Introducing ICBe: Very High Recall and Precision Event Extraction from Narratives about International Crises

A Preprint

Rex W. Douglass *
University of California, San Diego

Thomas Leo Scherer
University of California, San Diego

J. Andrés Gannon
Vanderbilt University

Erik Gartzke
University of California, San Diego

Jon Lindsay
Georgia Institute of Technology

Shannon Carcelli
University of Maryland

Jonathan Wilkenfeld
University of Maryland

David M. Quinn
University of Maryland

Catherine Aiken
Georgetown University

Jose Miguel Cabezas Navarro
Universidad Mayor

Neil Lund
University of Maryland

Egle Murauskaite
University of Maryland

Diana Partridge
University of Maryland

September 9, 2022

Abstract

How do international crises unfold? We conceptualize of international relations as a strategic chess game between adversaries and develop a systematic way to measure pieces, moves, and gambits accurately and consistently over a hundred years of history. We introduce a new ontology and dataset of international events called ICBe based on a very high-quality corpus of narratives from the International Crisis Behavior (ICB) Project. We demonstrate that ICBe has higher coverage, recall, and precision than existing state of the art datasets and conduct two detailed case studies of the Cuban Missile Crisis (1962) and Crimea-Donbas Crisis (2014). We further introduce two new event visualizations (event icongraphy and crisis maps), an automated benchmark for measuring event recall using natural language processing (sythnetic narratives), and an ontology reconstruction task for objectively measuring event precision. We make the data, online appendix, replication material, and visualizations of every historical episode available at a companion website www.crisisevents.org and the github repository.

Keywords Diplomacy · War · Crises · International Affairs · Computational Social Science

*Correspondence should be addressed to Rex W. Douglass at `rexdouglass@gmail.com`.

1 Introduction

If we could record every important interaction between countries in all of diplomacy, military conflict, and international political economy, how much unique information would this chronicle amount to, and how surprised would we be to see something new? In other words, what is the entropy of international relations? This record could in principle be unbounded, but the central conceit of social science is that there are structural regularities that limit what actors can do, their best options, and even which actors are likely to survive (Brecher 1999; Reiter 2015). If so, then these events can be systematically measured, and accordingly, massive effort is expended in social science attempting to record these regularities.[2] Thanks to improvements in natural language processing, more open-ended efforts have begun to capture entire unstructured streams of events from international news reports.[3] How close these efforts are to accurately measuring all or even most of what is essential in international relations is an open empirical question, one for which we provide new evidence here.

Our contribution is a high coverage ontology and event dataset for key historical episodes in 20th and 21st-century international relations. We develop a large, flexible ontology of international events with the help of both human coders and natural language processing. We apply it sentence-by-sentence to an unusually high-quality corpus of historical narratives of international crises (Brecher 1999; Brecher, James, and Wilkenfeld 2000; Wilkenfeld and Brecher 2000; James 2019; Iakhnis and James 2019). The result is a new lower bound estimate of how much actually happens between states during pivotal historical episodes. We then develop several methods for objectively gauging how well these event codings reconstruct the information contained in the original narrative. We conclude by benchmarking our event codings against several current state-of-the-art event data collection efforts. We find that existing systems produce sequences of events that do not contain enough information to reconstruct the underlying historical episode. The underlying fine-grained variation in international affairs is unrecognizable through the lens of current quantification efforts.

This is a measurement paper that makes the following argument — there is a real-world unobserved latent concept known as international relations, we propose a method for systematically measuring it, we successfully apply this method producing a new large scale set of measurements, those measurements exhibit several desirable kinds of internal and external validity, and those measurements out-perform other existing approaches. The article organizes that argument into eight sections: task definition; corpus; priors/existing state of the art; ICBe coding process; internal consistency; case study selection; recall; and precision. A final section concludes.

2 Task Definition

We consider the measurement task of abstracting discrete events about a historical episode in international relations. The easiest way to convey the task is with an example. Figure 1 shows a narrative account of the Cuban Missile Crisis (1962) alongside a mapping from each natural language sentence to discrete machine readable abstractive events. Formally, a historical episode, H, is demarcated by a period of time $[T_{start}, T_{end}] \in T$, a set of Players $p \in P$, and a set of behaviors they undertook during that time $b \in B$. International Relations, IR, is the system of regularities that govern the strategic interactions that world actors make during a historical episode, given their available options, preferences, beliefs, and expectations of choices made by others. We observe neither H nor IR directly. Rather the Historical Record, HR, produces documents $d \in D$ containing some relevant and true (as well as irrelevant and untrue) information about behaviors that were undertaken recorded in the form of unstructured natural language text. The task is to combine informative priors about IR with an unstructured corpus D to produce a series of structured discrete events, $e \in E$, that have high coverage, precision, and recall over what actually took place in history, H.

[2] See work on crises (Brecher and Wilkenfeld 1982; Beardsley et al. 2020), militarized disputes (Palmer et al. 2021; Gibler 2018; Maoz et al. 2019), wars (Sarkees and Wayman 2010; Reiter, Stam, and Horowitz 2016), organized violence (Ralph Sundberg and Mihai Croicu 2016; Pettersson and Eck 2018), political violence (Raleigh et al. 2010), sanctions (Felbermayr et al. 2020), trade (Barari and Kim, n.d.), and international agreements (Kinne 2020; Owsiak, Cuttner, and Buck 2018; Vabulas and Snidal 2021), dispute resolution (Vabulas and Snidal 2021; Frederick, Hensel, and Macaulay 2017), and diplomacy (Moyer, Turner, and Meisel 2020; Sechser 2011).

[3] See Li et al. (2021); Halterman (2020); Brandt et al. (2018); Boschee et al. (2015); Hegre et al. (2020); Grant et al. (2017). On event-extraction from images and social-media see Zhang and Pan (2019) and Steinert-Threlkeld (2019).

Figure 1: Case Study 1: Cuban Missile Crisis (1962) - ICB Narrative vs. ICBe Events

S	Natural Language Sentences (ICB Corpus)	Machine Readable Events (ICBe)
4	When the U.S. discovered the presence of Soviet military personnel in Cuba on 7 September 1962 it called up 150,000 reservists.	mobilization 100ks discover fact - deployment to area 100s;1ks
5	The Soviets mobilized on the 11th.	mobilization 1ks
6	Although persistent rumors circulated concerning the deployment of Soviet missiles in Cuba, Soviet Ambassador Anatoly Dobrynin denied the charges, and Premier Khrushchev gave his personal assurances that ground-to-ground missiles would never be shipped to Cuba.	
8	The U.S. crisis was triggered on 16 October when the CIA presented to President Kennedy photographic evidence of the presence of Soviet missiles in Cuba.	discover fact;start of crisis - deployment to area fortify start of crisis - deployment to area
9	The U.S. responded with a decision on the 20th to blockade all offensive military equipment en route to Cuba.	blockade coastline 10ks blockade coastline
10	When this was announced on 22 October, a crisis was triggered for Cuba and the USSR. An urgent meeting of the UN Security Council was requested by both the U.S. and Cuba on the 22nd, and by the USSR the next day.	start of crisis - appeal - meeting
11	On the 23rd as well, the Soviets accused the United States of violating the UN Charter and announced an alert of its armed forces and those of the Warsaw Pact members.	raise in alert 100ks accuse - violate terms of treaty
12	That day Cuba responded by condemning the U.S. blockade and declaring its willingness to fight.	- coastline disapprove
13	A resolution was adopted on the 23rd by the OAS calling for the withdrawal of the missiles from Cuba and recommending that member-states take all measures, including the use of force, to ensure that the government of Cuba would not continue to receive military material.	end military cooperation demand - withdraw from area 1ks
14	On 24 October the Security Council adopted a resolution requesting the Secretary-General to confer with the parties.	express intent - mediation express intent - mediation
15	On that same day, U Thant began mediation by sending identical letters to Khrushchev and Kennedy which proposed that the Soviet Union and the United States enter into negotiations, during which period both the shipment of arms and the quarantine would be suspended.	mediation
16	Moscow's major response to the crisis was a letter from Khrushchev to Kennedy on 26 October offering the removal of Soviet offensive weapons from Cuba and the cessation of further shipments in exchange for an end to the U.S. quarantine and a U.S. assurance that it would not invade Cuba.	offer happens - withdraw from area -C- will happen end blockade offer happens - -C- will happen end blockade
17	The situation was exacerbated on the 27th when a U.S. U-2 surveillance plane was shot down.	border violation airspace 1s 1s
18	That day another Khrushchev letter was received in Washington offering the removal of Soviet missiles from Cuba in exchange for the removal of U.S. missiles from Turkey.	offer happens - withdraw from area -C- will happen withdraw from area offer happens - -C- will happen
19	U.S. mobilization and aerial reconnaissance flights were stepped up.	border violation;mobilization mobilization airspace
20	And on the 27th President Kennedy sent the Soviet premier an acceptance of the proposals contained in the letter of 26 October while making no reference to Khrushchev's second letter of the 27th.	accept -
21	The following day Khrushchev notified the U.S. government that he had ordered work on the missile sites in Cuba stopped.	accept -
23	At the same time he warned Washington that U-2 reconnaissance flights over Cuba must be stopped as well.	demand - withdraw from area airspace
24	The crisis continued at a lower level of intensity for several more weeks due to Cuban President Castro's demands concerning a U.S. pledge not to invade his country.	demand - de-mobilization;lower alert body of water;coastline 100ks
25	On 30 October U Thant began talks in Havana, and Kennedy agreed to lift the quarantine for the duration of the talks.	end blockade coastline 10ks accept - end blockade coastline
26	When Cuba rejected UN inspection, the U.S. resumed the quarantine and air surveillance.	reject - unspecified cooperation reject blockade airspace;coastline 10ks
27	The Kremlin sent Deputy Premier Anastas Mikoyan to Cuba on 2 November to try to persuade Castro to allow UN inspection.	discussion;meeting
28	When this proved unsuccessful, a U.S.-USSR agreement was reached on 7 November allowing U.S. inspection and interception of Soviet ships leaving Cuba and the photographing of the missiles.	sign formal agreement
29	The following day the superpowers negotiated the removal of the IL-28 bombers which Castro had claimed were Cuban property.	withdraw from area 100s sign formal agreement
30	Castro's agreement was conveyed to the U.S. on 20 November 1962, which terminated the Missile crisis for all three actors.	end of crisis
31	The U.S. naval quarantine was lifted immediately, but aerial surveillance continued until the agreement was completely carried out.	end blockade body of water;coastline 10ks

