Rather, versions through MID 2.1 (1816-1992) provide for each (bilateral or multilateral) dispute a list of participants with start and end dates for each, the highest level of hostility reached by each, and the side of the dispute on which each participated. To use this 1816-1992 COW data it is necessary to convert the MIDs into a dyadic form, creating variables marking dispute onset and hostility levels between pairs of states. Our method for doing this conversion is detailed below.

Zeev Maoz released a data set (the “DyadMID” data set) which begins with conversions such as we carry out, but then also examines the historical record of several MIDs to try to ascertain which dyads were in fact engaged in direct hostilities. This dataset does not presuppose that all of the actors on one side confronted all actors on the other side. Maoz has examined each possible dyad in a MID and created the set to include only the dyads that

actually confronted one another in a dispute with appropriate time periods. Both our converted version, and Maoz’s data set, are available within EUGene.

The COW MID data set version 3.0 includes a dyadic version as part of the main COW data. This data set begins in 1993, but disputes that were ongoing on January 1 1993 were backdated with version 3.0 coding methods. Thus version 3.0 contains a dyadic version of all MIDs that were ongoing on or began after January 1, 1993, through the end of 2001.

Meshing COW MID Data, Maoz’s Dyadic MID Data, and MID 3.0 data

EUGene version 3.0 meshes the COW converted data, Maoz’s dyadic data set, and the COW 3.0 data in a manner intended to allow users to ignore differences between them. EUGene allows users of MID variables to select either “COW MIDs” or “Maoz MIDs” under the conflict variables section of the program. If the user selects “COW MIDs,” then data from 1816-1992 will be taken from the EUGene conversion of the COW MID data to dyadic form. If the user selects “Maoz MIDs,” then data from 1816-1992 will be taken from Maoz’s dyadmid set. **In either case, data from 1993 forward will be taken from the COW dyadic version of the MID 3.0 data.** A single set of MID variables will be output from 1816-2001, but with component data taken from the appropriate source. For instance, if the user selects “MID Onset” as a variable in a nondirected dyad analysis and selects the Maoz dyadic data version, then a single onset variable will be included in EUGene’s output, but from 1816-1992 the data will come from Maoz’s dyadmid data set, and from 1993 forward it will come from the COW v3.0 dyadic MID data. For cases that were ongoing on January 1, 1993, and so began in 1992, the COW v3.0 version of the case is used.

The MID 3.0 data contains some new variables that were not previously available. Values on these variables before 1993 (or 1992 in the case of some of the MIDs that were ongoing on January 1993) are missing.

Adjusting “Highest Action” level from Maoz dyadic data

One important variable within the Maoz set has been recoded in EUGene’s output to be consistent with the newer v3.0 MID data. In Maoz’s dyadic data set, the highest action type taken against a target state uses the older COW 2.1 coding scheme, which ranges from 1 to 23 for actions ranging from “no action” to “join interstate war.” In the MID 3.0 set, this variable takes values ranging from 0 to 21. Values on the “highest action” variable from the Maoz data have been updated to the COW 3.0 scheme. The meaning of the action level values are as follows (the hostility level corresponding to each action type are in brackets []):

MID v 2.1 coding (Original Maoz set)		MID v3.0 coding (EUGene v3.03+ output)	
Action Code	Behavior	Action Code	Behavior
		0	No militarized action [1]
1	No militarized action [1]	1	Threat to use force [2]
2	Threat to use force [2]	2	Threat to blockade [2]
3	Threat to blockade [2]	3	Threat to occupy territory [2]
4	Threat to occupy territory [2]	4	Threat to declare war [2]
5	Threat to declare war [2]	5	Threat to use CBR weapons [2]
6	Threat to use nuclear weapons [2]	6	Threat to join war
7	Show of troops [3]	7	Show of force [3]
8	Show of ships [3]	8	Alert [3]
9	Show of planes [3]	9	Nuclear alert [3]
10	Alert [3]	10	Mobilization [3]

11	Nuclear alert [3]	11	Fortify border [3]
12	Mobilization [3]	12	Border violation [3]
13	Fortify border [3]	13	Blockade [4]
14	Border violation [3]	14	Occupation of territory [4]
15	Blockade [4]	15	Seizure [4]
16	Occupation of territory [4]	16	Attack [4]
17	Seizure [4]	17	Clash [4]
18	Clash [4]	18	Declaration of war [4]
19	Raid [4]	19	Use of CBR weapons [4]
20	Declaration of war [4]	20	Begin interstate war [5]
21	Use of CBR weapons [4]	21	Join interstate war [5]
22	Interstate war [5]		
23	Joins interstate war [5]		

Converting COW MID Data to Directed Dyads - pre 1993 data

For pre-1993 (pre version 3.0) MID data, EUGene computes dispute dyads by initially computing a dyad record involving each state on side A with each state on side B for each year that the dispute was occurring. The software then checks the dates that each state is involved in the dispute to ensure that they were involved on opposite sides at the same time; only dispute-dyad-years in which both states were involved in the dispute on opposite sides are considered dispute-dyads. We thus code every pair of states that cross the side A / side B boundary as dyads involved in the dispute. We assume that all of the states on each side confronted all of the states on the other side in some manner. Similarly, we code as dispute initiations all of the states involved on the first day of the dispute on side A against all of the states involved on the first day of the dispute on side B (see the next section for possible modifications about the identity of initiators). NOTE: To make these conversions, some data is used from the MID A data set, while other data comes from MID B data set; this data is merged to obtain all MID output.

Creating dyadic MIDs – 1993+ data (MID 3.0)

The COW project has used customized routines created within *EUGene* to create the COW dyadic MID data set. To create dyadic MIDs for version 3.0, EUGene uses incident level data from the MID 3.0 data set, coupled with higher level MID information. In particular, EUGene uses the participant-incident level file to note exactly what pairs of states engaged in incidents against one another. Only those states that were engaged on opposite sides of at least one incident in a MID are coded as members of a dyadic MID. [See the COW project website at <http://cow2.la.psu.edu> for more information on the definition of incidents, the creation of the MID 3.0 dyadic data set, and particular coding rules for generating start and end dates and hostility levels for the dyadic MID level.]

Creating Dyadic MID Variables

The core of the dyadic MID 3.0 data set is the identification of pairs of states actually engaged in MID behavior. In addition, several variables in the dyadic MID data are adjusted for the dyadic nature of the interaction, and may show different values within a dyad. Some variables cannot be converted to a specifically dyadic form, however, and these variables take on the same value as the overall MID for each dyadic MID.

Variables adjusted for dyadic interactions

Dyadic MID Number: In addition to the number of the overall MID, v3.12 added calculation of a dyadic MID number, which has a separate number for each dyad (of states on opposite sides, in regular dyad-MID interaction) within a MID. There is also a different number for the different interactions that occur when a state in a MID exits and reenters the MID. When this happens, it is a new dyadic MID, with a new number. Note that the coding of an ongoing dispute depends on this dyadic MID number (so if a state exits and reenters a dispute, with a corresponding new dyadic MID number, it is considered a new MID onset and not an ongoing MID).

InSideA: This marker identifies which state in the dyad took the first militarized action against the other state in the dyad, that is, the first mover in the dyadic MID.

Start and end date of the dyadic MID: taken as the start date of the first incident in the MID, and the last day of the last incident in the MID, involving the two states in the dyad on opposite sides of any incident within the overall MID.

Highest hostility level and highest action: reflect the hostility levels reached only across those incidents where the two states in a dyad were on opposite sides.

Revisionist: The revisionist variable marks those states who were revisionist in incidents involving the other state in the dyad. That is, a state in a dyadic MID is coded revisionist if the state, in any incident involving the two states in the dyad, has a non-0 value for some issue. So a state is non-revisionist if and only if it has no issues in any incident involving it and the other state in the dyad during the MID. States could be revisionist in their aims against some states, but not others.

Revision type: Reflects revisions sought only in those incidents involving the other state in the dyad. Since only two revision types may be coded, and there may be many incidents, *EUGene* keeps the 2 most serious revision types (ranked as 1 > 2 > 3 > 4) across all incidents involving the two states in the dyad.

Reciprocation: Reciprocation is marked when both states take a directed incident-level action against the other state in the dyad.

Unadjusted variables (from overall MID/state information)

SideA: The SideA variable marks states who were on the side of the overall first-mover in the MID.

Fatalities (state level): reported fatalities for states in dyadic disputes are simply the fatalities that come from the overall dispute. They are NOT adjusted for dyadic interaction (there is no way to accurately garner in a multi-state engagements who inflicted casualties on whom). No new fatalities are computed in dyadic MIDs, and fatalities are not divided in any way among dyadic MIDs. Thus the same value will appear for a state in all dyadic MIDs that emerge from a single overall MID. Users using the fatality variable should use caution and use only one value for each overall MID, as adding state-level fatality values from the dyadic MID level will overcount fatalities. [If the user selects the Maoz version of dyadic MIDs for the pre-1993 period, Maoz has adjusted some fatalities for the dyadic disputes. Post 1992 dyadic MIDs do not have an adjustment.]

Fatalities (overall): Similar to state level fatalities, overall fatalities represent the total fatalities of the overall MID, not divided by state or dyad. The same value will appear for all dyadic MIDs within an overall MID.

Originator: Reports states that were on the originating side in the overall MID.

Number of states in MID: Reports the total number of states involved on each side in the overall MID, not necessarily against the particular target in a dyadic MID.

Name: Comes from overall MID coding.

Outcome: Comes from overall MID coding. Not adjusted for dyadic outcome. [If the user selects the Maoz version of dyadic MIDs for the pre-1993 period, Maoz has adjusted some outcomes for the dyadic disputes. Post 1992 dyadic MIDs do not have an adjustment.]

Settlement: Comes from overall MID coding. Not adjusted for dyadic outcome. [If the user selects the Maoz version of dyadic MIDs for the pre-1993 period, Maoz has adjusted some settlements for the dyadic disputes. Post 1992 dyadic MIDs do not have an adjustment.]

Other notes

For the role variable, if a state switches sides in a MID, then “role” is reported for the first involvement in the MID.

Excluding or Including Problematic Cases

EUGene creates data sets that may include or exclude cases based on a variety of settings made under the “Case/Conflict Exclusions” tab of the main settings window. Under this tab, you may select one (and only one) conflict data set on which to base case exclusions (although you may choose to output conflict variables from multiple conflict sets).

By default, *EUGene* will output data on directed dyads only 1) when there is no ongoing MID at the beginning of the year, and 2) when the reverse directed dyad does NOT have an initiation in it, and 3) when the states in the dyad are not joiners. We also code dispute initiation in the default case only when an initiation is a new initiation in a dyad. We exclude certain cases by default because we believe that they are problematic in terms of logical and empirical equivalency in relation to most dyads. However, we provide users with a large number of options to include or exclude a range of cases that we feel are problematic.

If the user does not choose to output MIDs on the variable page, then none of the following options will be available, and all cases – that is, all relevant dyads and years given the chosen subset of years and population of cases – will be output.

