Replication Materials for
Words of War: Negotiation as a Tool of Conflict

Eric Min

Last updated June 14, 2024

This document contains all information regarding the files necessary to replicate all results in the main manuscript and the appendix.

The document is divided into four main sections. In the first, we list and briefly describe the function of every single item in the replication file. In the second, we specify which R file produces a table, figure, or set of results that appears in the main text and appendix. In the third, we describe all objects in an R workspace that contains the results of our supervised learning and structural topic models. In the fourth, we describe every covariate in the CSV files that contain relevant data.

Technical Specifications **2**
Overview of Files **2**
RData Files 2
CSV Files 2
R Files 3
Objects in R Workspace **5**
korea_confusion_data.RData 5
Variables in CSV Files **6**
wardata_book.csv 6
negstarts.csv 9
iwbd1.0.csv 10
tpidata_book.csv 11
korea_report_data.csv 12
korea_speech_data.csv 13

Technical Specifications

All analyses were performed on a PC with Windows 11 Pro, using R version 4.3.1.

Overview of Files

RData Files

File	Description
korea_confusion_data.RData *(/data/)*	R workspace containing confusion matrices from all supervised learning models for sincere negotiation behavior during Kaesong and Panmunjom talks in the Korean War.

CSV Files

File	Description
wardata_book.csv *(/data/)*	The core war-day-level data on negotiations and battlefield activity for 92 interstate wars, 1816-2003.
negstarts.csv *(/data/)*	Data on the start and end dates for 189 negotiation periods across for 92 interstate wars, 1816-2003.
iwbd1.0.csv *(/data/)*	A copy of the Interstate War Battle Dataset, version 1.0 from Min 2022.
tpidata_book.csv *(/data/)*	War-day-level data on negotiations, battlefield activity, and third-party diplomatic interventions as coded in the International Conflict Management dataset (Bercovitch 2006) for post-1945 wars.
korea_report_data.csv *(/data/)*	Daily-level data of Korean War battlefield activity from daily operations reports generated by the United Nations Command.
korea_speech_data.csv *(/data/)*	Statement-level data of all statements made by the United Nations Command and Communist delegations during negotiations at Kaesong and Panmunjom during the Korean War. The data file also features predicted values of whether each statement made features sincere or insincere negotiation behavior as supplied by a random forest model, which had the best out-of-sample predictive performance.

R Files

File	Description
load_packages.R *(/function_code/)*	Installs and loads all necessary packages.
war_plots.R *(/function_code/)*	Creates plots illustrating five measures of battlefield activity for a given war.
relabel_vars.R *(/function_code/)*	Code used across all "coef_plot" files to help relabel variables to make them more accessible in all coefficient plots.
msm_coef_plot.R *(/function_code/)*	Code to produce coefficient plots for multistate models.
msm_coef_table.R *(/function_code/)*	Code to produce coefficient tables for multistate models.
hazard_coef_plot.R *(/function_code/)*	Code to produce coefficient plots for hazard/survival models.
battle_coef_plot.R *(/function_code/)*	Code to produce coefficient plots for Poisson regressions.
combine_rev_results.R *(/function_code/)*	Code to combine all results regarding changing battlefield trends during windows of time before and after negotiations that fail to end a war.
korea_coef_plot.R *(/function_code/)*	Code to produce coefficient plots for all analyses involving the Korean War in Chapter 6.
korea_heat_map.R *(/function_code/)*	Code to produce heat maps of predicted probabilities of sincere negotiation behavior, as well as to return predicted probabilities of sincere negotiation behavior at the lowest and highest observed levels of recent imbalance and recent costs, involving the Korean War in Chapter 6.
ch0_rep.R *(/analysis_code/)*	Replication code for the introduction chapter.
ch2_rep.R *(/analysis_code/)*	Replication code and appendix material for Chapter 2 ("Quantifying Two Centuries of War").
ch3_rep.R *(/analysis_code/)*	Replication code and appendix material for Chapter 3 ("Fighting to Talk").
ch4_rep.R *(/analysis_code/)*	Replication code and appendix material for Chapter 4 ("Talking to Fight").
ch6_rep.R *(/analysis_code/)*	Replication code and appendix material for Chapter 6 ("The 'Talking War' in Korea").

The code to replicate all tables, figures, and other statistical output in the book, as well as all material in the Online Appendix, is organized by chapter. Note that five chapters involve data analysis: the introduction (Chapter 0), Chapter 2, Chapter 3, Chapter 4, and Chapter 6. An R file exists for each chapter, adopting the naming convention "chX_rep.R," where X is replaced by the appropriate chapter number.