3 Corpus

For our corpus, D, we select a set of unusually high-quality historical narratives from the International Crisis Behavior (ICB) project ($n = 471$)(SI Appendix, Table A1)(Brecher et al. 2017; Brecher and Wilkenfeld 1997).[4] Their domain is 20th and 21st-century crises, defined as a change in the type, or an increase in the intensity, of disruptive interaction with a heightened probability of military hostilities that destabilizes states' relationships or challenges the structure of the international system (Brecher and Wilkenfeld 1982).[5] Crises are a significant focus of detailed single case studies and case comparisons because they provide an opportunity to examine behaviors in IR short of, or at least prior to, full conflict (Holsti 1965; Paige 1968; Allison and Zelikow 1971; Snyder and Diesing 1977; Gavin 2014; George and Smoke 1974; Brecher and Wilkenfeld 1982; Gaddis 1987; Brecher and James 1988). Case selection was exhaustive based on a survey of world news archives and region experts, cross-checked against other databases of war and conflict, and non-English sources (Kang and Lin 2019; Brecher et al. 2017, 59). Each narrative was written by consensus by a small number of scholars, using a uniform coding scheme, with similar specificity (Hewitt 2001). The corpus is unique in IR because it is designed to be used in a downstream quantitative coding project.

4 Prior Beliefs about IR, Ontological Coverage, and the Existing State of the Art

Next we draw informative prior beliefs about the underlying process of IR that we expect to govern behavior during historical episodes and their conversion to the historical record. We organize our prior beliefs along two overarching axes, summarized in detail by Table 1.

The first axis (rows) represents the types of information we expect to find in IR and forms the basis for our proposed ontology. We employ a metaphor of a chess game, with players (polities, rebel groups, IGOs, etc.), pieces (military platforms, civilians, domains), and behaviors (think, say, do). Precise sequencing is required to capture gambits (sequences of moves) and outcomes (victory, defeat, peace, etc.), while precise geo-coding is required to understand the chessboard (medium of conflict). We find 472 actors and 117 different behaviors, and provide a full codebook in the online material.[6]

We base our informed priors primarily on two sources of information. The first is the extensive existing measurement efforts of IR which we provide citations to alongside each concept. Second, we performed preliminary natural language processing of the corpus and identified named entities and behaviors mentioned in the text. Verbs were matched to the most likely definition found in Wordnet (Miller 1995), tallied, and then aggregated into a smaller number hypernyms balancing conceptual detail and manageable sparsity for human coding (SI Appendix, Table A2).

The second axis (columns) compares the very high ontological coverage of ICBe to existing state of the art systems in production and with global coverage. They begin with our contribution ICBe, alongside other event-level datasets including CAMEO dictionary lookup based systems (Historical Phoenix (Althaus et al. 2019); ICEWS (Boschee et al. 2015; Hegre et al. 2020); Terrier (Grant et al. 2017)), the Militarized Interstate Disputes Incidents dataset, and the UCDP-GED dataset (Ralph Sundberg and Mihai Croicu 2016; Pettersson and Eck 2018; Sundberg and Melander 2013).[7] The final set of columns compares episode-level datasets beginning with the original ICB project (Brecher et al., n.d.; Brecher and Wilkenfeld 1982; Beardsley et al. 2020); the Militarized Interstate Disputes dataset (Palmer et al. 2021; Gibler 2018; Braithwaite 2010, 2009), and the Correlates of War (Sarkees and Wayman 2010). With the exception of large scale CAMEO dictionary based systems, the existing state of the art quantitative datasets ignore the vast majority of the information content found in international relations.[8]

[4] The Online Appendix is at the ICBEventData Github Repository.

[5] On near crises see Iakhnis and James (2019).

[6] See the Github Repository ICBEventData.

[7] Additional relevant but dated or too small of an overlap in domain include BCOW (Leng and Singer 1988), WEIS (McClelland 1978), CREON (Hermann 1984), CASCON (Bloomfield and Moulton 1989), SHERFACS (Sherman 2000), Real-Time Phoenix (Brandt et al. 2018), and COfEE (Balali, Asadpour, and Jafari 2021) (see histories in Merritt (1994) and Schrodt and Hall (2006)).

[8] See (Balali, Asadpour, and Jafari 2021) for a recent review of ontological depth and availability of Gold Standard example text.