Ongoing Dispute Year Options

Note that these rules apply only to output created with the dyad-year as the unit of analysis. When outputting the dispute-dyad or dispute-dyad-year, all dispute initiations are included in the output, whether there was an ongoing dispute or not. Note that in the case of dyadic MIDs created because a state exited and reentered the MID, each re-entry is a new MID, and is not considered ongoing.

Include All Dyads with an Ongoing MID

Output will include one line of data for all directed-dyad-years, even if there is a MID ongoing between the two states in the dyad at the beginning of the year.

Drop All Dyads with an Ongoing MID

Output will never include data for directed-dyad-years in which there was a MID ongoing between the two states at the beginning of the year, even if there is a new MID initiated by one of the states against the other during the year.

Include Ongoing Dispute Dyad Year iff New Dispute

Output will include a directed-dyad-year either if 1) there is no ongoing dispute at the beginning of the year, or 2) if there is an ongoing dispute but there was also a new, different dispute initiation in that year. This option ensures that all dyad-years with a dispute initiation are included in the output. However, note that if this option is selected, then only some directed dyads of those new disputes are included. For example, if there is an ongoing dispute between country code 2 and 200 at the beginning of 1838 (and so 1838 would usually

not be included in the output file), and there was a new initiation by 2 vs. 200 in 1838, then only dyad-year 2 vs. 200 in 1838 is included in the data. Directed dyad 200 vs. 2 in 1838 would not be included in the output, as this was an ongoing dispute directed-dyad year without a new initiation.

Statistical/theoretical note: the decision to include any ongoing dispute years should be made carefully. In years with an ongoing dyadic dispute, the involved states may not have an opportunity to make a decision to initiate a MID. As a result, including such years in an analysis and coding that no dispute was initiated may actually be including a case with irrelevant information, since there is no opportunity (or, if the ongoing dispute ends during the year, less opportunity) for an initiation. Coding an ongoing dispute year as a new initiation has its own conceptual problems (see below). Concerning the second option, to include ongoing dispute years if and only if there is a new initiation, it is clear in such years that there was an opportunity for a new initiation. However, such an opportunity may have been present in other years with an ongoing dispute, namely when the dispute ended in that year, but we cannot be certain of this.

Treating Ongoing Dispute Years as Initiations

Some analyses of international relations have treated subsequent years of a single ongoing militarized dispute as if they were new conflict initiations. We disagree, and by default *EUGene* codes dispute initiation only when an initiation is a new initiation in a dyad-year. That is, *EUGene* normally codes the output variable “initiation” as a “1” only in the first (initial) year of a MID; subsequent dyad-years would be coded as a “0” for initiation. [As discussed separately, subsequent dyad-years which have an ongoing dispute at the beginning of the year are also dropped, unless you specify to report them in the output.] We believe that the decision to *continue* a MID for a period of time may be quite different from the decision to *initiate* a MID. If the user wishes to treat ongoing years of the same dispute as if they were new dispute initiations rather than a situation requiring a different model, the user can mark “Treat Ongoing Dispute Years as Initiations.” If this option is not marked, then ongoing years of the same dispute will receive a “0” on the initiate variable. If a different dispute starts in the same year that a different dispute was ongoing, though, the initiate variable will still be marked as a “1”.

Statistical note: treating ongoing dispute years as new initiations this has the effect of increasing the effects of multi-year disputes on results of statistical analysis, and conflates decisions to continue a dispute with decisions to initiate disputes. It may be that the choice to continue an ongoing dispute is a different choice problem than the decision to initiate a new dispute.

Target vs. Initiator Dyads

When one state initiates a MID vs. another, an initiator and a target are designated. If an initiator A starts a dispute against a target B, then it may be problematic to include the directed dyad B vs. A for analysis. The reason is that when A initiates vs. B, it may remove B’s ability to initiate a dispute against A in that same year (the B vs. A dyad is censored). *EUGene* provides options to include or exclude this reverse direction if desired.

More fully, in some dyad years, there is a dispute initiation by state A against state B, and an initiation by state B against state A. This indicates that whichever dispute was first ended in time for a second dispute (with the reverse direction) to be initiated. However, it is usually the case that when state A initiates a dispute against state B, we do not observe state B initiating a dispute against A in the same year. There may be two reasons for this. First, it may be the case that state B has chosen not to initiate a dispute against A, even though it had opportunity to do. Second, however, it may be that when state A initiates a dispute against state B, this takes away state B’s opportunity to initiate a dispute against A for the year. Especially if the dispute by A against B continues through the end of the year, then B does not have the same opportunity for dispute initiation against A that it would have if B had not

initiated. Essentially, if A initiates against B, the observation of initiation of B against A may be *censored*. Even though conditions may have been perfect for B to initiate against A, because A initiated first, we could not observe the initiation. Given this problem, we have three possible combinations of initiation in a dyad year:

A initiates against B, and B initiates against A: In this case, both A and B had opportunity for dispute initiation and took advantage of it.

A does not initiate against B, and B does not initiate against A: In this case, both A and B had opportunity for dispute initiation but neither took advantage of it.

A initiates against B, but B does not initiate against A: In this case, there are two possibilities: 1) both A and B had opportunity for dispute initiation, but only A took advantage of it, or 2) A had the opportunity for initiation and took it, but B did not have the same opportunity.

In the first two cases it is clear that both directed-dyad years should be included in a data set for analysis, because we fully observe the decisions about initiation/non-initiation made by the two states. However, in the third case it is less clear how to include cases. It is clear that the A vs. B directed dyad should be included, because we fully observe A's initiation vs. B. But we do not know how to code the "initiation" variable for the B vs. A dyad because it is censored. If the case is included in the data set with "initiation" coded as a 0, indicating no initiation, it may include cases in the data that overstated B's willingness to accept the status quo and understated its desire to initiate a dispute. On the other hand, if the case is omitted, we may be omitting cases where there truly was no desire on B's part to initiate a dispute, and so we are discarding some information that is relevant for our analysis.

Drop Target vs. Initiator Directed Dyads if no new MID

By default, *EUGene* opts for DISCARDING the reverse dyad of B vs. A in years where A initiated a dispute against B but where we do not observe a B vs. A initiation. Cases where both A and B initiated, or where neither A nor B initiated, ARE included in the directed-dyad output.

Keep Target vs. Initiator Directed Dyads if no new MID

If the user wants to include all B vs. A dyads (with a "0" coded for "initiation") in cases where A initiated a dispute against B but where we do not observe a B vs. A initiation, the user should mark the "Keep Target vs. Initiator Directed Dyads if no new MID" check box.

Joiners

A "Joiner" is considered to be any state who enters a MID after the first day. Note that because of the way the MID data is set up, it is possible for "joiners" to be voluntary joiners who choose to intervene on one side of a MID, or they may actually be targets of the expansion of a MID. "Joiners" really means "latecomers" and should not be read to imply a voluntary choice to become involved.

Drop all Joiner Dyads

By default, *EUGene* drops directed dyads A vs. B in the data set if A is a joiner into a dispute (on either the initiating or target side) against side B. The argument for dropping joiners is that the process of making a decision to join an ongoing dispute is not the same as making a decision to initiate a dispute.

Include all Joiner Dyads

If this option is marked, then all directed-dyad years where one or both states are joiners will be included in the output data. By default (see codings for "**Error! Reference source not found.**") these directed dyad-years are NOT marked as dispute initiations, but they ARE

included in the data set (hence they will have a “0” on the “initiate” variable, and would typically be considered status quo outcome directed dyad-years, although if you mark an appropriate option under the “treat joiners as initiators” section then some or all joiners will be marked as initiators as well). Note that the inclusion of joiners is handled slightly differently for the directed-dyad year data and for directed-dispute-dyad year data. For outputting directed-dispute-dyads, including joiners will include dyads where the first state was a joiner on the initiating side. When outputting directed dyads, if you choose to include joiners, then all dyads where state 1 was a joiner on either the initiating or target side will be included in the output.

Joiner Variables

As discussed previously, checking “Mark Joiners” under a MID variable page will output several variables concerning joining. In directed dyad output, variables *cwjomidi* and *cwjomid* mark whether, if this is a joining dyad (where at least one state is a joiner in the key MID), state A was on the initiating side or target side against B in either the key dispute. That is, this will mark dyads where *ccode1* was on the initiating side against *ccode2*, but one of the two states was not an originator (was not involved on day 1), or where A was on the target side against B but again where one state was not an originator. Variables *cwjoanyi* and *cwjoanyt* mark whether state A was on the initiating (or target) side against state B in the given year and where at least one was a joiner, in any MID in the given year. In nondirected output, *cwjoany* marks whether EITHER A or B is a joiner in ANY MID in the given year, while *cwjomid* marks whether EITHER A or B is a joiner in the key identified MID in the given year. Similar variables are output for the Maoz dyadmid conflict set.

Apparent Anomalies in Joiner Codings

Q: Seemingly anomalous cases exist where states are shown as being both an originator and a joiner. How is this possible? Aren’t these categories mutually exclusive?

In fact, these categories are not mutually exclusive. In particular, note the difference between the *cwjoanyi* and *cwjomidi* variables. The first reports whether in this directed dyad, in any MID in this year, one of the states was a latecomer, and *ccode1* was on the initiating side of the MID. The second reports information about joining in the key MID of the dyad. A state may join 1 MID, but originate another. In addition, in the current formulation the joining variables identify dyads where either state was a joiner, not just the first state of a dyad. So the behavior of both states in multiple MIDs may be relevant. Finally, the “i” and “t” designations in variables *cwjomidi* and *cwjomid* refer to whether *ccode1* was on the initiating or target side of the MID, and either of the states was a joiner, not that *ccode1* joined the initiators or targets. In some cases, *ccode1* could be listed as an originator and a joiner, if the TARGET in this directed dyad was a joining state (became involved late). In this case the dyad involves at least one joiner, but *ccode1* was in fact an originator.

As an example, if a state (e.g. France) initiates a conflict, and then later another state (e.g. Brazil) joins the target side of the dispute, the following things are true:

- ⇒ France is considered an originator (and initiator);
- ⇒ Brazil is NOT considered an originator (and is on the target side);
- ⇒ The dyad is considered a joining dyad (one state is a latecomer);
- ⇒ In the Brazil-France directed dyad, Brazil is listed as a joiner on the target side of the MID;
- ⇒ In the France-Brazil directed dyad, France is listed as a joiner on the initiating side of the MID.

The apparently odd final case is explained by rephrasing: In the France-Brazil directed dyad, the dyad is a joining dyad, and France was on the initiating side of the MID. In the Brazil-France directed dyad, the dyad is a joining dyad, and Brazil was on the target side of the MID.

Implications for variable construction

If you want a variable identifying the actual latecomers, so a variable that indicates “ccode1 was a latecomer and joined the initiating side of this MID” then stata code to create this variable would be:

```
gen ccode1_joined_initiators = 0
replace ccode1_joined_initiators = 1 if (cworig1==0 & cwjomidi ==1)
{Note that in Stata, “&” indicates AND, while “|” indicates OR }
```

Similarly, if you want to create a variable identifying an actual latecomer that was on the target side, so “ccode1 was a latecomer and joined the target side of this MID” then code to create this variable would be:

```
gen ccode1_joined_targets = 0
replace ccode1_joined_targets = 1 if (cworig1==0 & cwjomidt ==1)
```

Note that we cannot tell in directed-dyad-year output whether ccode2 was a joiner on side A or side B; cwjomidi and cwjomidt refer only to the side that ccode1 was on. So we cannot compute similar variables for “ccode2_joined_initiators” etc.