The table below summarizes which basic files are used in each of the main replication R code files for each chapter.

	Intro (ch0_rep.R)	**Chap. 2** (ch2_rep.R)	**Chap. 3** (ch3_rep.R)	**Chap. 4** (ch4_rep.R)	**Chap. 6** (ch6_rep.R)
CSV Files					
wardata_book.csv	x	x	x	x	
negstarts.csv		x	x	x	
iwbd1.0.csv		x			
tpidata_book.csv			x		
korea_report_data.csv					x
korea_speech_data.csv					x
R files					
war_plots.R		x			
msm_coef_plot.R			x	x	
msm_coef_table.R			x	x	
hazard_coef_plot.R			x		
relabel_vars.R			x*	x*	
battle_coef_plot.R				x	
combine_rev_results.R				x	
korea_coef_plot.R					x
korea_heat_map.R					x
RData file					
korea_confusion_data.RData					x

The asterisks for the "relabel_vars.R" file are meant to indicate that the file is not directly mentioned in the replication files for Chapters 3 or 4. Instead, "relabel_vars.R" is called by R files that contain functions which are used to produce results in those chapters.

Objects in R Workspace

korea_confusion_data.RData

R workspace containing confusion matrices from all supervised learning models for sincere negotiation behavior during Kaesong and Panmunjom talks in the Korean War.

- *confGB:* Performance metrics for gradient boosting models.
- *confNB:* Performance metrics for Naïve Bayes models.
- *confNN:* Performance metrics for neural network models.
- *confRF:* Performance metrics for random forest models.
- *confSVM:* Performance metrics for support vector machines.

Variables in CSV Files

wardata_book.csv

The core war-day-level data on negotiations and battlefield activity for 92 interstate wars, 1816-2003.

- *id:* A unique identifier at the war-date level, using war names from the Correlates of War.
- *id2:* A unique identifier at the war-date level, using war numbers from the Correlates of War.
- *warName:* War name, using names from the Correlates of War.
- *warNum:* War number, using numbers from the Correlates of War.
- *date:* Date of the war-day.
- *dayNum:* A running count of the days elapsed in a war.
- *dayNumNext:* A variable defined as *dayNum* + 1, which is used in the hazard/survival models.
- *logDayNum:* A variable defined as log(*dayNum*), which is used to help restore the proportional hazards assumption in some hazard/survival models.
- *post45:* A binary variable for whether a war-day is before January 1, 1945 or on/after January 1, 1945 (0 = no, 1 = yes).
- *propDone:* A variable indicating what proportion of the war has elapsed as a function of the war's overall duration.
- *propDoneR:* The same as *propDone*, but rounded down to the nearest 0.05.
- *neg:* A binary variable for whether a negotiation took place on the war-day (0 = no, 1 = yes).
- *negInt:* A binary variable for whether an internal negotiation took place on a war-day (0 = no, 1 = yes).
- *negExt:* A binary variable for whether an external negotiation took place on a war-day (0 = no, 1 = yes).
- *negIE:* A three-level variable for whether an internal negotiation, external negotiation, or no negotiation took place on a war-day.
- *negState3:* A three-level state variable for whether a war-day involved no negotiations (1), negotiations (2), or war termination (3).
- *negStateIE:* A three-level state variable for whether a war-day involved no negotiations (1), internal negotiations (2), or external negotiations (3).
- *outSumN:* Battlefield momentum over the previous N days, where N can take the value of 30, 60, or 90. Note that the analysis in the main manuscript uses N=60.