Table 1: Ontological coverage of ICBe versus existing State of the Art

		Events Datasets						Episodes Datasets		
	Concept	ICBe	Phoenix	Terrier	ICEWs	MIDs Incidents	UCDP-GED	ICB	MIDs	COW
Domain	Start (Brecher 1993; Hegre and Sambanis 2006; Iakhnis and James 2019; Sambanis 2004)	1918	1945	1977	1995	1993	1989	1918	1816	1816
	End (Kreutz 2010; Weisiger 2016/ed)	2017	2019	2018	2020	2010	2015	2017	2014	2007
	N	32k	8.5M	28.4M	17.5M	9.6K	128k	1K	5.9K	1K
	Coders	Hand	Automated (CAMEO)			Hand	Hand	Hand	Hand	Hand
	Corpus	ICB	News			Mix	News	Mix	Mix	Mix
	Date	Event	Article			Event	Article	Event	Event	Event
	Location	Event	Event			Actor	Event	Actor	Event	Actor
Players	States (Fazal 2011; Paul, Ikenberry, and Hall 2020; Ryan 2021; Spruyt 1996)	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Subnational Actors (Haffar 2002; Hsu et al. 2020; Kuznetsov 2014; McMillan 2008)	✓	✓	✓	✓		✓			✓
	IGO/NGO (Bush and Hadden 2019; Kim 2017; McCleary and Barro 2008; Olter 2021)	✓	✓	✓	✓			✓		
	Civilians (Ben-Yehuda and mishali–ram 2006; Bueno de Mesquita and Smith 2012)	✓	✓	✓	✓		✓			
Pieces	Fatalities (Lacina 2006; Lacina and Gleditsch 2005)	✓				✓	✓	✓	✓	✓
	Force Size (Carafano 2014; Goertz and Diehl 1986; McNabb Cochran and Long 2017)	✓								
	Force Domain (Gartzke and Lindsay 2019; Horowitz 2020; Lanoszka and Hunzeker 2016; Lindsay and Gartzke 2020)	✓	✓	✓	✓					
	Geography (location, territorial change) (Carter 2010)	✓								
Think	Alert (Start/End Crisis) (Lupton 2018)	✓						✓		
	Wishes (Desire/Fear) (Goldgeier and Tetlock 2001; Mercer 2005)	✓						✓		
	Evaluation (Victory/Defeat) (Stein and Russett 1980)	✓						✓		
	Aims (Territory, Policy, Regime, Preemption) (Sullivan 2007)									
	Awareness (Discover, Become Convinced) (Ramsay 2017; Wirtz 2006; Yarhi-Milo 2013)	✓								
Say	React to past event (Praise, Disapprove, Accept, Reject, Accuse) (O'Neill 2018; Risse 2000; Trager 2016)	✓	✓	✓	✓					
	Request future event (Appeal, Demand) (Zartman and Faure 2005)	✓	✓	✓	✓	✓				
	Predict future event (Promise, Threaten, Express Intent, Offer Without Condition) (Davis 2000; Sechser 2011)	✓	✓	✓	✓	✓			✓	
	Predict with condition (Offer, Ultimatum) (R. Powell 2002)	✓								
Do — Unarmed	Government (Leadership/Institution Change, Coup, Assasination) (Goemans, Gleditsch, and Chiozza 2009; Harkness 2016; Jones and Olken 2009; Matanock 2017; J. M. Powell and Thyne 2011)	✓	✓	✓	✓					
	By Civilians (Protest/Riot/Strike) (Chenoweth, Hendrix, and Hunter 2019)	✓	✓	✓	✓					
	Against Civilians (Terrorism, Domestic Rights, Mass Killing, Evacuate) (Eck and Hultman 2007; LaFree and Dugan 2007)	✓	✓	✓	✓					
	Diplomacy (Discussion, Meeting, Mediation, Break off negotiations, Withdraw/Expel Diplomats, Propoganda) (Beardsley 2011)	✓	✓	✓	✓					
	Legal Agreements (Sign Agreement, Settle Dispute, Join War on Behalf of, Ally, Mutual Defense Pact, Open Border, Cede Territory, Allow Inspections, Political Succession, Leave Alliance, Terminate Treaty) (Leeds and Anac 2005; Gibler and Sarkees 2004; Owsiak, Cuttner, and Buck 2018)	✓	✓	✓	✓					
	Violate Agreement (Violate Terms of Agreement) (Leeds 2003/ed)	✓								
	Mutual Cooperation or Directed Aid (Economic cooperation or Aid, Military Cooperation, Intelligence Cooperation, Unspecified) (Leeds 1999)	✓	✓	✓	✓					
	Directed Aid (General Political Support, Economic Aid, Humanitarian Aid, Military Aid, Intelligence Aid, Unspecified Aid) (de Mesquita and Smith 2007; Yarhi-Milo, Lanoszka, and Cooper 2016)	✓	✓	✓	✓					
Do — Armed	Preparation (Alert, Mobilization, Fortify, Exercise, Weapons Test) (Lai 2004)	✓	✓	✓	✓					
	Maneuver (Deployment, Show of Force, Blockade, No Fly Zone, Border Violation) (Allen, Flynn, and Martinez Machain 2021)	✓	✓	✓	✓					
	Combat (Battle/Clash, Attack, Invasion/Occupation, Bombard, Cease Fire, Retreat) (Fortna 2018; Min 2021)	✓	✓	✓	✓					
	Strategic (Declare War, Join War, Continue Fighting, Surrender, End War, Withdraw from War, Switch Sides) (Levy and Thompson 2011; Reiter 2009)	✓	✓	✓	✓					
	Autonomy (Assert Political Control Over, Assert Autonomy Against, Annex, Reduce Control Over, Decolonize) (Frederick, Hensel, and Macaulay 2017; Hensel 1996; Schultz 2015)	✓	✓	✓	✓					

5 ICBe Coding Process

The ICBe ontology follows a hierarchical design philosophy where a smaller number of significant decisions are made early on and then progressively refined into more specific details (Brust and Denzler 2020).[9] Each coder was instructed to first thoroughly read the full crisis narrative and then presented with a custom graphical user interface (SI Appendix, Fig. B1). Coders then proceeded sentence by sentence, choosing the number of events (0-3) that occurred, the highest behavior (thought, speech, or activity), a set of players (P), whether the means were primarily armed or unarmed, whether there was an increase or decrease in aggression (uncooperative/escalating or cooperative/de-escalating), and finally one or more non-mutually exclusive specific activities. Some additional details like location and timing information was always collected while other details were only collected if appropriate, e.g. force size, fatalities, domains, units, etc. A unique feature of the ontology is that thought, speech, and do behaviors can be nested into combinations, e.g. an offer for the U.S.S.R. to remove missiles from Cuba in exchange for the U.S. removing missiles from Turkey. Through compounding, the ontology can capture what players were said to have known, learned, or said about other specific fully described actions.

Each crisis was typically assigned to 2 expert coders and 2 novice coders with an additional tie-breaking expert coder assigned to sentences with high disagreement.[10] For the purposes of measuring intercoder agreement and consensus, we temporarily disaggregate the unit of analysis to the Coder-Crisis-Sentence-Tag (n=993,740), where a tag is any unique piece of information a coder can associate with a sentence such as an actor, date, behavior, etc. We then aggregate those tags into final events (n=18,783), using a consensus procedure (SI Appendix, Algorithm B2) that requires a tag to have been chosen by at least one expert coder and either a majority of expert or novice coders. This screens noisy tags that no expert considered possible but leverages novice knowledge to tie-break between equally plausible tags chosen by experts.

6 Internal Consistency

We evaluate the internal validity of the coding process in several ways. For every tag applied we calculate the observed intercoder agreement as the percent of other coders who also applied that same tag (SI Appendix, Fig. B3). Across all concepts, the Top 1 Tag Agreement was low among novices (31%), moderate for experts (65%), and high (73%) following the consensus screening procedure.

We attribute the remaining disagreement primarily to three sources. First, we required coders to rate their confidence which was observed to be low for 20% of sentences- half due to a mismatch between the ontology and the text ("survey doesn't fit event"-45%) and half due to a lack of information or confused writing in the source text ("more knowledge needed"-40%, "confusing sentence"-6%). Observed disagreement varied predictably with self reported confidence (SI Appendix, Fig. B4). Second, as intended agreement is higher (75-80%) for questions with fewer options near the root of the ontology compared to agreement for questions near the leafs of the ontology (50%-60%). Third, individual coders exhibiting nontrivial coding styles, e.g. some more expressive applying many tags per concept while others focused on only the single best match. We further observed unintended synonymity, e.g. the same information can be framed as either a threat to do something or a promise not to do something.

7 Case Study Selection

The remaining two qualities we seek to measure are recall and precision of ICBe events in absolute terms and relative to other existing systems. We provide full ICB narratives, ICBe coding in an easy to read icongraphic form, and a wide range of visualizations for every case on the companion website. In this paper, we focus on two deep case studies. The first is the Cuban Missile Crisis (Figure 1) which took place primarily in the second half of 1962, involved the United States, the Soviet Union, and Cuba, and is widely known for bringing the world to the brink of nuclear war (hereafter Cuban Missiles). The second is the Crimea-Donbas Crisis (SI Appendix Figure D1) which took place primarily in 2014, involved Russia, Ukraine, and NATO, and within a decade spiraled into a full scale invasion (herafter Crimea-Donbas). Both cases involve a superpower in crisis with a neighbor, initiated by a change from a friendly to hostile regime, with implications for economic and

[9]This process quickly focuses the coder on a smaller number of relevant options while also allowing them to apply multiple tags if the sentence explicitly includes more than one or there is insufficient evidence to choose only one tag. The guided coding process also allows for the possibility that earlier coarse decisions have less error than later fine-grained decisions.

[10]Expert coders were graduate students or postgraduates who collaboratively developed the ontology and documentation for the codebook. Undergraduate coders were students who engaged in classroom workshops.

military security for the superpower, risked full scale invasion, and eventually invited intervention by opposing superpowers. We choose these cases because they are substantively significant to 20th and 21st century international relations, widely known across scientific disciplines and popular culture, and are sufficiently brief to evaluate in depth.

8 Recall

Recall measures the share of desired information recovered by a sequence of coded events, $Pr(E|H)$, and is poorly defined for historical episodes. First, there is no genuine ground truth about what occurred, only surviving texts about it. Second, there is no *a priori* guide to what information is necessary detail and what is ignorable trivia. History suffers from what is known as the Coastline Paradox (Mandelbrot 1983) — it has a fractal dimension greater than one such that the more you zoom in the more detail you will find about individual events and in between every two discrete events. The ICBe ontology is a proposal about what information is important, but we need an independent benchmark to evaluate whether that proposal is a good one and that allows for comparing proposals from event projects that had different goals. We need a yardstick for history.