For case selection purposes, if you wish to exclude from analysis all dyads where either state was a joiner into the MID on which information is reported (the usual selection criterion), you should now use

```
drop if cwjomidi ==1 | cwjomidt ==1
```

This keeps only cases where both states were originators in the reported MID in this directed-dyad year, or there was no conflict (the status quo).

If for some reason you wish to keep cases where one of the states was involved on day 1 but one was a joiner (dropping cases of joining unless one was an originator), then the relevant command would be

```
drop if (cwjomidi ==1 | cwjomidt ==1) & ~( cworig1==1 | cworig2==1)
```

Relationship to other MID variables

By definition, it does not happen that there is both initiation and either cwjomidi or cwjomidt coded. That is, joining only happens in non-initiation dyads (if initiation is coded as both sides being originators). But in a small number of cases, dyad-year data suggest that there is both an initiation and cwjoanyi or cwjoanyt in the year, because of multiple MIDs occurring in a dyad during the year, one of which is initiation and another of which involves joining.

The originator status of state 2 should bear no relation to cwjomidi, cwjomidt. That is, the second state might be an originator, but the first state might have been a joiner onto side i or side t. The joiner variables only tell you about whether, if this was a “joining dyad,” that state 1 was on side i or side t.

Combining Include/Exclude Specifications

As is evident in this section, *EUGene* allows a large number of combinations of specifications on criteria such as including joiners or only originators, targets or only initiators, keeping or dropping ongoing years, and treating subsequent years of a dispute as new initiations. The possible interactions among these choices are not always obvious. In this section we present some of the ways in which these selections about joiners and targets interact in producing data. If you wish to be certain that you understand the data set that you are generating in terms of included and excluded cases, and in terms of when the dependent variable of dispute initiation is coded a “1,” you should examine this section.

Let’s assume that we are using a sample data set of disputes following the format of the v2.1 MID data set with the following information:

```

1, "USR", 365, 31, 5, 1816, 30, 3, 1818, 1, 1, 2, 6, 22, 5, -1, 2.1
1, "UKG", 200, 15, 6, 1816, 30, 3, 1818, 1, 0, 0, 6, 22, 5, 0, 2.1
1, "TUR", 640, 31, 5, 1816, 30, 3, 1818, 0, 0, 0, 6, 22, 5, -1, 2.1
1, "AUH", 300, 15, 6, 1816, 30, 3, 1818, 0, 0, 0, 0, 7, 3, 0, 2.1

```

This data set has only one dispute, dispute number “1.” These data indicate 2 states on each side of this hypothetical dispute. On side A, the initiating side, are Russia (“USR”) and the UK (“UKG”). On the target side are Turkey and Austria-Hungary. The dispute is initiated on day 1 by Russia against Turkey. The UK and Austria are joiners who become involved on a subsequent day.

Dyad-Year Output

Basic Output

Assume the following options are marked (in all cases, side A is treated as the initiator):

- treat only “true” initiators (originators) as initiators;
- drop target vs. initiator dyads if no new MID;
- include all joiner dyads.

Output is:

ccode1	ccode2	year	ongoing	initiate	hostlev1	hostlev2	midnum	joiner_i	joiner_t
300	640	1816	0	0	0	0	0	0	0
300	200	1816	0	0	3	5	1	0	1
300	365	1816	0	0	3	5	1	0	1
640	300	1816	0	0	0	0	0	0	0
640	200	1816	0	0	5	5	1	0	1
200	300	1816	0	0	5	3	1	1	0
200	640	1816	0	0	5	5	1	1	0
200	365	1816	0	0	0	0	0	0	0
365	300	1816	0	0	5	3	1	1	0
365	640	1816	0	1	5	5	1	0	0
365	200	1816	0	0	0	0	0	0	0

Only 365 vs. 640 in 1816 (Russia vs. Turkey) is coded as an initiation. 300 vs. 200, 300 vs. 365, and 640 vs. 200 are “joiner_t” dyads, that is, a dyad where the first state was a joiner in a dispute against the second, and that first joiner state (300, 300, and 640 respectively) was on the target side in the dispute. However, note that they are NOT initiations (because only “true” initiators (originators) are being considered initiators). Similarly, 200 vs. 300, 200 vs. 640, and 365 vs. 300 are “joiner_i” dyads but not initiations because these dyads did not dispute on day 1, but the states were subsequently on opposite sides with states 200 and 365 on the initiating side. Since there is no dispute between them, 300 vs. 640, 640 vs. 300, 200 vs. 365, and 365 vs. 200 are all status quo dyads with no initiation and no hostility level > 0. Finally note that there is NO case listed for 640 vs. 365 because 640 is the target of an initiation, and the specifications are to drop target vs. initiator dyads without a new MID.

Keep Targets

If the specification is changed to keep target vs. initiator dyads:

- treat only “true” initiators (originators) as initiators;
- keep target vs. initiator dyads if no new MID;
- include all joiner dyads

then 1 case is added to the output (all other cases remain the same). Note that in this case, though, the initiate variable is a 0. Even though 640 vs. 365 (Turkey vs. USSR) is included in the output data set, Turkey did not initiate vs. USSR. If you wish to analyze a dependent variable such as dispute ONSET (rather than the directed form, initiation), then you would code your own dependent variable using the hostility level variables, which do show that there was significant hostility between Turkey and Russia, regardless of the direction of initiation. The added case is:

cocode1	cocode2	year	ongoing	initiate	hostlev1	hostlev2	midnum	joiner_i	joiner_t
640	365	1816	0	0	5	5	1	0	0

Drop Joiners

If the specification is changed to drop both joiners and target vs. initiator dyads:

- treat only “true” initiators (originators) as initiators;
- drop target vs. initiator dyads if no new MID;
- drop all joiner dyads.

then all of the cases that involved states joining a dispute are dropped, leaving a smaller set of data saved to the output file:

cocode1	cocode2	year	ongoing	initiate	hostlev1	hostlev2	midnum	joiner_i	joiner_t
300	640	1816	0	0	0	0	0	0	0
640	300	1816	0	0	0	0	0	0	0
200	365	1816	0	0	0	0	0	0	0
365	640	1816	0	1	5	5	1	0	0
365	200	1816	0	0	0	0	0	0	0

Coding Joiners As Initiators

If the specification is changed to drop both joiners and target vs. initiator dyads:

- treat joiners and originators as initiators;
- keep target vs. initiator dyads if no new MID;
- keep all joiner dyads.

cocode1	cocode2	year	ongoing	initiate	hostlev1	hostlev2	midnum	joiner_i	joiner_t
300	640	1816	0	0	0	0	0	0	0
300	200	1816	0	1	3	5	1	0	1
300	365	1816	0	1	3	5	1	0	1
640	300	1816	0	0	0	0	0	0	0
640	200	1816	0	1	5	5	1	0	1
640	365	1816	0	0	5	5	1	0	0
200	300	1816	0	1	5	3	1	1	0
200	640	1816	0	1	5	5	1	1	0
200	365	1816	0	0	0	0	0	0	0
365	300	1816	0	1	5	3	1	1	0
365	640	1816	0	1	5	5	1	0	0
365	200	1816	0	0	0	0	0	0	0

Note that now, ALL of the joiner dyads are coded as a “1” on the “initiate” variable. If you wanted to only consider the directed dyad from the initiator to target side as actual initiation in these joiner cases (and not consider target side vs. initiating side to be an initiation), you would generate the data with this specification and then either 1) drop the “joiner_t” dyads, or 2) recode “initiate” to be “0” when “joiner_t” = 1.

Coding Joiners As Initiators, Dropping Joiners

As you would expect, setting specifications as immediately above except for dropping joiners would drop the 6 directed dyads where one of the “joiner” conditions is true.

- treat joiners and originators as initiators;
- keep target vs. initiator dyads if no new MID;
- drop all joiner dyads.

cocode1	cocode2	year	ongoing	initiate	hostlev1	hostlev2	midnum	joiner_i	joiner_t
300	640	1816	0	0	0	0	0	0	0
640	300	1816	0	0	0	0	0	0	0
640	365	1816	0	0	5	5	1	0	0
200	365	1816	0	0	0	0	0	0	0
365	640	1816	0	1	5	5	1	0	0
365	200	1816	0	0	0	0	0	0	0

Dispute Dyad Output

Basic output

Dispute-Dyad output functions somewhat differently with regards to these specifications. The dispute-dyad output ALWAYS includes ONLY cases of directed dispute initiation. For that reason, there is not an option to modify including or not including target vs. initiator dyads. Target vs. initiator directed dyads never constitute a directed initiation, and so are never included in the output. So if A initiates vs. B, and B does not initiate another MID vs. A, the B vs. A dyad is not included in the output. Note, though, that if A initiates vs. B and B also initiates vs. A, both A vs. B and B vs. A ARE included in output.

So from the above sample data, if we set

- treat only “true” initiators (originators) as initiators;
- drop all joiner dyads.

we obtain one case in the output:

cocode1	cocode2	year	ongoing	initiate	hostlev1	hostlev2	midnum	joiner_i	joiner_t
365	640	1816	0	1	5	5	1	0	0

Include Joiners

But if we change this to include joiners,

- treat only “true” initiators (originators) as initiators;
- include all joiner dyads.

we STILL obtain only one case in the output, because the output only includes dispute initiations, and as specified joiner dyads are not initiations:

cocode1	cocode2	year	ongoing	initiate	hostlev1	hostlev2	midnum	joiner_i	joiner_t
365	640	1816	0	1	5	5	1	0	0

Treat Joiners as Initiators

Only if we set the specification to treat joiners as initiators would those other dyads be included. Note that all of these are “joiner_i” dyads, because the directed dyad-dispute output considers only the initiator side vs. target side is considered an initiation.

cocode1	cocode2	year	ongoing	initiate	hostlev1	hostlev2	midnum	joiner_i	joiner_t
200	300	1816	0	1	5	3	1	1	0
200	640	1816	0	1	5	5	1	1	0
365	300	1816	0	1	5	3	1	1	0
365	640	1816	0	1	5	5	1	0	0

Program Files

EUGene uses a variety of files for input, logging, and saving results.

Log File

Given appropriate input files and a correct installation, *EUGene* should operate without errors. However, few complicated programs are ever distributed bug-free, and problems with input files and data could lead to errors. *EUGene* generates a log file containing error messages generated by most internal errors. If an error should occur and you cannot figure out why, this log file may be useful to either the user or to *EUGene*'s programmers in figuring out what happened. This file is normally called "error.log" and is saved in the main *EUGene* program directory. A complete path to the location of the error log may be specified in the initialization file "*EUGene.ini*".