- *AOSN:* Recent battlefield imbalance over the previous N days, where N can take the value of 30, 60, or 90. This measure is simply the absolute value of *outSumN*. Note that the analysis in the main manuscript uses N=60.
- *outAll:* Battlefield position over the entire course of the war up to the war-day in question.
- *AOA:* Overall battlefield imbalance over the course of the war up to the war-day in question. This measure is simply the absolute value of *outAll*.
- *apropdiff:* Battlefield inconsistency, which is measured by taking the absolute value of the difference between CINC ratio (*cincP* below) and the proportion of battles won by the war initiator by the war-day in question.
- *nBattle:* The number of active battles taking place on a war-day.
- *nBattleTrend:* The linear trend in number of active battles in the previous 60 days or since the start of the war (whichever is closer).
- *terminate:* A binary variable for whether the war-day is the final day of the war (0 = no, 1 = yes).
- *doeWinLose:* A measure of relative capabilities constructed using the Dispute Outcome Expectations (DOE) score by Carroll and Kenkel (2019). See Appendix 3.4 for more information.
- *apropdiffDOE:* A measure of inconsistency produced using the *doeWinLose* measure instead of the *cincP* measure (see below).
- *ceasefire:* A binary variable for whether a ceasefire was in place on a given war-day (0 = no, 1 = yes).
- *issues:* The sum of issue salience measures for both belligerents. The issue salience measure can range from 0 (no salience) to 2 (highest salience) for each side. This measure takes the sum of these measures across both sides. If one side has multiple states, the highest issue salience is used to represent the entire side.
- *contig:* A binary variable for whether the belligerents (or in multilateral wars, any two belligerents) share a land or water border (0 = no, 1 = yes).
- *cincP:* The ratio of CINC scores, constructed by taking the sum of all CINC scores on the war initiator's side and dividing it by the sum of all CINC scores across all active belligerents.
- *polBinA:* A binary variable for whether the war initiator features a democracy, where a state is considered a democracy if it has a Polity score of 6 or higher (0 = no, 1 = yes).
- *nuk:* A binary variable for whether any active belligerent is a nuclear power (0 = no, 1 = yes).
- *maj:* A binary variable for whether any active belligerent is a major power (0 = no, 1 = yes).
- *minor:* A binary variable for whether all active belligerents are minor—that is, not major—powers (0 = no, 1 = yes).

- *keyAllies:* A binary variable for whether any active belligerent has an alliance with a major power in pre-1945 wars or one of the two superpowers—the United States or the Soviet Union—in post-1945 wars (0 = no, 1 = yes).
- *nBellig:* The number of active belligerents in a war on a given war-day.
- *logBattles:* The (logged) number of total battles that have taken place by the given war-day.
- *anyDipEnemy:* A binary variable for whether any diplomatic relations exist between any two opposing belligerents.

negstarts.csv

Dataset of negotiation periods across 92 interstate wars.

- *id:* A unique identifier at the war-date level, using war names from the Correlates of War. The date in this identifier corresponds to the first day of the negotiation period.
- *warName:* War name, using names from the Correlates of War.
- *resNum:* The ordinal count of the negotiation period in the war.
- *startdate:* The first day of the negotiation period.
- *enddate:* The last day of the negotiation period.
- *negLength:* The overall length of the negotiation period.
- *pressure:* A binary variable indicating whether the negotiation period was borne of third-party pressures (0 = no, 1 = yes). Note that this variable is used to determine whether the war-days associated with a negotiation period are internal or external in the "wardata_book.csv" file.
- *endwar:* A binary variable indicating whether the negotiation period was immediately followed by the termination of conflict (0 = no, 1 = yes).

iwbd1.0.csv

The Interstate War Battle Dataset (1823-2003), version 1.0, from Min 2021. See the original publication and online appendices for more information on data creation and definitions.

- *cowNum:* War number, using numbers from the Correlates of War.
- *iwdNum:* War number, using numbers from the Interstate War Dataset (Reiter, Stam, and Horowitz 2016).
- *warName:* War name, using names from the Correlates of War.
- *battleName:* Unique name for a battle.
- *attacker:* The name of the belligerent (or set of belligerents) that started the attack in the battle.
- *defender:* The name of the belligerent (or set of belligerents) that were attacked in the battle.
- *startYear:* The year in which the battle began.
- *startMonth:* The month in which the battle began.
- *startDay:* The day on which the battle began.
- *startDate:* The date on which the battle began.
- *endYear:* The year in which the battle ended.
- *endMonth:* The month in which the battle ended.
- *endDay:* The day on which the battle ended.
- *endDate:* The date on which the battle ended.
- *battleLength:* The overall length of the battle in days.
- *victor:* The name of the belligerent (or set of belligerents) that won the battle.
- *victorWarLevel:* Whether the battle was won by the overall war initiator or overall war target.
- *victorBattleLevel:* Whether the battle was won by the battle attacker or battle defender.

tpidata_book.csv

The core war-day-level data on negotiations and battlefield activity for 92 interstate wars, 1816-2003, with additional third-party intervention data added. Note that this dataset is very similar to “wardata_book.csv,” but it is limited to post-1945 wars and includes two additional variables that do not exist in “wardata_book.csv.” Both variables are derived from the International Conflict Management (ICM) dataset (Bercovitch 1999). Only variables that are new to “tpidata_book.csv” will be described below; refer to the entry on “wardata_book.csv” for information on all other measures.

- *intervention:* A binary variable for whether a third-party diplomatic intervention took place on a given war-day, according to the ICM dataset (0 = no, 1 = yes).
- *tpi_outcome_2:* A variable for whether a war-day involved any diplomatic concessions by an active belligerent, according to the ICM dataset (0 = no, 1 = yes).

korea_report_data.csv

Dataset of battlefield activity during the Korean War, as described by daily military operations reports from the United Nations Command.