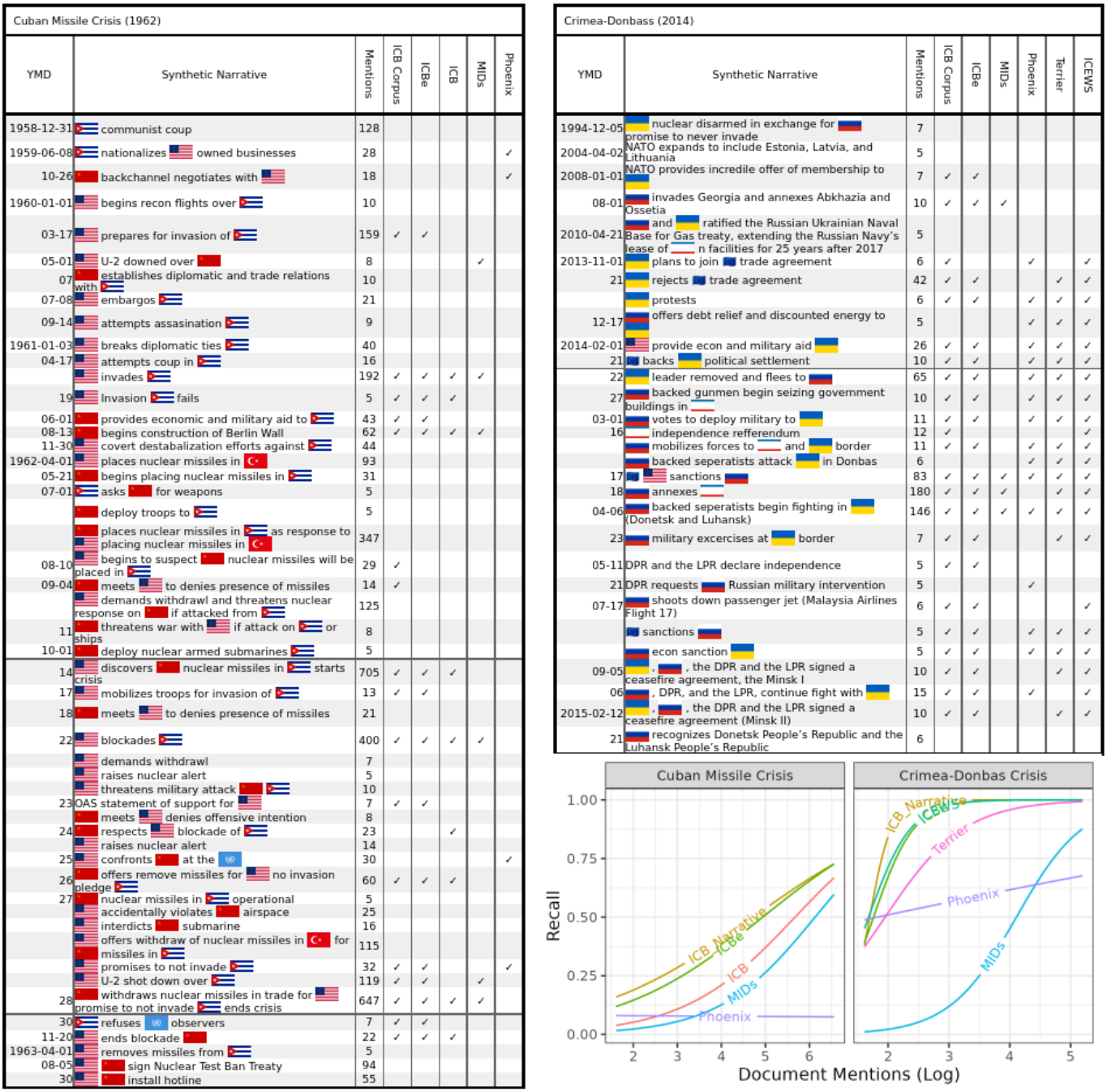
Our strategy for dealing with both problems is a plausibly objective yardstick called a synthetic historical narrative. For both case studies, we collect a large diverse corpus of narratives spanning timelines, encyclopedia entries, journal articles, news reports, websites, and government documents. Using natural language processing (fully described in SI Appendix, Algorithm C1), we identify details that appear across multiple accounts. The more accounts that mention a detail, the more central it is to understanding the true historical episode. The theoretical motivation is that authors face word limits which force them to pick and choose which details to include, and they choose details which serve the specific context of the document they are producing. With a sufficiently large and diverse corpus of documents, we can vary the context while holding the overall episode constant and see which details tend to be invariant to context. Intuitively, a high quality event dataset should have high recall for context invariant details both because of their broader relevance and also because they are easier to find in source material.

Synthetic historical narratives for Cuban Missiles (51 events drawn from 2020 documents) and Crimea-Donbas (30 events drawn from 971 documents) appear in Figure 2. Each row represents a detail which appeared in at least five documents along with an approximate start date, a hand written summary, the number of documents it was mentioned in, and whether it could be identified in the text of our ICB corpus, in our ICBe events, or any of the competing systems.

From them, we draw several stylized facts. First, there is substantial variation in which details any one document will choose to include. Our ground truth ICB narratives included 17/51 and 23/30 of the events from the synthetic narrative, while including other details that are not in the synthetic narrative. Second, mentions of a detail across accounts is exponentially distributed with context invariant details appearing dozens to hundreds of times more than context dependent details. Third, crisis start and stop dates are arbitrary and the historical record points to many precursor events as necessary detail for understanding later events, e.g. the U.S. was in a *de facto* grey scale war with Cuba before it invited Soviet military protection (Cormac and Aldrich 2018) and Ukraine provided several security guarantees to Russia that were potentially undone, e.g. a long term lease on naval facilities in Crimea. Fourth, we find variation between the two cases. Cuban Missiles has a cleaner canonical end with the Soviets agreeing to withdraw missiles while Crimea-Donbas meekly ends with a second cease fire agreement (Minsk II) but continued fighting. The canonical narrative of Cuban Missile also includes high level previously classified details, while the more recent Crimea-Donbas case reflects primarily public reporting.

We find substantive variation in recall across systems. Recall for each increases in the number of document mentions which is an important sign of validity for both them and our benchmark. The one outlier is Phoenix which is so noisy that it's flat to decreasing in mentions. The two episode level datasets have very low coverage of contextual details. The two other dictionary systems ICEWs and Terrier have high coverage, with ICEWs outperforming Terrier. ICBe strictly dominates all of the systems but ICEWs in recall though we note that the small sample sizes mean these systems should be considered statistically indistinguishable. Importantly our corpus of ICB narratives has very high recall of frequently mentioned details giving us confidence in how those summaries were constructed, and ICBe lags only slightly behind showing that it left very little additional information on the table.

Figure 2: Measuring Recall with Synthetic Historical Narratives



Notes: Synthetic narratives combine several thousand accounts of each crisis into a single timeline of events, taking only those mentioned in at least 5 or more documents. Checkmarks represent whether that event could be hand matched to any detail in the ICB corpus, ICBe dataset, or any of the other event datasets.

9 Precision

The other side of event measurement is precision, the degree to which a sequence of events correctly and usefully describes the information in history, $Pr(H|E)$. It does little good to recall a historical event but too vaguely (e.g. MIDs describes the Cuban Missile crisis as a blockade, a show of force, and a stalemate) or with too much error (e.g. ICEWS records 263 "Detonate Nuclear Weapons" events between 1995-2019) to be useful for downstream applications. ICBe's ontology and coding system is designed to strike a balance so that the most important information is recovered accurately but also abstracted to a level that is still useful and interpretable. You should be able to lay out events of a crisis on a timeline, as in Figure 3 and Figure 4, and read off the macro structure of an episode from each individual move. We call this visualization a crisis map, a directed graph intersected with a timeline, and provide crisis maps for every event dataset for each case study (SI Appendix, Fig. D3 and D4) and all crises on the companion website.

We further want to verify individual event codings, which we can do in the case of ICBe because each event is mapped to a specific span of text. We develop the iconography system for presenting event codings as coherent statements that can be compared side by side to the original source narrative as for Cuban Missiles (Figure 1), Crimea-Donbas (SI Appendix Table D1), and for every case on the companion website. We further provide a stratified sample of event codings alongside their source text (SI Appendix Table D2).

We find both the visualizations of macro structure and head-to-head comparisons of ICBe codings to the raw text to strongly support the quality of ICBe, but as with recall we seek a more objective detached universal benchmark. Our proposed measure is a reconstruction task to see whether our intended ontology can be recovered through only unsupervised clustering of sentences they were applied to. Figure 4 shows the location of every sentence from the ICBe corpus in semantic space as embeded using the same large language model as before, and the median location of each ICBe event tag applied to those sentences.[11] Labels reflect the individual leaves of the ontology and colors reflect the higher level coerce branch nodes of the ontology. If ICBe has high precision, substantively similar tags ought to have been applied to substantively similar source text, which is what we see both in two dimensions in the main plot and via hierarchical clustering on all dimensions in the dendrogram along the righthand side.[12]

Finally, how does ICBe's precision compare to the existing state of the art? The crisis-maps reveal the episode level datasets like MIDs or the original ICB are too sparse and vague to reconstruct the structure of the crisis (SI Appendix Figure D3 and D4). On the other end of the spectrum, the high recall dictionary based event datasets like Terrier and ICEWs produce so many noisy events (several hundreds thousands) that even with heavy filtering their crisis maps are completely unintelligible. Further, because of copyright issues, none of these datasets directly provide the original text spans making event level precision difficult to verify.

