

A Wiki-based Dataset of Military Operations with Novel Strategic Technologies (MONSTr)

J Andrés Gannon* and Kerry Chávez†

April 18, 2023

Abstract

Research on strategies and force employment in modern warfare is prolific, but siloed. While some examine boots on the ground, others focus on aerial bombing or unpiloted platforms. Consequently, most studies focus on the effects of one approach, seldom considering it in lieu of or conjunction with others. Furthermore, there is less knowledge on the origins and implementations of these strategic choices analyzed in isolation. The primary reason for these gaps lies with data limitations. This paper introduces a comprehensive dataset on the universe of United States military operations from 1989-2021 from a single source: Wikipedia. Using automated extraction techniques on its two structured knowledge databases—Wikidata and DBpedia—we uncover information about individual operations within nearly every post-1989 military intervention described in existing academic datasets. The data we introduce offers unprecedented coverage and granularity that enables analysis of myriad factors associated with when, where, and how the United States employs military force. We describe the data collection process, demonstrate its contents and validity, and discuss its potential applications to existing theories about force employment and strategy in war.

Keywords: military intervention, use of force, dataset

*Stanton Nuclear Security Fellow, Council on Foreign Relations, jgannon@cfr.org

†Texas Tech University, Department of Political Science, kerry.chavez@ttu.edu

For feedback on earlier drafts, the authors thank Christa Almonte, Jonathan Caverley, Rex Douglass, Erik Gartzke, Nadiya Kostyuk, Kendrick Kuo, Ashley Leeds, Erik Lin-Greenberg, Timothy MacDonald, Sara Plana, Jeff Randall, Tom Scherer, Ryan Shandler, Erin Staine-Pyne, Alison Thompson, and Sanne Verschuren. Excellent research assistance was provided by Thomas Brailey, Zoe Coutlakis, Allison Lilley, Amanda Madany, Christie Marquez, David McCrum, Jordan Merkel, Matthew Miltimore, Dominick Nguyen, Peyton Olszowka, Cole Reynolds, Caitlen Rodriguez, Yiyi Sun, Qitao Wu, and Lisa Yen. This research was sponsored by Office of Naval Research Grant N00014-14-1-0071, the APSA Centennial Center, and the UCSD Undergraduate Research Apprenticeship Program (URAP). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the view of the Office of Naval Research. The dataset is available at <http://military-operations.com/> and replication code at <https://github.com/jandresgannon/MONSTr>.

Despite being critical of the War on Terror, United States (US) President Obama amplified it upon assuming office and personally ensconced himself in the drone “kill list” approval chain. Then-Director of National Intelligence Blair explained this pivot: “It is the politically advantageous thing to do—low cost, no US casualties, gives the appearance of toughness” (Becker and Shane 2012). Critics argue that these benefits are outweighed by strategic costs such as backlash, radicalization, and increased terrorist recruitment (Kilcullen and Exum 2009). Obama’s decision was not, however, whether or not to use drones. It was whether to use drones rather than insert or maintain boots on the ground in the ongoing wars in Iraq and Afghanistan that he campaigned on ending. It was whether to use drones rather than piloted flights outside of active war theaters.

Current research on how states use military force either focuses on the most aggregate level of the war or explains a particular tool in isolation. Consequently, variation in how states design and deploy force structures over a multi-year campaign remains unexplained. So does the relationship between complementary or substitutable types of force. As a result, academic understanding of how states fight is limited, being highly generalized on one side and siloed into individual domains or platforms on the other. There is a sprawling gap between these, where the actual warfighting occurs. Scholars interested in the prosecution of conflict contend with a dearth of data on the means of force across a comprehensive list of military operations. This paper debuts our attempt to overcome it.

We seek to remedy these shortcomings by providing a dataset that captures the means of force across a more expansive universe of US military events using a familiar, yet under-utilized data source: Wikipedia. Existing datasets focus on higher levels of aggregation, namely the broad military intervention, in part because collecting information at a more granular level is intensive and oft incomplete.¹ Aware of the stigma attached to Wikipedia,

¹For instance, Allen and Martinez Machain (2019, 549) remark that “[i]deally, the unit of analysis for this project would be at the campaign level. Unfortunately, while most modern Western militaries operate in terms of campaigns, data for all countries over time was [sic] not available at the campaign level”.

we nonetheless note that what we gain is worth the tradeoffs. By scraping the Wikiverse for pages on post-1989 US military operations that occurred under the auspices of an intervention, our dataset identifies over 300 observations that existing datasets aggregate into broader wars. Using Wikidata and DBpedia, the structured bedrocks of Wikipedia, we also extract a host of fine-grained covariates concerning each military operation’s dates, belligerents, location, and means used. Since these variables are less subject to human interpretation than those like conflict outcome or objective, concerns about data quality from a source generally considered taboo are mitigated (though still addressed). We hope to convince that such a Wiki-based dataset on Military Operations with Novel Strategic Technologies (MONSTr) is not such a monster after all.