Input and Configuration Files

Raw input data and configuration information are included in a series of data files that must be available to *EUGene* upon initialization. It is important for the user not to attempt to alter these files, or to be very careful when doing so! Alterations to the configuration file may lead to *EUGene* being unable to find necessary input files, either initially upon startup or later as procedures are executed. Any alterations performed on the data files may render them unreadable unless the user is careful to maintain spacing, commas, and tabs as in the original files. In addition, changes to data files may render calculations performed by *EUGene* inaccurate and unreliable unless carefully documented and reported, and unless all recalculations are done in the necessary order.

In particular, when you change data in these files, you must be certain to consider the implications of this for subsequent data files. If you change state membership dates, you should recompute capabilities data and tau data so that it is correctly recorded as missing (or not) in those data files. Similarly, if you change capabilities data, you should (theoretically) recompute expected utility according to *The War Trap*, then risk scores, then expected utility according to *War and Reason* methods. The configuration file is saved in the main Windows directory of the PC when *EUGene* is installed. Raw input files are in the "INPUTDAT" subdirectory under the main directory where *EUGene* was installed.

Configuration Information -- file "eugene.ini": This file is read from the main directory where *EUGene* is installed, typically "C:\EUGene". The file contains file names and directory paths to other necessary input files and certain other information necessary for program initialization. Lines beginning with ";" are treated as comments (i.e. ignored). Each line consists of a key name plus a file name, path, or numeric value enclosed in quotation marks. Recognized key values and their definitions are as follows.

error_file: full path to the file where *EUGene* will record any error messages.
user_data_subdirectory : subdirectory name for user added data sets.
raw_cow_capabilities_data : relative path to raw version of COW capabilities data.
modified_cow_capabilities_data : relative path to file containing any modified capabilities data.
distance_data : relative path to data on capital city latitude/longitude.
colonial_contiguity_data : relative path to file with both direct and colonial contiguity data.

As of 12/17/2003, this file was assembled from Diehl's direct contiguity data (COW

contiguity data v 3.0) with an internal update to 2002, and updated COW colonial contiguity data (v3.0).

nation_data : relative path to interstate membership data.

major_power_data : relative path to interstate membership data.

cow_alliance_data : relative path to "sequence number" version of COW alliance data. (no longer relevant with v3.0)

alliance_seq_data : relative path to file listing ccodes and sequence numbers. (no longer relevant with v3.0)

dyadic_alliance_data : relative path to dyadic alliance file listing.

WTR_risk_data : name of raw data file with risk scores computed by Bueno de Mesquita (1985).

MID_Case_Data : name of COW MID data with basic case (one record per MID) information, normally distributed by COW as "mida_210.txt".

MID_Actor_Data : name of COW MID data with basic country-dispute information (one record per country per dispute), normally distributed by COW as "midb_210.txt"

MID_Name_Data : name of COW MID data with names of MIDs when available, distributed by COW as "midc_210.txt"

MID_Name_Data : name of COW MID data with names of MIDs when available, distributed by COW as "midc_210.txt"

MID_Name_Data : name of COW MID data with names of MIDs when available, distributed by COW as "midc_210.txt"

MID_Participant_Incident_Data: name of COW MID data file with participant-incident level data.

MID_Data_format; marker indicating whether the data being read (and marked by the previous variables) is in v3.0 format (which includes participant-incident data) or v2.1 (which has only MIDA, B, and C files). Recognized values are "2.1" or "3.0".

MID_Participant_Incident_Data_First_Year: Year that the participant-incident data file begins to have information.

MID_Format_2_1_Data_Last_year: Last year of the input MID data sets that does NOT have participant incident data, that is, the last year of the data in the pre-2003 (v2.1) format.

Maoz_Dyadic_MID_Data : name of Maoz-generated dyadic MID update data

ICB_Dyadic_Crisis_Data: name of the ICB dyadic crisis data file.

COW_War_Data: name of the COW dyadic wars data file.

werner_peace_years_file_name: name of the file containing Suzanne Werner's peace years correction information.

Minimum_Distance_Data: name of data file containing minimum distance between states data.

polity3_data : name of Jagers and Gurr Polity III data file.

syscap_data : name of intermediate (binary) file with national capability index computed by *EUGene*.

tau_data : name of intermediate (binary) file with tau-b scores.

s_data : name of intermediate (binary) file with *S* scores (contains both weighted and unweighted *S*).

risk_Tau_data : name of intermediate (binary) file with risk scores computed by *EUGene*, with calculations based on tau-b scores.

risk_S_unweighted_data : name of intermediate (binary) file with risk scores computed by *EUGene*, with calculations based on unweighted *S* scores.

security_alliance_Tau_data : name of intermediate (binary) file with security alliance information generated by *EUGene*, with calculations based on tau-b scores.

security_alliance_S_unweighted_data : name of intermediate (binary) file with security alliance information generated by *EUGene*, with calculations based on unweighted *S* scores.

EU_War_Trap_Tau_data : name of intermediate (binary) file with expected utility scores following *The War Trap* methods (as modified and discussed above) computed by *EUGene*, with calculations based on tau-b scores.

EU_War_Trap_S_unweighted_data : name of intermediate (binary) file with expected utility scores following *The War Trap* methods (as modified and discussed above) computed by *EUGene*, with calculations based on unweighted S scores.

EU_War_and_Reason_Tau_data : name of intermediate (binary) file with expected utility scores following *War and Reason* methods computed by *EUGene*, with calculations based on tau-b scores.

EU_War_and_Reason_S_unweighted_data : name of intermediate (binary) file with expected utility scores following *War and Reason* methods computed by *EUGene*, with calculations based on S scores.

EU_War_and_Reason_S_weighted_data : name of intermediate (binary) file with expected utility scores following *War and Reason* methods computed by *EUGene*, with calculations based on capability-weighted S scores.

EUGene_ftp_site: internet address that *EUGene* uses to access/download additional user data sets.

Saved_Settings: file name where *EUGene* saves/loads settings when user selects “save settings” or “load settings” from the “File” menu.

bibliography: file name containing bibliography information that is put into the log/command file.

HelpFiles_FileMenu_name, HelpFiles_RecomputeMenu_name, etc: Help files read *EUGene* for on-line help.

first_nation_year : first year any state is a nation according to interstate members list to be read by *EUGene*.

last_nation_year : last year any state is a nation according to interstate members list to be read by *EUGene*.

first_capability_year : first year any state has capability data in the capability data set to be read by *EUGene*.

last_capability_year : last year any state has capability data in the capability data set to be read by *EUGene*.

first_alliance_year : first year any state has alliance data in the main dyadic alliance data sets to be read by *EUGene*.

last_alliance_year : last year any state has alliance data in the main dyadic alliance data sets to be read by *EUGene*.

first_alliance_seq_year: first year any state has alliance data in the sequence number version of the alliance data sets to be read by *EUGene* (no longer relevant).

last_alliance_seq_year: last year any state has alliance data in the sequence number version of the alliance data sets to be read by *EUGene* (no longer relevant).

first_polity3_year : first year any state has polity III data in the polity III data set to be read by *EUGene*.

last_polity3_year : last year any state has polity III data in the polity III data set to be read by *EUGene*.

first_wtr_risk_year : first year any state has data from the “War Trap Revisited” article data set.

last_wtr_risk_year : last year any state has data from the “War Trap Revisited” article data set.

first_MID_year : first year any state has MID data in the MID data set to be read by *EUGene*.

last_MID_year : last year any state has MID data in the MID data set to be read by *EUGene*.

first_ICB_year: first year any state has MID data in the dyadic ICB data set to be read by *EUGene*.

last_ICB_year: last year any state has MID data in the dyadic ICB data set to be read by *EUGene*.

first_contiguity_year: first year any state has data in the colonial/direct contiguity data set to be read by *EUGene*.

last_contiguity_year: last year any state has data in the colonial/direct contiguity data set to be read by *EUGene*.

first_mindist_year: first year any state has data in the minimum distance data file.

last_mindist_year: last year any state has data in the minimum distance data file.

Raw capability data -- file "NMC_3.0.csv": this file is v3.0 of the COW National Capabilities data file. The COW file and codebook are available from the Correlates of War 2 website (cow2.la.psu.edu). This dataset codes for each international system member (state), for each year between 1816 and 2001, demographic, military and industrial indicators, along with major power status. The raw data file here is in comma separated variable format, and has (in order) variables for country abbreviation, country code, year, iron and steel production, military expenditure, military personnel, energy consumption, total population, urban population, the aggregate CINC score, and the version number.

Modifications to Capabilities data -- file "cowmod10-2003.csv": Some users may wish to modify the capabilities data for their own runs. *EUGene* has the ability to do this. For computing the capabilities index, the user can choose to either 1) use only the officially released COW data on capabilities, as given by COW; 2) drop the "energy" category from the capability index (the energy category is perhaps the most volatile category and in an earlier version contained the most missing data of the 6 capability categories, especially up to about 1850); 3) include modifications to the COW data as provided by the user in the file "cowmod1-2003.csv". If the "include modified data" option is specified, then the input file "cowmod10-2003.csv" must be present in the "INPUTDAT" subdirectory. The capabilities data that are in this file will replace (in memory) the original COW capabilities for the country-years included. This file has the same format and variables as the "NMC_3.0.csv" file. However, it is currently blank given the release of the newest COW capabilities data (v3.0).

EUGene was previously distributed with a modified cowmod9-1999.csv file, which has changes to values for energy consumption for the UK between 1816 and 1853. The original COW data file has a few values for energy for UK between 1816 and 1854, but many are missing. This is a problem because the UK has high values for energy for the years where data is available. There is a large discrepancy from year to year when energy data changes from being available to unavailable. For instance, if the COW data is left as is, the UK's national capability index varies from .34 in one year to .23 in the next (UK has 90% of the total energy in the system when energy data is reported). Since we know Great Britain was still using energy from year to year, the modified capability data file has energy data interpolated between 1816 and 1854 in those years where the COW data does not have it. The years and values can be found in Appendix B.

Raw Alliance data -- Beginning with v3.0, *EUGene* is distributed with COW version 3.0 alliance data files (only). The format of the file is required to be that of the multiple-alliance per dyad year file posted on the COW 2 website, namely with the variables country code 1, country code 2, year, alliance number, alliance type, and a version number. Alliance type ranges from 1 to 3 and designates the most committed alliance (lowest alliance value where defense pact=1, neutrality pact=2, entente=3) that states 1 and 2 were involved in during a given year. Alliance number is the identification number given in the latest COW listing for the alliance leading to that type. It is acceptable for there to be multiple alliances listed for a pair of states in any given year in this data set. *EUGene* will automatically select the most-committed level of alliance present in each dyad-year for computations and output. (Note that because COW has employed the same rule as *EUGene* to create single record per alliance year data, namely taking the highest level agreement, *EUGene* could alternately employ the single record data set.)