However, given their high recall on our task and the global and real-time coverage of dictionary based event systems, we want to take seriously the possibility that some functional transformation could recover the precision of ICBe. For example, (Terechshenko 2020) attempts to correct for the mechanically increasing amount of news coverage each year by detrending violent event counts from Phoenix using a human coded baseline. Others have focused on verifying precision for ICEWs on specific subsets of details against known ground truths, e.g. geolocation (Cook and Weidmann 2019), protest events (80%) (Wüest and Lorenzini 2020), anti-government protest networks (46.1%) (Jäger, n.d.).

We take the same approach here in Figure 5, selecting four specific CAMEO event codings and checking how often they reflect a true real world event. We choose four event types around key moments in the crisis. The start of the crisis revolves around Ukraine backing out of trade deal with the EU in favor of Russia, but "sign formal agreement" events act more like a topic detector with dozens of events generated by discussions of a possible agreement but not the actual agreement which never materialized. The switch is caught by the "reject plan, agreement to settle dispute", but also continues for Victor Yanukovych for even after he was removed from power because of articles retroactively discussing the cause of his of his removal. Events for "use conventional military force" capture a threshold around the start of hostilities and who the participants were but not any particular battles or campaigns. Likewise, "impose embargo, boycott, or sanctions" captures the start of waves of sanctions and from who but are effectively constantly as the news coverage does not distinguish between subtle changes or additions. In sum, dictionary based methods on news corpora tend to have high recall because they parse everything in the news, but for the same reason their specificity for most event types is too low to back out individual chess like sequencing that ICBe aims to record.

[11] We preprocess sentences to replace named entities with a generic Entity token.

[12] Hierchcial clustering on cosine similarity and with Ward's method.

Figure 3: Crisis Map: Cuban Missile Crisis

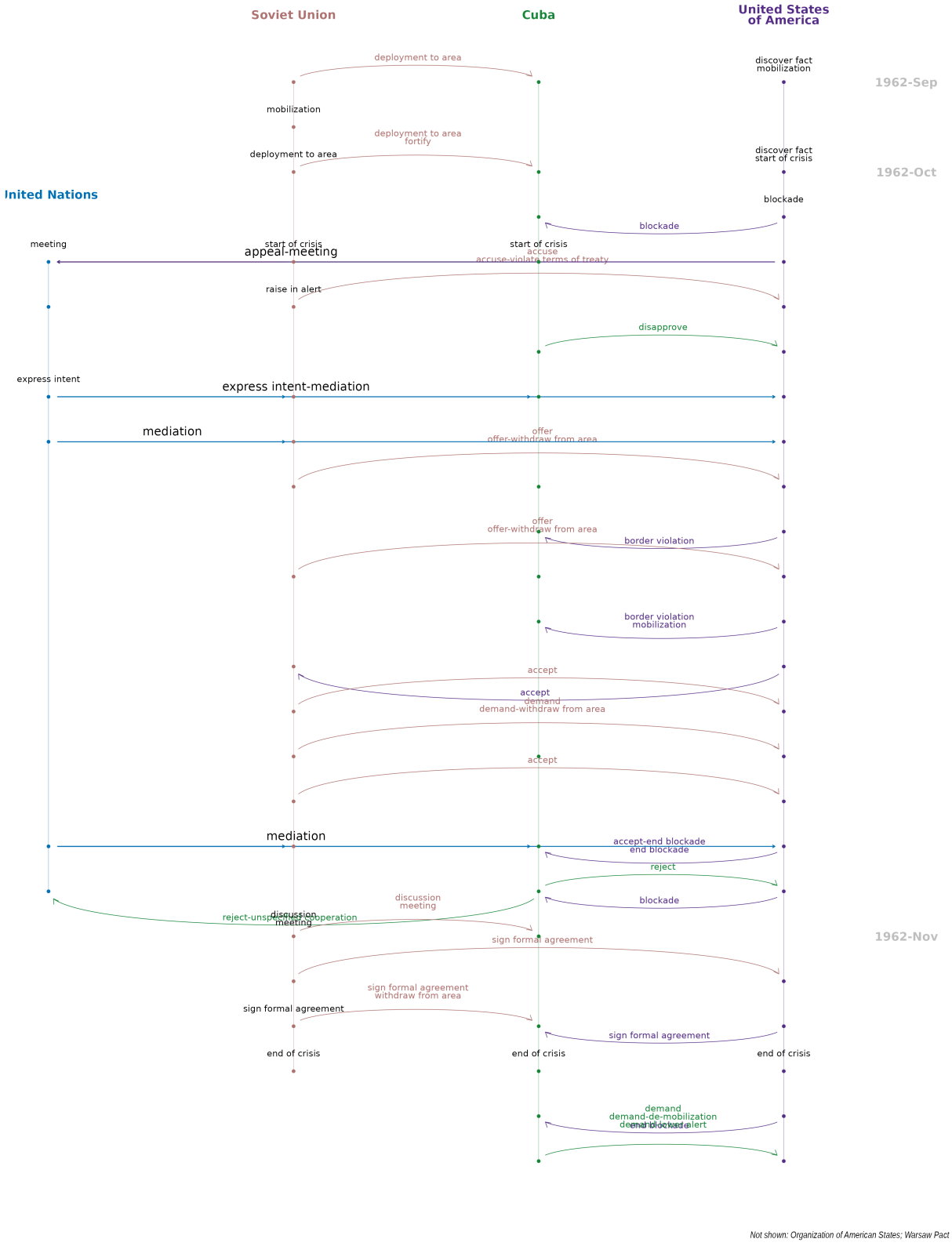
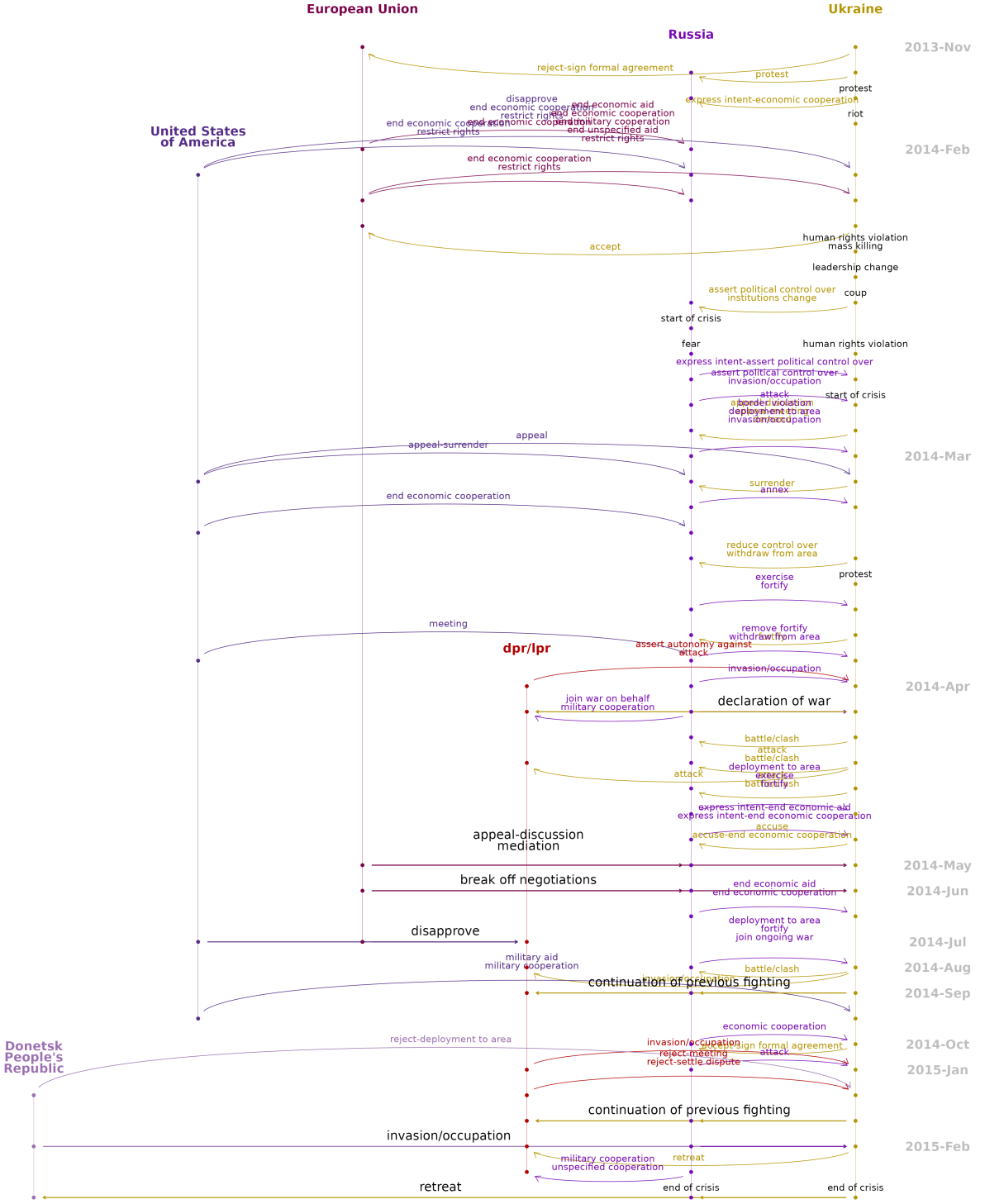
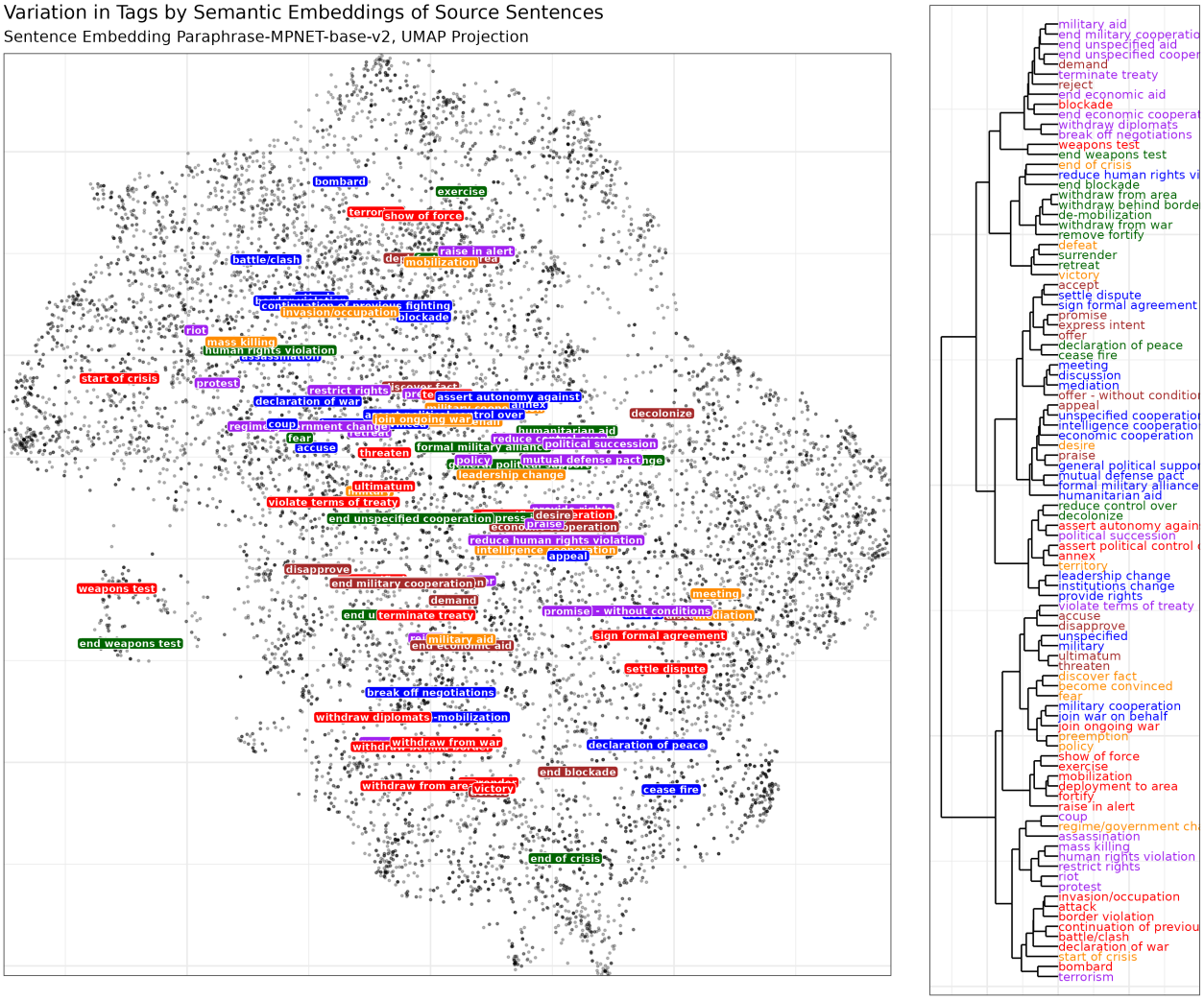