The paper proceeds with a section on our motivation, in which we trace the mismatch between our research interests and existing theory and data. Next, it outlines our approach in collecting, validating, and organizing the data gleaned from our Wikipedia extraction endeavor. Following this, we describe the dataset’s distinctive features to convey the value added, suggesting how scholars might use them to extend or improve studies on how combatants fight. Specifically, we detail its increased granularity and depth, measures of the means of force for each operation, geographic precision, and thoroughness in identifying the roster of combatants on all sides. The penultimate section presents a data demonstration, predicting the means of military force with theoretically derived determinants drawn from existing scholarship. We conclude with a summary discussion, admission of limitations, and suggestions for future research using the MONSTr dataset.

Motivation

A central question in international security concerns how strong states fight, and why. Sometimes they choose losing strategies against weak actors ([Arreguín-Toft 2001](#)). Military my-

opia arguments ([Gentry 2002](#); [Lyall and Wilson 2009](#)), capital-rich force structure narratives ([Gartzke 2001](#); [Caverley 2009](#)), and uncertainty in pre-war estimates of costs and success ([Sullivan and Koch 2009](#)) explain many of these cases. Yet sometimes, despite bureaucratic inertia, economic factor abundance, and the fog of war, advanced militaries deploy ideal force structures against innovative weak enemies. Empirical tests of these theories are limited, however, by inadequate data required for validation.

We identified three barriers to testing theoretical models of how states fight with existing data. First, we are interested in the planning and execution of military strategy *during* armed conflict ([Wallace 2008](#)). Most of the research and data has focused on its beginning—bargaining, causes, onsets—and end—duration, outcomes, aftermaths. Notwithstanding the profound knowledge they have yielded, these approaches black box the prosecution of war. This is surprising since we think its conduct has much to tell us about its causes and consequences. Some studies breach the black box of military force to analyze specific technological approaches within, but typically focus on the possession of platforms rather than their use, offering insight only on the latent potential for force ([Gannon 2021](#)).² Furthermore, the majority bore a peephole onto a single platform or force structure rather than pop open the box. Some look at nuclear platforms ([Gartzke, Kaplow and Mehta 2014](#)), naval power ([Crisher and Souva 2014](#)), aerial bombing ([Pape 1996](#)), air power more broadly ([Horowitz and Reiter 2001](#); [Martinez Machain 2015](#); [Allen and Martinez Machain 2019](#)), ballistic missiles ([Reiter and Mettler 2013](#)), cruise missiles ([Early et al. 2022](#)), or drones ([Fuhrmann and Horowitz 2017](#)).

This leads to the second barrier—siloed scholarship that inhibits the idea of substitutability ([Morgan and Palmer 2000](#)). At a high level, the substantial foreign policy literature reflects this standpoint by depicting a state’s toolkit in broad categories—military, diplomatic, economic, cultural, etc. That same logic applies within each category. Leaders craft the use

²Recent exceptions that focus on use include [Martinez Machain \(2015\)](#), [Post \(2019\)](#), and [Gannon \(2022\)](#).

of military force not as isolated dichotomous decisions, but from a menu of substitutable (or complementary) options. Being domain- or platform-specific, existing datasets on the means of force are difficult to harmonize. Rather than peepholes into the blackbox of military force, our research questions on force employment and strategy require it to be open and well-lit. Once again, our inquiries on how and why states fight fall between existing theory and data, generalized on one side, specific but partitioned on the other.

Finally, holding the universe of military actions and the idea of substitutability in mind, a concern about the relationships across observations crops up. If we seek to explain why an operation featured solely unpiloted systems, it is relevant if it is part of a decapitation campaign in enemy territory. Or if a sizable contingent of troops is deployed for a manpower surge in Iraq, we naturally expect operations nested within the surge campaign to feature a higher frequency of ground troops. While scholars recognize this theoretically, existing datasets rarely identify the bureaucratic and strategic relationships of wars, campaigns, and operations. Testing explanations of force employment choices outside the strategic contexts in which they occur might lead to bias in estimating independent causal impacts.