Contiguity data -- file "Unified Contiguity 2002.csv": this file was constructed from the 2002 versions of the COW direct and colonial contiguity data sets. This is a combined file that contains both direct and colonial contiguity relationships. Contiguity is classified into five separate types, following the COW specifications for the types of contiguity: 1) land contiguity; 2) contiguous for up to 12 miles of water; 3) contiguous for 13-24 miles of water; 4) contiguous for 25-150 miles of water; 5) contiguous for 151-400 miles of water. The organization of the file is described in the documentation to the COW contiguity data file, but is as follows. The file has 7 variables: 1) state A ccode, 2) state A's possession ccode, if any; 3) type of contiguity; 4) state B ccode; 5) state B's possession ccode, if any; 6) begin year of contiguity relationship; 7) end year of contiguity

relationship. If the record is a direct contiguity relationship, then the possession ccodes will be blank. Format of file follows the 1993 data file form as described in "Contiguity 1993.doc".

City Location data – file "latlong2002.raw": the "latlong" file was constructed by the authors of *EUGENE*. The file codes the latitude and longitude coordinates of state capitals. *EUGENE* then calculates distance. When a state is large, for example, the U.S., more than one city is included. When the capital city changes, multiple entries are contained in the file along with the appropriate time period that they should be used. Variables in the file are in order country code, city name, latitude hours, latitude minutes, latitude seconds, "n" or "s" for north/south latitude, longitude hours, longitude minutes, longitude seconds, "e" or "w" for east/west longitude, first year this city should be used, last year this city should be used, full name of country. **This file must be updated if dates in nation membership file are updated or modified.** If proper distances for a country in a year are to be output, then each country must have an entry in this file that contains that year. Also, if multiple entries over time appear for a state (for instance, the capital of Italy shifts from Turin to Rome in 1871), they must appear in time order. The file "latlong2002.raw" replaced the prior version of this file for version 3.0.

Nation Membership data -- files "states2002.csv" and "majors2002.csv": List of nations and years of membership in the international system, and list of major power dates. As of v1.13 of *EUGene*, we use the standard format of these two files from the COW system membership data, as downloaded from the Peace Science Society web site (pss.la.psu.edu) in August, 1998. Note 1: if this file is updated, the file containing latitude/longitude information must be updated as well. Note 2: The 2002 version of the state membership data set reordered the first two variables of the data set, and so the code within *EUGene* was changed. This means it will be hard to go back to the 1997 state data. So, I have reordered the variables within the 1997 version of the state data set, so that it may again be used. [Versions of *EUGene* through v1.12 used the 1994 version of the COW system membership data. There are a few small differences between the two versions, which affect perhaps 25 country-years (out of over 11,000 country years in the international system). While we do not necessarily recommend re-running analysis that used data sets created with v1.12, we do recommend using the new data when possible.]

War Trap Revisited Risk data -- file "WTRRisk.txt". Risk data from "War Trap Revisited" (Bueno de Mesquita 1985) calculations, as obtained from input files to Bueno de Mesquita's TOLSTOY program. This file contains risk attitude data for states in their home region and sometimes in other regions, from 1816 to 1970.

Polity III data – file "polity3.dat". Jagers and Gurr polity data. This file should remain in the flat text format in which Jagers and Gurr distributed in the May 1996 version. The order of the variables in that file (with one country-year per line) is as follows:

```
CaseNum ccode country_abbreviation year autoc democ xrreg xrcomp xropen mono
xconst parreg parcomp cent
```

User Dyad List – file "userdyad.txt" (or other name of user's choosing). This file contains a list of dyads that will be read in and used by *EUGene* to set up the population of dyad-years for output, if the user specifies the cases for output as "Dyads Read from User File." The format of this file is as follows: each line must contain exactly 4 integers, corresponding to 1) the ccode of the first country in the dyad, 2) the ccode of the second country in the dyad, 3) the first year for which you want this dyad to be output, and 4) the last year for which you want this dyad to be output. So sample lines from this file might be as follows:

```
2 365 1946 1980
365 2 1946 1980
200 220 1816 1918
```

These lines would designate that the user wants in the output file the US vs. USSR directed dyad from 1946 to 1980, the USSR vs. US directed dyad from 1946 to 1980, and the UK vs. France directed dyad from 1816 to 1918. The ccode and year values must be separated by spaces, tabs, or commas. Note that dyads entered in this file are treated as directed, so that if you enter ccode 2

vs. ccode 365, you will get only the directed dyad of 2 vs. 365. If you want the reverse direction as well, enter 365 vs. 2 also.

Online Help files -- files "filemenu.hlp", "recomenu.hlp", etc. Contain text of various help screens accessed within *EUGene*. These files are in rich text format.

MID data files. To create dyadic MID data, *EUGene* uses the following files, with the following specifications:

1. MIDA (currently MIDA_3.02.csv)
 - Variables, in order:
DispNum, StDay, StMon, StYear, EndDay, EndMon, EndYear, Outcome, Settle, Fatality, FatalPre, MaxDur, MinDur, HiAct, HostLev, Recip, NumA, NumB, Link1, Link2, Link3, Ongo2001, Version
 - File must have a header line with variable names.
 - File must be saved as comma-delimited.
2. MIDB (currently MIDB_3.02.csv)
 - Variables, in order:
DispNum, StAbb, ccode, StDay, StMon, StYear, EndDay, EndMon, EndYear, SideA, RevState, RevType1, RevType2, Fatality, FatalPre, HiAct, HostLev, Orig, Version
 - File must have a header line with variable names.
 - File must be saved as comma-delimited.
3. MIDIP (currently MIDIP_3.02.csv)
 - Variables, in order:
DispNum, IncidNum, StAbb, ccode, StDay, StMon, StYear, EndDay, EndMon, EndYear, InSide A, SideA, Fatality, FatalPre, Action, HostLev, RevType1, RevType2, Version
 - File must have a header line with variable names.
 - File must be saved as comma-delimited.
 - Data should be sorted by MID#, and then participant-mid-id #. (e.g.3551001)
Incident records need to be in chronological order; if a date is missing (-9) then *EUGene* takes the first incident in the list as first, so care must be taken in sorting to ensure that incidents with a -9 date do not automatically come first.

EUGene does not employ the MIDI file.

EUGene does, however, read dispute names from the MIDC_210.txt file (to be replaced by v3.0 names when available).

Intermediate Files

EUGene calculates several intermediate data files that are saved to disk. These files are distributed with the *EUGENE* program to avoid the user having to run each procedure to obtain output, and contain values pre-calculated using the default options described in this document. These files are binary, and cannot be edited or directly viewed by the user.

Percent System Capabilities file: This file contains the percentage of system capabilities, COW national capabilities index, for all states and years. The file is saved as "syscapv2.dat" in *EUGene*'s "Intermediate Data Files" subdirectory.

Tau-b scores file: This file contains regional and global tau-bs for all dyad years. This file is saved as "tau.dat" in *EUGene*'s "Intermediate Data Files" subdirectory.

S scores file: This file contains regional and global S scores for all dyad years. This file is saved as "s.dat" in *EUGene*'s "Intermediate Data Files" subdirectory.

Expected Utility data file based on The War Trap: This file includes the expected utility scores as calculated following *The War Trap* methods. There are 2 files with this data, “EuwartrpTau.dat” and “EuwartrpS.dat” in *EUGene*’s “Intermediate Data Files” subdirectory. These data contain expected utility using Tau and S scores respectively.

Risk Attitude Scores file: This file includes risk attitude data for all states for each region and year, as calculated following equations in “The War Trap Revisited.” This file is saved as “riskv2Tau.dat” in *EUGene*’s “Intermediate Data Files” subdirectory when calculated using Tau. An S-based version, based on unweighted S, is named “riskv2S18161992.dat”.

Expected Utility data file based on War and Reason: This file includes expected utility scores as calculated by the equations in *War and Reason*. This file is saved as “EUWarReasonTau.dat” in *EUGene*’s “Intermediate Data Files” subdirectory. S-based versions of this data are named “EUWarReasonSUnweighted.dat” and “EUWarReasonSUnweighted.dat”.

Security Alliances file: This file includes all data relating to security alliance structures for all dyad years. Because it is inaccessible except through internal procedures, we are not distributing this file, but it is available if desired. This file is saved as “security_alliancesTau.dat”. An S version of this data to be named “security_alliancesS.dat” will be available in a future update.

Known Bugs and Problems

1. *Estimated time to completion not always accurate.* The estimated time to completion is only a rough estimate, and may change rapidly at times. As the procedure being estimated approaches completion, the estimate will normally converge to 0, but at times there will be discrepancies. This is due to the fact that countries and dyads are distributed unevenly across time and space, with the number of dyads increasing dramatically after 1945. *EUGene* attempts to estimate the remaining number of dyads, and looks at completion time, dynamically rather than pre-calculating the number of cases that must be processed. This results in the rapid changes, and in some cases the inability to estimate, the time remaining.
2. *Memory violation/access error after user interrupt.* When the user interrupts *EUGene* by using the “Stop” button, the program sometimes fails to clear all of memory, resulting in a Windows memory access violation. If this occurs, the program will normally terminate and leave the user in Windows 95/NT. However, if a Windows error message appears and the program keeps running, the best solution is simply to exit and restart *EUGene* in order to ensure that none of the internal data structures have been corrupted. You may be able to keep running successfully, but we recommend exiting and restarting to avoid any possible complications. If possible, write down any details of what you were doing when you interrupted the program and notify the programmers.
3. *Trace window remains onscreen.* Normally when *EUGene* finishes executing, it will clear away all of the onscreen windows that it created while performing processing. On some occasions, especially if the program run is interrupted, the last of those trace windows is not removed. If a trace window remains onscreen, it will not interfere with *EUGene*’s continuing functioning; as soon as a new procedure starts, windows will return to normal.

If you believe you have found a bug, please contact the authors (see “Contacting the Authors” on page 1).

Internal Details for Programmers

The following contains information about how *EUGene* functions internally for programmers looking to understand and/or modify the code. If code is found that is in error, please notify the primary program author, D. Scott Bennett. WHEN POSSIBLE, he will try to answer questions about how the program was developed and on internal details of the program that will allow extensions and program verification.

In general, files containing the program code for procedures and windows are in various .pas and .dfm files in the "source" directory. These files are brought together through the eugene32.dpr Delphi project file. All of these files are text files, but within Delphi will be properly formatted, and associated in the proper order if the project file is opened. For programmers looking at the program, the "mdiframe" unit contains the main menus for the program. From there, code and actions can be traced through the program. The unit "eutypes1.pas" contains the most fundamental data structures referred to throughout the rest of the code. The "PagedOutput" unit contains the main selection windows. The "euinoutd.pas" unit contains most of the code that actually outputs data. The other units contain types, procedures, and windows for various particular data structures and I/O windows.

Details on some program structures and methods:

Debug_level: When running, *EUGene* checks a setting contained in the debug_level array to know what sort of trace messages (how detailed) to put on screen. When the debug_level is set high (this is a constant, and so is set at compile time), the program will print out many details as it iterates. At a low level, trace output will contain very little information. Internally the debug_level is set up as an array[1..10] of boolean, so when debug[10]=true, debugging is set to level 10, when debug[1]=true, debugging is set to level 1. Some levels and their output are:

Level 2 or greater: trace output reports how many partitions are necessary for the run.