Figure 4: Crisis Map: Crimea-Donbas



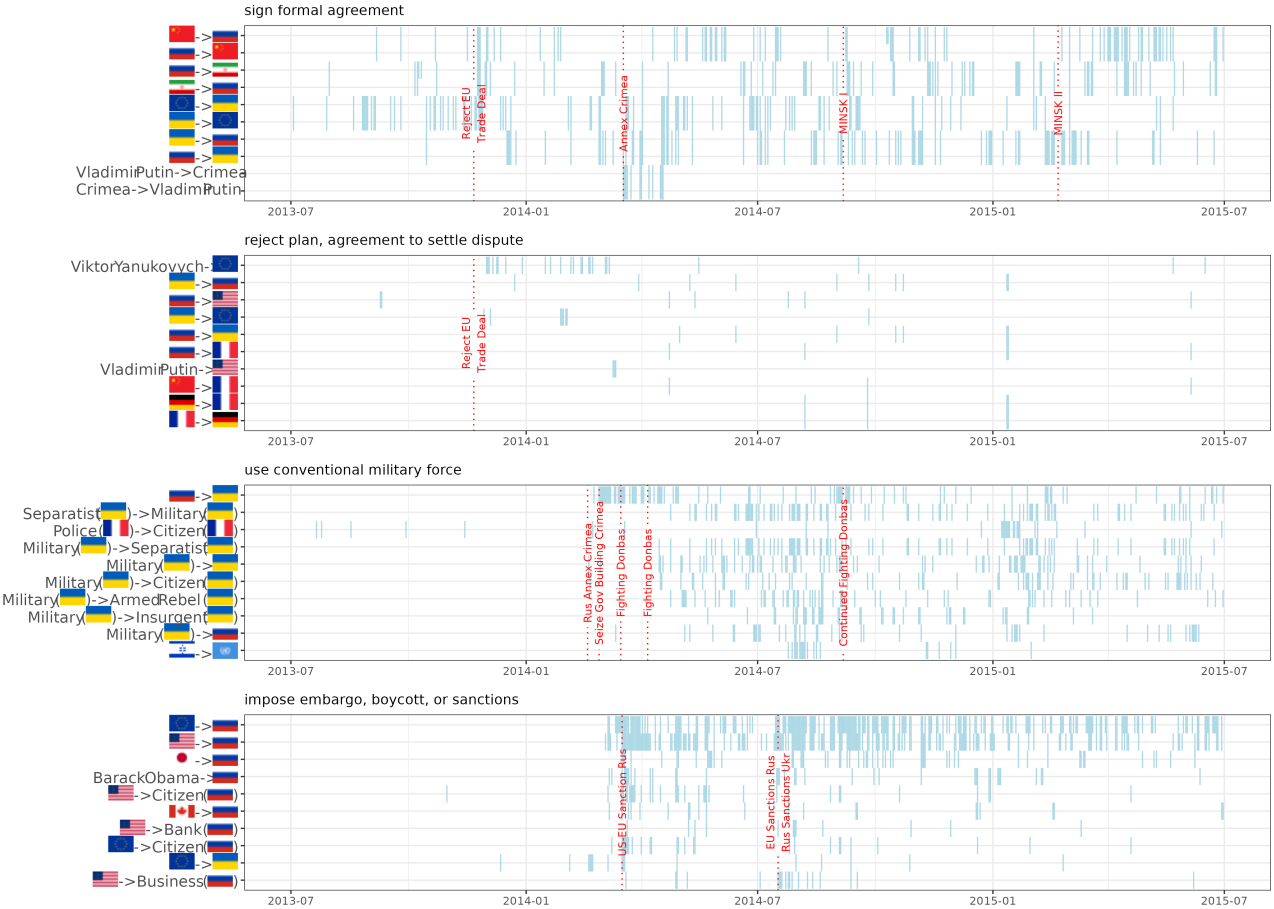
Not shown: France; G8; Germany; Luhansk People's Republic; Malaysia; NATO; Organization for Security and Co-operation in Europe; Trilateral Contact Group on Ukraine; United Nations

Figure 5: ICBe event codings in comparison to Semantic Embeddings from source sentences



Notes: Dots represent individual ICB narrative sentences, as embeded by the Paraphrase-MPNET-base-v2 large language model and flattened into two dimensions with UMAP. Text labels reflect individual leaves of the ICBe ontology, and colors represent intermediate branches of the ontology. Label placement is the median of all of the sentences that tag was applied to by the coders. The dendrogram shows hiearchical clustering of the tags. If ICBe precision is high, the sentences tags were applied to ought to say similar things, and the intended shape of the ontology ought to be visually recognizable.