- *date:* The date of the war.
- *dayPosN:* The proportion of all status words used in reports that express positive movement (battlefield advances) for the UNC. The measure is an exponentially weighted moving average using the previous N days, where N can take the value of 30, 60, or 90. Note that the analysis in the main manuscript uses N=60.
- *dayNegN:* The proportion of all status words used in reports that express negative movement (battlefield retreats or withdrawals) for the UNC. The measure is an exponentially weighted moving average using the previous N days, where N can take the value of 30, 60, or 90. Note that the analysis in the main manuscript uses N=60.
- *rollBothLossNl:* The logged number of all recent UNC casualties and Communist POWs. The measure is an exponentially weighted moving average using the previous N days, where N can take the value of 30, 60, or 90. Note that the analysis in the main manuscript uses N=60.
- *rollUNCLossNl:* The logged number of all recent UNC casualties. The measure is an exponentially weighted moving average using the previous N days, where N can take the value of 30, 60, or 90. Note that the analysis in the main manuscript uses N=60.
- *rollPOWNl:* The logged number of all recent Communist POWs taken in by the United Nations Command. The measure is an exponentially weighted moving average using the previous N days, where N can take the value of 30, 60, or 90. Note that the analysis in the main manuscript uses N=60.

korea_speech_data.csv

A dataset containing all statements made by the United Nations Command and Communist delegations during negotiations in Kaesong and Panmunjom, as well as contemporaneous battlefield activity.

- *date:* The date of the war.
- *phase:* A variable indicating whether the war-day in question is part of the first, second, or third round of negotiations.
- *meetTypeMain:* A variable indicating whether a given statement was made at the plenary, sub-delegation, staff officer, or other level.
- *item:* A variable indicating whether a given statement is explicitly associated with discussion of an agenda item.
- *speaker:* A variable indicating whether a given statement was made by the United Nations Command or the Communist states.
- *meetID:* A unique identifier for an individual meeting.
- *speakMeetID:* A unique identifier for the speaker and meeting.
- *speechID:* A unique identifier for an individual statement.
- *speech:* The text of the statement made in negotiations.
- *meanClass:* A binary classification of whether a given statement is considered to be insincere (0) or sincere (1) according to a random forest model.
- *riN:* A measure of recent imbalance based on the last N days of battlefield activity. The measure is constructed by taking the absolute value of the difference between *dayPosN* and *dayNegN*, defined below. Note that N can take the value of 30, 60, or 90; the analysis in the main manuscript uses N=60.
- *dayPosN:* The proportion of all status words used in reports that express positive movement (battlefield advances) for the UNC. The measure is an exponentially weighted moving average using the previous N days, where N can take the value of 30, 60, or 90. Note that the analysis in the main manuscript uses N=60.
- *dayNegN:* The proportion of all status words used in reports that express negative movement (battlefield retreats or withdrawals) for the UNC. The measure is an exponentially weighted moving average using the previous N days, where N can take the value of 30, 60, or 90. Note that the analysis in the main manuscript uses N=60.
- *rollBothLossNl:* The logged number of all recent UNC casualties and Communist POWs. The measure is an exponentially weighted moving average using the previous N days, where N can take the value of 30, 60, or 90. Note that the analysis in the main manuscript uses N=60.
- *rollUNCLossNl:* The logged number of all recent UNC casualties. The measure is an exponentially weighted moving average using the previous N days, where N can take the value of 30, 60, or 90. Note that the analysis in the main manuscript uses N=60.

- *rollPOWNl:* The logged number of all recent Communist POWs taken in by the United Nations Command. The measure is an exponentially weighted moving average using the previous N days, where N can take the value of 30, 60, or 90. Note that the analysis in the main manuscript uses N=60.
- *logCombatSorties:* The logged number of combat sorties conducted by the United Nations Command on a given day.
- *logRepLength:* The logged number of total words in the day's United Nations Command daily operations report.
- *winter:* A binary variable for whether a statement was made during the winter months of December, January, or February (0 = no, 1 = yes).
- *eisenhower:* A binary variable for whether a statement was made after Eisenhower's inauguration on January 20, 1953 (0 = no, 1 = yes).
- *stalin:* A binary variable for whether a statement was made during the time following Stalin's death on March 5, 1953 (0 = no, 1 = yes).
- *election:* A binary variable for whether a statement was made during the time of the 1952 presidential election, between July 11 and November 4, 1952 (0 = no, 1 = yes).