Level 3+: Trace output reports details of partition division.

Level 4+: Trace output reports Memory calculations, e.g. how much heap was needed for particular structures.

Run Partitioning: The data files contain information on about 200 countries for the years between 1816 and 1992, although not all data are available for all years. A data matrix of all countries in all years is quite large for some variables, and few computers would be capable of holding all alliance, capability, statehood, risk, and expected utility data for all country-years in memory simultaneously. As a result, the program parcels out the data to be processed in smaller amounts by only processing a small number of years at one time. Creating arrays and processing the data in parts is essential because some computers may not have enough memory to hold all of the needed information without crashing. Therefore, the *EUGene* calculates how much data can be fit into the amount of available memory on the PC and then processes data in chunks until all years have been analyzed.

Nation_Array: There are 200 countries, but country codes run from about 1 to 1000. While it would be nice to index all of the internal arrays (of capabilities, risk, etc.) by ccode, this multiplies by 4 to 5 times the amount of memory that would be required for static arrays. As a result, there is a "nation_array" structure that acts as intermediary between country codes and a shorter index used to create *EUGene's* static array structures. *EUGene* first reads the nation list from input data, and indexes all countries into an array by county code number. The program then creates a unique index for all countries, ranging from 1 to about 200 (this is created dynamically, and additional countries could be added to the input nation list up to a hard-coded limit of 250 states without a problem). This gives each country a unique mapping or index number that is then used throughout the rest of the program.

Objects: *EUGene* uses a lot of object oriented programming, for most but not quite all of its internal structures.

Dispute Data: *EUGene* takes the country-dispute records from the input MID data set and converts them to directed dispute records. Note that these records are created as all records of all initiator-side states vs. all target-side states (or all revisionist side vs. SQ states, if the user specifies revisionists). What this means for programming purposes is that you must think carefully about how to access that array using calls to procedures like "is_target" because the desired information is about the second member of an ordered dyad, but to get it you may need to look at dyads using information about the first member (yeah, right!). For instance, if you want to know in dyad 2 vs. 731 in 1950 that state 2 joined the target side, you actually need to be certain

that the programming will use 731 vs. 2, because any target status of 2 will only be recorded in the 731 vs. 2 dyad. Since 2 vs. 731 was not an initiator vs. target dyad, it isn't coded. This is an issue just for internal programming of additional MID-based procedures.

Inheritance in Dispute Data: For v3.0 the structures for the dispute data were significantly reworked, with several previously standalone objects being integrated in the declarations through inheritance. This was done to facilitate the creation of the conflict_exclusion set, which must fit the dispute, war, and icb data sets. More careful attention now has to be paid, however, to what sections of code some MID procedures are actually accessing, since base declarations are overridden in subsequent structures.

Meshing dispute data structures: As noted previously, COW computed or Maoz dyadic dispute data is integrated with the computed dyadic disputes from MID 3.0. The way this works procedurally is that the dispute 3.0 data is read, and dyadic disputes computed. Then the pre-1993 is read and integrated, by reading the data into a separate standalone structure, then copying dyadic disputes (from Maoz or COW computation) into the new structure.

Constants: At the top of the file EUTypes1.pas, there are a number of predefined constants giving parameters for input values and so on. If significant modification of *EUGene* is to be made, or if the program is to be used after 2025 or on data before 1800, many of these constants will have to be changed.

Changes in v3.0: in v3.0 a number of procedures were reworked, particularly in the output unit. Where there used to be 3 separate routines for outputting the three possible units of analysis, the procedures are now integrated, with only the final loop to go through the cases differencing.

Additional risk procedures only available to programmers. Programmers looking at the MDIFrame main run screen will see that it has a menu titled **Risk/Security Conversion (Programmers)** and another titled **Custom Procedures**. These menus are set at run time to be hidden and invisible to users, but programmers can access by changing the "Frame.RiskSecurityProcs1.Visible := false;" and "Frame.CustomProceduresMenu.Visible := false" lines in the *EUGene32.dpr* main project file. The Risk/security conversion menu has 4 options; the custom procedures menu has several more. These will never need to be accessed by users, and the procedures that actually generate data under these submenu do not need to be run again. But, the procedures under risk/security menu are documented here.

Create New Format Risk Data. This procedure was used to convert risk data files from v1 risk format (in one of the earliest *EUGene* versions) to v2. Since this has been done once, it does not have to be done again.

Show Risk Data. This procedure just shows a couple of years of risk data on screen, just listing ccode-year-risk score for years that are hardcoded in the program.

Create Baseline Security/Alliance Data. This procedure was used to create an initial security file where best and worst alliances were set to the actual alliance pattern. Since the best/worst alliance file now exists, and will be overwritten with the program's finding of what the best/worst alliances are when risk data is generated, this does not need to be run ever again.

Show Current Security/Alliance Data. This procedure just shows a couple of years of best security alliance data onscreen. [Note: Programmers can draw on this to convert for eventual user output of security alliance patterns.]

Legal Notice

Copyright

EUGene Copyright 1997-2007 D. Scott Bennett Jr. and Allan C. Stam, III
All Rights Reserved

Conditions of Use

EUGene, a software program to generate expected utility data, is distributed as freeware by the authors. YOU MAY NOT FURTHER REDISTRIBUTE ANY PART OF THIS PROGRAM, INCLUDING DATA FILES, SOURCE CODE, HELP AND DOCUMENTATION FILES, AND THE EXECUTABLE PROGRAM. Delphi source code for *EUGene* is included as part of distribution, and may be modified and recompiled in the course of conducting research on the expected utility theory of war. However, NEITHER THE MODIFIED CODE NOR THE RESULTING COMPILED PROGRAMS MAY BE REDISTRIBUTED EXCEPT WITH PERMISSION OF *EUGENE'S* AUTHORS.

Program Extensions and Modifications

If you wish to make modifications or extensions that lead to you changing the program code or using portions of *EUGene's* code in other programs, you are permitted to do so. However, if you modify the program, use segments of the code in other programs, generate new results, and report them, any modifications you make must be precisely documented, and *EUGene* must be cited as the initial source of your extensions. In addition, modified code cannot be distributed without the express permission of Scott Bennett and Allan Stam. If you run *EUGene* in its unmodified form and change program options to generate new data and output, express permission to publish those results is not needed, but the settings that you make should be precisely documented (to allow replication), and *EUGene* should be cited.

Disclaimer of Warranty

This software and manual are distributed "as is" and without warranties as to performance of merchantability or any other warranties whether expressed or implied. Because of the various hardware and software environments into which this program may be put, no warranty of fitness for a particular purpose is offered. The user must assume the entire risk of using the program. Any liability of the author will be limited exclusively to product replacement.

EUGene was developed using Delphi, versions 1 through 7, 2005, and 2007 by Borland International, Inc., Inprise, and CodeGear.

APPENDIX A

Modified Values in Modified Capabilities Data File

These commands show the values for energy that are changed if the user selects the "Modified Capability" calculation under the "calculate system capabilities" recalculate option and uses the default *EUGene* modified capability data file. There are 1) changes for energy values for the United Kingdom, and 2) changes for 1991-1993 for energy and urban population.

In the 1991-1993 capabilities data, there are significant missing data on urban population, energy and (in 1993) Iron/Steel production and total population. There are always missing data on these variables, but in these years the data become missing to an extent that it distorts any balance between countries. The problem with the pattern of missing data becomes apparent when the data are sorted by time rather than by country. The difficulty in 1991-1993 is that when there are only valid data on a small number of countries, they gain a very large share of the % of system capabilities in that area, and if this is high enough, it will cause them to show up as quite large overall, because the one component pulls up the average across the 6 components dramatically. For example, in 1993 only 1 country (Macedonia) has any non-zero, non-missing data on urban population. As a result, it has 100% of the system urban population, and this 100% gets averaged with tiny percentages in other area. The resulting inflation of the countries' capabilities is unwarranted. In 1992, 6 countries have urban population data. In 1991, 34 countries have this data (a few others have correct values of 0). But when some countries have correct values of 0 while all or most others are missing, this poses a different problem for the CINC scores, namely that states with 0 values receive a 0 as their % of system total, which deflates their CINC scores. As long as other states are receiving some positive amount, this is correct. But if no (or few) states have positive capabilities in these areas, then the states with 0 are being pulled down unnecessarily.

These problems are especially noticeable for 1991-1993. Our solution in the problematic cases is to recode ALL countries as having missing data. The component of the CINC score then drops out of estimation. We recognize that this introduces some over time distortion since scores will be computed across 4 or 5 components rather than 6, but we believe this solution is superior to the even more dramatic distortions that result from leaving all components in. As a result, for 1991-1993, we code all urban population figures as missing values in our modified capability data file. For 1992-1993, we code all energy values as missing. In 1993, Iron/Steel production is recoded as missing (here, all data is either missing or 0; we recode it all to missing). In 1993, we also recode total population to missing (about 2/3 of the data is missing on total population in 1993; the sum of world population in 1992 is 5.262 bn., while in 1993 it is 874 mn.). This leaves in our modified data only the 2 military components of the CINC scores intact. For this reason, WE RECOMMEND NOT CONDUCTING ANALYSIS FOR 1993 for any variables that require accurate CINC scores.

Changes for 1816-1853 UK are as follows. Entries with a "*" are the values that are actually in the original COW data file:

```
* if (ccode=200) and (year=1816) energy=15958.
if (ccode=200) and (year=1817) energy=16393.5.
if (ccode=200) and (year=1818) energy=16829.
if (ccode=200) and (year=1819) energy=17264.5.
* if (ccode=200) and (year=1820) energy=17700.
if (ccode=200) and (year=1821) energy=18114.
if (ccode=200) and (year=1822) energy=18528.
* if (ccode=200) and (year=1823) energy=18942.
if (ccode=200) and (year=1824) energy=20462.
* if (ccode=200) and (year=1825) energy=21982.
if (ccode=200) and (year=1826) energy=22043.2.
if (ccode=200) and (year=1827) energy=22104.4.
if (ccode=200) and (year=1828) energy=22165.6.
if (ccode=200) and (year=1829) energy=22226.8.
* if (ccode=200) and (year=1830) energy=22288.
if (ccode=200) and (year=1831) energy=23302.2.
```

if (ccode=200) and (year=1832) energy=24316.4.
if (ccode=200) and (year=1833) energy=25330.6.
if (ccode=200) and (year=1834) energy=26344.8.
* if (ccode=200) and (year=1835) energy=27359.
if (ccode=200) and (year=1836) energy=28403.6.
if (ccode=200) and (year=1837) energy=29448.2.
if (ccode=200) and (year=1838) energy=30492.8.
if (ccode=200) and (year=1839) energy=31537.4.
* if (ccode=200) and (year=1840) energy=32582.
if (ccode=200) and (year=1841) energy=34889.2.
if (ccode=200) and (year=1842) energy=37196.4.
if (ccode=200) and (year=1843) energy=39503.6.
if (ccode=200) and (year=1844) energy=41810.8.
* if (ccode=200) and (year=1845) energy=44118.
if (ccode=200) and (year=1846) energy=45041.6.
if (ccode=200) and (year=1847) energy=45965.2.
if (ccode=200) and (year=1848) energy=46888.8.
if (ccode=200) and (year=1849) energy=47812.4.
* if (ccode=200) and (year=1850) energy=48736.
if (ccode=200) and (year=1851) energy=51930.25.
if (ccode=200) and (year=1852) energy=55124.5.
if (ccode=200) and (year=1853) energy=58318.75.