Figure 6: ICEWs Events by Day by Type during the Crimea-Donbas Crisis



Notes:Unit of analysis is the Dyad-Day. Edges <-> indicates undirected dyad and -> indicates directed dyad. Top 10 most active dyads per category shown. Red text shows events from the synthetic narrative relative to that event category. Blue bars indicate an event recorded by ICEWs for that dyad on that day.

10 Conclusion

We investigated event abstraction from narratives describing key historical episodes in international relations. We synthesized a prior belief about the latent unobserved phenomena that drive these events in international relations and proposed a mapping to observable concepts that enter into the observed historical record. We designed an ontology with high coverage over those concepts and developed a training procedure and technical stack for human coding of historical texts. Multiple validity checks find the resulting codings have high internal validity (e.g. intercoder agreement) and external validity (i.e. matching source material in both micro-details at the sentence level and macro-details spanning full historical episodes). Further, these codings perform much better in terms of recall, precision, coverage, and overall coherence in capturing these historical episodes than existing event systems used in international relations.

We release several open-source products along with supporting code and documentation to further advance the study of IR, event extraction, and natural language processing. The first is the International Crisis Behavior Events (ICBe) dataset, an event-level aggregation of what took place during the crises identified by the ICB project. These data are appropriate for statistical analysis of hard questions about the sequencing of events (e.g. escalation and de-escalation of conflicts). Second, we provide a coder-level disaggregation with multiple codings of each sentence by experts and undergrads that allows for the introduction of uncertainty and human interpretation of events. Further, we release a direct mapping from the codings to the source text at the sentence level as a new resource for natural language processing. Finally, we provide a companion website that incorporates detailed visualizations of all of the data introduced here (www.crisisevents.org).

Funding

This work was supported by a grant from the Office of Naval Research [N00014-19-1-2491] and from the Charles Koch Foundation [20180481].

Acknowledgements

We thank the ICB Project and its directors and contributors for their foundational work and their help with this effort. We make special acknowledgment of Michael Brecher for helping found the ICB project in 1975, creating a resource that continues to spark new insights to this day. We thank the many undergraduate coders for their patience and dedication. Thanks to the Center for Peace and Security Studies and its membership for comments. Special thanks to Rebecca Cordell, Philip Schrodt, Zachary Steinert-Threlkeld, and Zhanna Terechshenko for generous feedback. Thank you to the cPASS research assistants that contributed to this project: Helen Chung, Daman Heer, Syeda ShahBano Ijaz, Anthony Limon, Erin Ling, Ari Michelson, Prithviraj Pahwa, Gianna Pedro, Tobias Stodiek, Yiyi 'Effie' Sun, Erin Werner, Lisa Yen, and Ruixuan Zhang. This project was supported by a grant from the Office of Naval Research [N00014-19-1-2491] and benefited from the Charles Koch Foundation's support for the Center for Peace and Security Studies.

Author Contributions

Conceptualization: R.W.D., E.G., J.L.; Methodology: R.W.D., T.L.S.; Software: R.W.D.; Validation: R.W.D., T.L.S.; Formal Analysis: R.W.D., T.L.S.; Investigation: S.C., R.W.D., J.A.G., C.K., N.L., E.M., J.M.C.N., D.P., D.Q., J.W.; Data Curation: R.W.D., D.Q., T.L.S., J.W.; Writing - Original Draft: R.W.D., T.L.S.; Writing - Review & Editing: R.W.D., J.A.G., E.G., T.L.S.; Visualization: R.W.D., T.L.S.; Supervision: E.G.; Project Administration: S.C., R.W.D., J.A.G., D.Q., T.L.S., J.W.; Funding Acquisition: E.G., J.L.

Data Availability Statement

This article's data, online appendix, replication material, and visualizations of every historical episode are available on the github repository ICBEventData and through the companion website www.crisisevents.org.

Conflicts of Interests

There are no conflicts of interest to disclose.

Allison, Graham T., and Philip Zelikow. 1971. *Essence of Decision: Explaining the Cuban Missile Crisis*. Vol. 327. 729.1. Little, Brown Boston.

Althaus, Scott, Joseph Bajjalieh, John F. Carter, Buddy Peyton, and Dan A. Shalmon. 2019. "Cline Center Historical Phoenix Event Data Variable Descriptions." *Cline Center Historical Phoenix Event Data*.

Balali, Ali, Masoud Asadpour, and Seyed Hossein Jafari. 2021. "COfEE: A Comprehensive Ontology for Event Extraction from Text." arXiv. https://doi.org/10.48550/arXiv.2107.10326.

Barari, Soubhik, and In Song Kim. n.d. "Democracy and Trade Policy at the Product Level: Evidence from a New Tariff-line Dataset," 16.

Beardsley, Kyle, Patrick James, Jonathan Wilkenfeld, and Michael Brecher. 2020. "The International Crisis Behavior Project." Oxford Research Encyclopedia of Politics. September 28, 2020. https://doi.org/10.1093/acrefore/9780190228637.013.1638.

Bloomfield, Lincoln P., and Allen Moulton. 1989. "CASCON III: Computer-aided System for Analysis of Local Conflicts." *MIT Center for International Studies, Cambridge*.

Boschee, Elizabeth, Jennifer Lautenschlager, Sean O'Brien, Steve Shellman, James Starz, and Michael Ward. 2015. "ICEWS Coded Event Data." *Harvard Dataverse* 12.

Braithwaite, Alex. 2009. "Codebook for the Militarized Interstate Dispute Location (MIDLOC) Data, v 1.0." *University College London*.

———. 2010. "MIDLOC: Introducing the Militarized Interstate Dispute Location Dataset." *Journal of Peace Research* 47 (1): 91–98. https://doi.org/10.1177/0022343309350008.

Brandt, Patrick T., Vito DOrazio, Jennifer Holmes, Latifur Khan, and Vincent Ng. 2018. "Phoenix Real-Time Event Data." http://eventdata.utdallas.edu.

Brecher, Michael. 1999. "International Studies in the Twentieth Century and Beyond: Flawed Dichotomies, Synthesis, Cumulation: ISA Presidential Address." *International Studies Quarterly* 43 (2): 213–64.

Brecher, Michael, and Patrick James. 1988. "Patterns of Crisis Management." *Journal of Conflict Resolution* 32 (3): 426–56.

Brecher, Michael, Patrick James, and Jonathan Wilkenfeld. 2000. "Crisis Escalation to War: Findings from the International Crisis Behavior Project." *What Do We Know About War*.

Brecher, Michael, and Jonathan Wilkenfeld. 1982. "Crises in World Politics." *World Politics* 34 (3): 380–417. https://doi.org/10.2307/2010324.

———. 1997. *A Study of Crisis*. University of Michigan Press.

Brecher, Michael, Jonathan Wilkenfeld, Kyle C. Beardsley, Patrick James, and David Quinn. 2017. "International Crisis Behavior Data Codebook." Codebook Version 12. http://sites.duke.edu/icbdata/data-collections/.

Brecher, Michael, Jonathan Wilkenfeld, Kyle Beardsley, Patrick James, and David Quinn. n.d. "International Crisis Behavior Data Codebook, Version 12," 69.

Brust, Clemens-Alexander, and Joachim Denzler. 2020. "Integrating Domain Knowledge: Using Hierarchies to Improve Deep Classifiers." http://arxiv.org/abs/1811.07125.

Cook, Scott J., and Nils B. Weidmann. 2019. "Lost in Aggregation: Improving Event Analysis with Report-Level Data." *American Journal of Political Science* 63 (1): 250–64. https://doi.org/10.1111/ajps.12398.

Cormac, Rory, and Richard J. Aldrich. 2018. "Grey Is the New Black: Covert Action and Implausible Deniability." *International Affairs* 94 (3): 477–94. https://doi.org/10.1093/ia/iiy067.

Felbermayr, Gabriel, Aleksandra Kirilakha, Constantinos Syropoulos, Erdal Yalcin, and Yoto V. Yotov. 2020. "The Global Sanctions Data Base." *European Economic Review* 129 (October): 103561. https://doi.org/10.1016/j.euroecorev.2020.103561.