APPENDIX B

Default Specifications for Pre-Calculated Data

EUGene is distributed with data for capabilities, tau and S scores, risk, and expected utility pre-calculated. Unless the user wants to modify the raw data or make different assumptions than those that went into those calculations, these data can simply be output using the "Output" menu. The default data and options used to calculate the intermediate files distributed with *EUGene* are as follows.

National Capabilities Index: Created from COW data, plus interpolated values for Energy in Britain, 1816-1854.

Tau-b and S Scores: Created using COW version 3.0 alliance data.

Expected Utility (War Trap version): Distance discounting set to capitals plus contiguity plus multiple locations. Alliance data used is from above.

Risk Scores: Distance discounting set to capitals plus contiguity plus multiple locations. Alliance data used is "dyadic alliance data" version. Algorithm used is genetic algorithm, 20 member population, mutation probability 0.05, 8 generation stability criteria, and 2 "cloned" patterns. Capability data is from above. Note that as of v2.40 (January 2002), these scores have not been updated using the new COW v3.0 alliance data.

Expected Utility (War and Reason version): Distance discounting set to capitals plus contiguity plus multiple locations. Risk scores used were those generated by *EUGene* as above. Capability data is as above. Note that for v3.0 (September 2003), these values were updated using the new COW v3.0 alliance data, but not new risk scores (which have not been recomputed).

APPENDIX C
Data Sources

The development and functioning of *EUGene* would not be possible without the data accumulated through the hard work and dedication of many researchers. The following table lists of original data sources *EUGene* uses and presents. In some cases, the data used is directly from an external source, while other data is *EUGene* output based on previously defined computations. If you use variables from *EUGene* that derive from these external sources, we ask that you cite the original sources as well as *EUGene*. Regardless of the source, *EUGene* users should consult the original citation for detailed information about the interpretation and limitations of particular output results.

Variable	Unit of Analysis	Source	Original Citation
Alliance	dyad-year	COW Alliance data set	Gibler and Sarkees (2002)
Capabilities	country-year	COW national capabilities index (CINC)	Singer, Bremer, and Stuckey (1972)
Contiguity/Distance	dyad-year	Contiguity from COW contiguity and colonial/dependency data sets; Distance computed in <i>EUGene</i>	Small and Singer (1982)
Dyad Duration	dyad-year	Calculated in <i>EUGene</i>	
Expected Utility (War Trap)	dyad-year	Calculated in <i>EUGene</i>	Bueno de Mesquita (1981)
Expected Utility (War and Reason)	directed dyad-year	Calculated in <i>EUGene</i>	Bueno de Mesquita (1985), Bueno de Mesquita and Lalman (1992)
International Interaction Game Equilibria	directed dyad-year	Calculated in <i>EUGene</i>	Bueno de Mesquita and Lalman (1992)
Home Region	country-year	COW system membership and region list	Small and Singer (1982)
Relevant Region of Conflict	directed dyad-year	Calculated in <i>EUGene</i>	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 <i>EUGene</i> ; Values pre-1816 from Werner (2000)	Beck, Katz, and Tucker (1997); Werner (2000)
Political Relevance	dyad-year	Calculated in <i>EUGene</i>	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 <i>EUGene</i>	Bueno de Mesquita (1975); Bueno de Mesquita and Lalman (1992)
Risk Attitude (<i>EUGene</i>)	country-year	Calculated in <i>EUGene</i>	Bueno de Mesquita (1985)
Risk Attitude (<i>War Trap Revisited Data</i>)		Bueno de Mesquita (through <i>Tolstoy</i> software)	Bueno de Mesquita (1985); Horn (1990)
System	country-year	Calculated in <i>EUGene</i>	Singer, Bremer, and Stuckey (1972)

Characteristics			
Tau-B	dyad-year	Calculated in <i>EUGene</i>	Bueno de Mesquita (1975); Bueno de Mesquita. (1981)
Tau-B with System Leader	country-year	Calculated in <i>EUGene</i>	Lemke and Reed (1998).

Bibliography

- 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* 42 (October); 1260-1288.
- Bennett, D. Scott. 1997. "Testing Alternative Models of Alliance Duration, 1816-1984." *American Journal of Political Science* 41:846-878.
- Bennett, D. Scott, and Allan Stam. 1998a. "Is Instrumental Rationality a Universal Phenomenon?" Paper presented at the 56th Annual Meeting of the Midwest Political Science Association, April, Chicago, IL.
- Bennett, D. Scott, and Allan Stam. 1998b. "Theories of Conflict Initiation and Escalation: Comparative Testing, 1816-1980." Presented at the Annual Meeting of the International Studies Association, March 17-21, Minneapolis, MN.
- Bennett, D. Scott, and Allan Stam. 1997a. "Theories of Conflict Initiation and Escalation: Comparative Testing, 1816-1984." Presented at the Annual Meeting of the American Political Science Association, August 28-31, Washington, DC.
- Bennett, D. Scott, and Allan Stam. 1997b. "Is Instrumental Rationality a Universal Phenomenon?" Paper presented at the 55th Annual Meeting of the Midwest Political Science Association, April 11, Chicago, IL.
- Bennett, D. Scott, and Allan Stam. 1995. "Comparative Testing and Theories of Dispute Escalation, 1816-1970." Presented at the Peace Science Society Annual Meeting, October, Columbus, OH.
- Bueno de Mesquita, Bruce. 1975. "Measuring Systemic Polarity." *Journal of Conflict Resolution* 19:187-216.
- Bueno de Mesquita, Bruce. 1978. "Systemic Polarization and the Occurrence and Duration of War." *Journal of Conflict Resolution* 22:241-267.
- Bueno de Mesquita, Bruce. 1981. *The War Trap*. New Haven: Yale University Press.
- Bueno de Mesquita, Bruce. 1985. "The War Trap Revisited." *American Political Science Review* 79: 156-177.
- Bueno de Mesquita, Bruce, and David Lalman. 1988. "Empirical Support for Systemic and Dyadic Explanations of International Conflict." *World Politics* 41:1-20.
- Bueno de Mesquita, Bruce, and David Lalman. 1992. *War and Reason*. New Haven: Yale.
- Bueno de Mesquita, Bruce, David Newman, and Alvin Rabushka. 1985. *Forecasting Political Events*. New Haven, CT: Yale.
- Fitzpatrick, Gary L., and Modlin, Marilyn J. 1986. *Direct-Line Distances (United States Edition)*. Metuchen, NJ: the Scarecrow Press, Inc.
- Gibler, Douglas M., and Meredith Sarkees. 2002. "Coding Manual for v3.0 of the Correlates of War Formal Interstate Alliance Data set, 1816-2000." Typescript.
- Goldberg, David E. 1989. *Genetic Algorithms in Search, Optimization, & Machine Learning*. New York: Addison-Wesley Publishing Company, Inc.
- Hays, William L. 1981. *Statistics* 3rd ed. New York: Holt Rinehart and Winston.
- Holland, John H. 1975. *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.
- Horn, Michael. 1990. *Tolstoy*. Computer program.
- Jagers, Keith, and Ted Robert Gurr. 1995. "Tracking Democracy's Third Wave with the Polity III Data." *Journal of Peace Research* 32:469-482.
- 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* 15(2); 163:213.

- Lemke, Douglas, and William Reed. 1998. "Power is not Satisfaction." *Journal of Conflict Resolution*. 42; 511-516.
- Maoz, Zeev, and Bruce Russett. 1993. "Normative and Structural Causes of the Democratic Peace." *American Political Science Review* 87:624-638.
- Signorino, Curtis S., and Jeffrey M. Ritter. 1999. "Tau-b or Not Tau-b: Measuring the Similarity of Foreign Policy Positions." *International Studies Quarterly* 43 (March): 115-144.
- Singer, J. David, and Melvin Small. 1966. "Formal Alliances, 1815-1939." *Journal of Peace Research* 3;1-31.
- Singer, J. David, Stuart Bremer, and John Stuckey. 1972. "Capability Distribution, Uncertainty, and Major Power War, 1820-1965." in Bruce Russett, ed. *Peace, War and Numbers*. Beverly Hills: Sage.
- Small, Melvin, and J. David Singer. 1969. "Formal Alliances, 1815-1965: An Extension of the Basic Data." *Journal of Peace Research* 6; 257-282.
- Small, Melvin, and J. David Singer. 1982. *Resort to Arms: International and Civil Wars, 1816-1980*. Beverly Hills: Sage Publications.
- Werner, Suzanne. 2000. "The Effects of Political Similarity on the Onset of Militarized Disputes, 1816-1985." *Political Science Quarterly* 53(2): 343-374.

Index

<p>#</p> <p>φ 20</p> <p>“</p> <p>“Alliance” Variable Tab 37</p> <p>“Conflict Data” Variable Tab 40</p> <p>“Expected Utility” Variable Tab 38</p> <p>“General” Variable Tab 35</p> <p>“Polity III” Variable Tab 37</p> <p>6</p> <p>-66 11</p> <p>7</p> <p>-77 11</p> <p>8</p> <p>-88 11</p> <p>9</p> <p>-9 11</p> <p>-99 11</p> <p>A</p> <p>Acquiescence by A Equilibrium Outcome 21</p> <p>Acquiescence by B Equilibrium Outcome 22</p> <p>adjusted capabilities 18</p> <p>Adjusting “Highest Action” level 64</p> <p>All Contiguous Dyads 31, 32</p> <p>All Dyads from All Countries 31, 32</p> <p>All Dyads within a Specified Distance s 32, 33</p> <p>All Dyads within Specified Region(s) 31, 33</p> <p>All Major Power vs. Any State Dyads 31, 32</p> <p>All Major Power vs. Major Power Dyads 31, 32</p> <p>All Major XE "Country-Year Selection:All Major Powers" Powers 30</p> <p>All Politically Relevant Dyads 31, 33</p> <p>All States 30</p> <p>All Years 30</p> <p>alpha 20</p> <p>Always Include Reverse Directed Dyad 10</p> <p>Anomalies in Joiner Codings 70</p> <p>Austria-Hungary 62</p> <p>Authors 1</p> <p>B</p> <p>Browser Menu 4</p> <p>Bug Reports 1</p> <p>Bugs 81</p> <p style="padding-left: 20px;">Memory violation/access error 81</p> <p style="padding-left: 20px;">time to completion 81</p>	<p>Trace window remains onscreen 81</p> <p>C</p> <p><i>cap93.csv</i> 77</p> <p>Capabilities 15</p> <p style="padding-left: 20px;">Modifying 78</p> <p>capabilities, adjusted vs. raw 18</p> <p>Capabilities, computing 23</p> <p>Capital to capital 23</p> <p>Capital to capital, with contiguity 23</p> <p>Capital to capital, with contiguity and location 23</p> <p>Capitulation by A Equilibrium Outcome 22</p> <p>Capitulation by B Equilibrium Outcome 22</p> <p>Case Selection 9</p> <p>CD Installation <i>See</i> Installation, CD</p> <p>citation 88</p> <p>Citation 2</p> <p>City Location Data 78</p> <p>Combining include/exclude specifications 71</p> <p>Command files 14</p> <p>Command Files 9</p> <p>Command Files Only 9</p> <p>Composite Capabilities Index 23</p> <p>Composite Capability Index 15</p> <p>Concentration 22</p> <p>Conditions of Use 83</p> <p>Conflict 10</p> <p>Contact 1</p> <p><i>contigu.raw</i> 78</p> <p>Contiguity 78</p> <p>Copyright 83</p> <p><i>Country-Year Data</i></p> <p style="padding-left: 20px;">As unit of analysis 7</p> <p>Country-Year Selection 30</p> <p style="padding-left: 20px;">All Major Powers 30</p> <p style="padding-left: 20px;">All States 30</p> <p style="padding-left: 20px;">Specific Set of Countries 31</p> <p>Country-Years Included 9</p> <p>COW Dyadic MID Data 40</p> <p><i>cowmod.csv</i> 78</p> <p>Create Command File Only 9</p> <p>Create Command File(s) 9</p> <p>Create command files 14</p> <p>Create Command Files Only 9</p> <p>Creating Dyadic MID Variables 65</p> <p>Creating Dyadic MIDs and Meshing MID Data Sets 63</p> <p>D</p> <p>Data Sources 88</p> <p>Default Specifications for Pre-Calculated Data 87</p> <p>Deinstalling EUGene 4</p>
--	--