Frederick, Bryan A, Paul R Hensel, and Christopher Macaulay. 2017. "The Issue Correlates of War Territorial Claims Data, 1816–20011." *Journal of Peace Research* 54 (1): 99–108. https://doi.org/10.1177/0022343316676311.

Gaddis, John Lewis. 1987. "Expanding the Data Base: Historians, Political Scientists, and the Enrichment of Security Studies." *International Security* 12 (1): 3–21. https://doi.org/10.2307/2538915.

Gavin, Francis J. 2014. "History, Security Studies, and the July Crisis." *Journal of Strategic Studies* 37 (2): 319–31. https://doi.org/10.1080/01402390.2014.912916.

George, Alexander L., and Richard Smoke. 1974. *Deterrence in American Foreign Policy: Theory and Practice*. Columbia University Press.

Gibler, Douglas M. 2018. *International Conflicts, 1816-2010: Militarized Interstate Dispute Narratives*. Rowman & Littlefield. https://books.google.com?id=_4VTDwAAQBAJ.

Grant, Christan, Andrew Halterman, Jill Irvine, Yan Liang, and Khaled Jabr. 2017. "OU Event Data Project," December. https://osf.io/4m2u7/.

Halterman, Andy. 2020. "Extracting Political Events from Text Using Syntax and Semantics."

Hegre, Håvard, Mihai Croicu, Kristine Eck, and Stina Högbladh. 2020. "Introducing the UCDP Candidate Events Dataset." *Research & Politics* 7 (3): 2053168020935257.

Hermann, Charles. 1984. "Comparative Research on the Events of Nations (CREON) Project: Foreign Policy Events, 1959-1968: Version 1." ICPSR - Interuniversity Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR05205.V1.

Hewitt, J. Joseph. 2001. "Engaging International Data in the Classroom: Using the ICB Interactive Data Library to Teach Conflict and Crisis Analysis." *International Studies Perspectives* 2 (4): 371–83. https://doi.org/10.1111/1528-3577.00066.

Holsti, Ole R. 1965. "The 1914 Case." *The American Political Science Review* 59 (2): 365–78. https://doi.org/10.2307/1953055.

Iakhnis, Evgeniia, and Patrick James. 2019. "Near Crises in World Politics: A New Dataset." *Conflict Management and Peace Science*, July, 0738894219855610. https://doi.org/10.1177/0738894219855610.

Jäger, Kai. n.d. "The Limits of Studying Networks with Event Data: Evidence from the ICEWS Dataset."

James, Patrick. 2019. "What Do We Know about Crisis, Escalation and War? A Visual Assessment of the International Crisis Behavior Project." *Conflict Management and Peace Science* 36 (1): 3–19. https://doi.org/10.1177/0738894218793135.

Kang, David C., and Alex Yu-Ting Lin. 2019. "US Bias in the Study of Asian Security: Using Europe to Study Asia." *Journal of Global Security Studies* 4 (3): 393–401.

Kinne, Brandon J. 2020. "The Defense Cooperation Agreement Dataset (DCAD)." *Journal of Conflict Resolution* 64 (4): 729–55. https://doi.org/10.1177/0022002719857796.

Leng, Russell J., and J. David Singer. 1988. "Militarized Interstate Crises: The BCOW Typology and Its Applications." *International Studies Quarterly* 32 (2): 155–73. https://doi.org/10.2307/2600625.

Li, Qian, Hao Peng, Jianxin Li, Yiming Hei, Rui Sun, Jiawei Sheng, Shu Guo, et al. 2021. "A Comprehensive Survey on Schema-based Event Extraction with Deep Learning." http://arxiv.org/abs/2107.02126.

Mandelbrot, Benoit B. 1983. *The fractal geometry of nature*. New York: Freeman.

Maoz, Zeev, Paul L. Johnson, Jasper Kaplan, Fiona Ogunkoya, and Aaron P. Shreve. 2019. "The Dyadic Militarized Interstate Disputes (MIDs) Dataset Version 3.0: Logic, Characteristics, and Comparisons to Alternative Datasets." *Journal of Conflict Resolution* 63 (3): 811–35. https://doi.org/10.1177/0022002718784158.

McClelland, Charles. 1978. "World Event/Interaction Survey, 1966-1978." *WEIS Codebook ICPSR* 5211.

Merritt, Richard L. 1994. "Measuring Events for International Political Analysis." *International Interactions* 20 (1-2): 3–33.

Miller, George A. 1995. "WordNet: A Lexical Database for English." *Communications of the ACM* 38 (11): 39–41. https://doi.org/10.1145/219717.219748.

Moyer, Jonathan D, Sara D Turner, and Collin J Meisel. 2020. "What Are the Drivers of Diplomacy? Introducing and Testing New Annual Dyadic Data Measuring Diplomatic Exchange." *Journal of Peace Research*, September, 0022343320929740. https://doi.org/10.1177/0022343320929740.

Owsiak, Andrew P, Allison K Cuttner, and Brent Buck. 2018. "The International Border Agreements Dataset." *Conflict Management and Peace Science* 35 (5): 559–76. https://doi.org/10.1177/0738894216646978.

Paige, Glenn D. 1968. *The Korean Decision, June 24-30, 1950*. Free Press.

Palmer, Glenn, Roseanne W McManus, Vito D'Orazio, Michael R Kenwick, Mikaela Karstens, Chase Bloch, Nick Dietrich, Kayla Kahn, Kellan Ritter, and Michael J Soules. 2021. "The Mid5 Dataset, 2011–2014: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science*, February, 0738894221995743. https://doi.org/10.1177/0738894221995743.

Pettersson, Therése, and Kristine Eck. 2018. "Organized Violence, 1989–2017." *Journal of Peace Research* 55 (4): 535–47. https://doi.org/10.1177/0022343318784101.

Raleigh, Clionadh, Andrew Linke, Håavard Hegre, and Joakim Karlsen. 2010. "Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature." *Journal of Peace Research* 47 (5): 651–60.

Ralph Sundberg, and Mihai Croicu. 2016. "UCDP GED Codebook Version 5.0." Department of Peace and Conflict Research, Uppsala University.

Reiter, Dan. 2015. "Should We Leave Behind the Subfield of International Relations?" *Annual Review of Political Science* 18 (1): 481–99. https://doi.org/10.1146/annurev-polisci-053013-041156.

Reiter, Dan, Allan C. Stam, and Michael C. Horowitz. 2016. "A Revised Look at Interstate Wars, 1816–2007." *Journal of Conflict Resolution* 60 (5): 956–76. https://doi.org/10.1177/0022002714553107.

Sarkees, Meredith Reid, and Frank Wayman. 2010. *Resort to War: 1816-2007*. CQ Press.

Schrodt, Philip A., and Blake Hall. 2006. "Twenty Years of the Kansas Event Data System Project." *The Political Methodologist* 14 (1): 2–8.

Sechser, Todd S. 2011. "Militarized Compellent Threats, 1918–2001." *Conflict Management and Peace Science* 28 (4): 377–401. https://doi.org/10.1177/0738894211413066.

Sherman, Frank L. 2000. "SHERFACS: A Cross-Paradigm, Hierarchical, and Contextually-Sensitive International Conflict Dataset, 1937-1985: Version 1." ICPSR - Interuniversity Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR02292.V1.

Snyder, Glenn Herald, and Paul Diesing. 1977. *Conflict Among Nations: Bargaining and Decision Making in International Crises*. Princeton University Press.

Steinert-Threlkeld, Zachary C. 2019. "The Future of Event Data Is Images." *Sociological Methodology* 49 (1): 68–75. https://doi.org/10.1177/0081175019860238.

Sundberg, Ralph, and Erik Melander. 2013. "Introducing the UCDP Georeferenced Event Dataset." *Journal of Peace Research* 50 (4): 523–32.

Terechshenko, Zhanna. 2020. "Hot Under the Collar: A Latent Measure of Interstate Hostility." *Journal of Peace Research* 57 (6): 764–76. https://doi.org/10.1177/0022343320962546.

Vabulas, Felicity, and Duncan Snidal. 2021. "Cooperation Under Autonomy: Building and Analyzing the Informal Intergovernmental Organizations 2.0 Dataset." *Journal of Peace Research* 58 (4): 859–69. https://doi.org/10.1177/0022343320943920.

Wilkenfeld, Jonathan, and Michael Brecher. 2000. "Interstate Crises and Violence: Twentieth-Century Findings." *Handbook of War Studies II*, 282–300.

Wüest, Bruno, and Jasmine Lorenzini. 2020. "External Validation of Protest Event Analysis." *Contention in Times of Crisis: Recession and Political Protest in Thirty Euro-Pean Countries*, 49–78.

Zhang, Han, and Jennifer Pan. 2019. "CASM: A Deep-Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media." *Sociological Methodology* 49 (1): 1–57. https://doi.org/10.1177/0081175019860244.