Delphi	2	Execution Time	2
Differences in Data	26	Exiting EUGene	14
<i>Directed Dispute-Dyad Data</i>		Expected Utility (The War Trap), calculating	17
As unit of analysis	8	<i>Expected Utility Calculations</i>	
Directed Dispute-Dyad Selection	33	menu selection	5, 6
One Case Per Directed Dispute Dyad Initiation	34	Extending EUGene	84
One Case Per Directed Dispute Dyad Year	34	F	
<i>Directed Dyad-Year Data</i>		File Extensions	14
As unit of analysis	7	File integrity	7
Directed Dyad-Year Selection	31	File Menu	<i>See Menus, File</i>
All Contiguous	31	Files	75
All Dyads from All Countries	31	cap93.csv	77
All Dyads within a Specified Distance	32, 33	contigu.raw	78
All Dyads within Specified Region(s)	31, 33	cowmod.csv	78
Dyads Read from User File	32	eugene.ini	75
Major Power Dyads	31	EUWarReaTau.dat	81
Major Powers vs. Any State	31	Euwartrp.dat	80
Politically Relevant	31	Input and Configuration Files	75
Specific Set	32	Intermediate	80
Disclaimer of Warranty	84	Log File	75
Dispute Dyad Output	74	majors.csv	79
Basic output	74	Online Help	79
Include Joiners	74	polity3.dat	79
Treat Joiners as Initiators	74	riskv2.dat	80
Dispute Unit	9	s.dat	80
Distance	23	security_alliances.dat	81
<i>Distance Discounting Options</i>	6	states.raw	79
Distance, calculating	17	syscap.dat	80
Distance, computing	23	tau.dat	80
Download installation	<i>See Installation</i>	userdyad.txt	79
Drop All Dyads with an Ongoing MID	67	WTRRisk.txt	79
Drop Joiner Directed Dyads	69	G	
Drop Target vs. Initiator Directed Dyads if no new MID	69	gamma	20
<u>Dyad Years with Ongoing Conflicts</u>	10	Generate Data Set Menu	<i>See Menus, Generate Data Set</i>
Dyadic MIDs	63	Genetic Algorithm	7, 25
<u>Dyads Involving Joiners</u>	11	Convergence	26
Dyads Read from User File	32, 33	Mutation	26
Dyad-Year Output	72	Number of Top Patterns to Maintain	26
Basic Output	72	Population	26
Coding Joiners As Initiators	73	Search Tolerance	26
Coding Joiners As Initiators, Dropping Joiners	73	Germany	62
Drop Joiners	73	H	
Keep Targets	72	Header Line	10
E		Help Menu	<i>See Menus, Help</i>
Equilibria (War and Reason)	21	Highest Action level	64
Equilibrium Computation, Logical vs. Backwards		Hypothetical Alliances	11
Induction	21	I	
<i>eugene.ini</i>	75	ICB Crisis Data	42
EUWarReaTau.dat	81	Include All Dyads with an Ongoing Dispute	10
Euwartrp.dat	80	Include All Dyads with an Ongoing MID	67
<u>Exclude Conflict Dyad-Years Based on Which Data Set?</u>	10		
Exclude/Include specifications, combining	71		

Include all Joiner Dyads	69	Converting COW MID Data to Directed Dyads	65
Include Dyads i vs. i	9	Marking Continuing Dispute Years as Initiations	68
Include Header Line	10	MID Data Sets	63
Include Ongoing Dispute Dyad Year iff New Dispute	10, 67	MID Variables	65
Include/exclude specifications, combining	71	Missing Values	11, 30
initialization	4	<i>Modifications to COW Capabilities</i>	6, 23, 78
Initiator / Multiple MID Settings	43	Modified Capability Values	85
Initiator Coding		Modified Values in Modified Capabilities Data File	85
Identity	44	Modifying EUGene	84
Timing	43	Movement	22
Input and Configuration Files	75	Multiple MIDs in Year	45
Installation	3	Mutation Probability	26
CD	3	N	
Download	3	Nation Membership data	79
intermediate files	7	<i>National Capabilities Index</i>	23
Intermediate Files	80	menu selection	5
Internal Details for Programmers	81	National Capabilities, computing	15
interrupting EUGene		Negotiation Equilibrium Outcome	22
potential problems	7	No Distance Discounting	23
Interrupting EUGene	13	Non-Directed Dispute Data Selection	34
J		Non-Directed Dispute-Dyad Selection	
Joiner Coding Issues	70	One Case Per Non-Directed Dispute Dyad Year	35
Joiner Variables	70	<i>Non-Directed Dyad-Year Data</i>	
Joiners	41, 69	As unit of analysis	7
K		Non-directed Dyad-Year Selection	32
Keep Target vs. Initiator Directed Dyads if no new		All Contiguous	32
MID	69	All Dyads from All Countries	32
key MID	45	Dyads Read from User File	33
Known Bugs and Problems	81	Major Power Dyads	32
L		Major Powers vs. Any State	32
Legal Notice	83	Politically Relevant	33
LIMDEP	11, 14, 30, 63	Specific Set	33
Loading Settings	4	Number of Iterations	25
Log File	75	Number of Top Patterns to Maintain	26
Logical Equilibrium Conditions	21	O	
M		One Case Per Directed Dispute Dyad Initiation	34
<i>majors.csv</i>	79	One Case Per Directed Dispute Dyad Year	34
Maoz Dyadic MID Data	40	One Case Per Non-Directed Dispute Dyad Year	35
Memory violation/access error	81	Ongoing Dispute Year Options	67
Menus		Ongoing Dispute Years	68
File	4	Online Help	13
Generate Data Set	7	Online Help files	79
Help	13	Output Choices	8
Overview	4	Output Destination	8
Recompute	5	Overview	1
Trace	13	P	
User Data	13	Peace Years	42
Meshing MID Data Sets	63	phi	20, 21
MID Inclusion	10	Polity III data	79
MID Data	40	missing values	63
		Polity III Data	

merging	62	Shortcuts, Creating	4
recoding	62	Software	
<i>polity3.dat</i>	79	reading data into other	14
Population Size	26	SPSS, Stata, LIMDEP	14
Pre-Calculated Data	87	Verification	26, 28
Prepare Data Set for Submission	13	Sources, Data	88
Program Extensions and Modifications	84	Specific Set of Countries	31
Program Files	75	Specific Set of Dyads	32, 33
program initialization	4	Specifications	2
Program Requirements	2	SPSS	14
Programmers	81	Stable Iterations for Convergence	26
R		Stata	14
random walk	7	<i>states.raw</i>	79
Random Walk	24	Status Quo Equilibrium Outcome	22
Number of Iterations	25	Steepest Descent	6, 24
Raw Alliance data	78	Tolerance	24
Raw capability data	77	Summary	1
Reading Output Data	14	<i>syscap.dat</i>	80
ReCompute Menu	5, <i>See</i> Menus, ReCompute	System Concentration	22
recomputing variables	6	System Movement	22
dependence on prior calculations	6	T	
Recomputing variables		Target vs. Initiator Dyad	68
options	6	tau (utility term)	20
ReComputing variables	5	tau.dat	80
Regional Identification	16	<i>Tau-b Scores</i>	
relevant region	15, 16, 17, 18, 20, 24, 38, 39	menu selection	5
Relevant Region	16, 36	Tau-b, calculating	15
Removing EUGene	4	Technical Support	1
Requirements	2	time to completion	81
retrospective sampling	9	Time, Execution	<i>See</i> Execution Time
<i>Risk Attitude Optimization Method</i>	6, 23	<i>Tolstoy</i>	1
Genetic Algorithm	25	Trace Menu	<i>See</i> Menus, Trace
Random Walk	24	Trace window	81
Steepest Descent	24	Transfer User Data Sets to/from Website	13
Risk Attitude, calculating	18	Treat Continuing Dispute Years as Initiations	10
Risk Attitude, computing	23	Treating Continuing Dispute Years as Initiations	68
<i>Risk Data Source</i>	7, 26	U	
Risk Scores	7	Uncertainty, calculating	17
Risk Search Tolerance	24	Uninstalling EUGene	4
risky2.dat	80	Unit of Analysis	7
Run Time	<i>See</i> Execution Time	Use Specific Random Number Seed	10
S		User Data Menu	<i>See</i> Menus, User Data
<i>S Scores</i>		User Data Sets	13
menu selection	5	User Dyad List	79
S, calculating	16	user interrupt	81
s.dat	80	User Variables	46
Sampling	9	<i>userdyad.txt</i>	79
Saving Settings	4	Utility (War and Reason), calculating	20
Search Tolerance	26	V	
security_alliances.dat	81	Variable Calculations	15
Selected Years	30	Variable Names	46
<u>Self-Referencing Dyads</u>	9	Variable Order	46
Setup	3		

Variable Separator	9	Y	
Variables	10	Years to Include in Output	9, 30
Variables Available	35	All Years	30
Variables to Include in Output	10	Years to Include In Output	
Verifying Calculations	26	Selected Years	30
W			
War started by A Equilibrium Outcome	21	α	20
War started by B Equilibrium Outcome	21		
War Trap Revisited Risk data	79		
Warranty	84	γ	20
Werner Peace Years Option	42		
<i>WTRRisk.txt</i>	79		
		τ	20