Our Approach

To surmount these barriers, we produced a dataset that features 1) measures of the means of military force across 2) a comprehensive and disaggregated list of US military operations that 3) captures dependence between observations using a consistent unit of analysis. Given the labor-intensive nature of manual data collection, we turned to the structured knowledge graphs of a familiar, underutilized source of information on political events: Wikipedia. There is virtually no scholar who promotes Wikipedia citations, but there is also virtually no scholar who does not queue up a Wiki page for initial information about a topic. While expert knowledge is required to measure contested or complex concepts like conflict outcome

or purpose, it is less necessary to identify objective datapoints like the location, participants, and dates of an event. When collected carefully, the dimensionality and depth of such information easily extracted through automated processes is worth the taboo trade-off.

We disciplined our collection efforts definitionally, in space, and in time to curate a dataset on US military operations from 1989 to January 2021. The unit of analysis is the military operation. There is consensus among academics, practitioners, and military personnel that there are three general levels of conflict: strategic, operational, and tactical (von Clausewitz 1976; Millett, Murray and Watman 1986; Zabecki 2006).³ While scholars study all three, most research and datasets address the strategic level where national policy and theater planning dwells (McMaster 1998; Millett and Murray 1988). Meanwhile, most military studies focus on the tactical level where soldiers make decisions, execute tasks, and learn (Harkness and Hunzeker 2015; Kollars 2015).

Since we are interested in phenomena between, our dataset dwells at the operational level that “links the tactical employment of forces to national strategic objectives...using operational art to determine how, when, where, and for what purpose military forces will be employed” (Chairman of the Joint Chiefs of Staff 2018). Luttwak (1981) asserts that “the operational level of war, as opposed to the tactical and strategic levels, is or ought to be of greatest concern to the analyst” (61). Several scholars agree that this level is more appropriate for studying military strategy and force employment (Allen and Martinez Machain 2019; Tecott Metz and Halterman 2021). While it includes both operations and campaigns (series of related operations), MONSTr is organized according to the former since military doctrine identifies operations as the primary building block of this level. We follow the Department of Defense (DoD) dictionary which defines the operation as “a sequence of tactical actions

³These levels are generalized for conceptual clarity, but boundaries between them can be blurry in reality. In defining the levels of warfare, US military Joint Publication 3-0 acknowledges that “there are no fixed limits or boundaries between these levels, but they help commanders visualize a logical arrangement of operations, allocate resources, and assign tasks to appropriate commands” (Chairman of the Joint Chiefs of Staff 2018, I-12).

with a common purpose or unifying theme” (Chairman of the Joint Chiefs of Staff 2021, GL-13).

Most existing datasets use the military intervention as the unit of analysis. To properly scope operations that fall within military interventions, we lean on the definitions in the exemplary Military Intervention by Powerful States (MIPS) and International Military Intervention (IMI) datasets. Like MIPS, we exclude threats, training exercises or mobilizations that demonstrate rather than apply force, ebbs and flows of personnel and material at permanent military bases, diplomatic cover and evacuation, and humanitarian operations (Sullivan and Koch 2009). IMI clarifies that events must be purposeful and political (Pearson and Baumann 1988; Pickering and Kisangani 2009). Thus, we omit accidental and unintentional clashes. For instance, if troops deployed along the Iraq border have an unexpected encounter with Syrians, we view it as simply that—a military operation in Iraq that entailed a contingent clash with Syrians who meddled at the border. Being outside the politically determined mission scope, such a clash (or airspace violation) does not constitute a separate observation. We generated our original list of Wiki events to scrape from interventions identified in these reputable datasets. Table A2 in the Appendix summarizes our nearness to MIPS and IMI across definitional fine points of military intervention, and our departures from the broader approaches of RAND (Kavanagh et al. 2019) and Military Intervention Project (MIP) (Kushi and Toft 2023).

Combining the gold standard DoD definition of a military operation with the scoping conditions of a military intervention from MIPS and IMI, an observation in our dataset is:

A series of tactical actions (battles, engagements, strikes) conducted by combat forces in an operational theater to achieve strategic or campaign objectives in the context of a political issue or dispute through action against a foreign adversary. Routine military movements and operations without a defined target like military training exercises, the routine forward deployment of military troops, non-combatant evacuation operations, and disaster relief are excluded.

With our focus on military operations, our dataset does not include drone strikes conducted by the Central Intelligence Agency (CIA), often in noncombat zones, even though they might be conceptualized as part of a military intervention. Falling outside the purview of the Pentagon, we contend that CIA operations abide by different constraints and covery. In fact, extensive criticism has been levied at the CIA program for its lack of military control and transparency.

Considering covery, by nature of being collectible from the internet our cases are open source. We find this coverage appealing in that the events it includes have met some threshold of public awareness to make it onto Wikipedia. Since we expect that public attitudes amid the domestic politics of war are meaningful factors in how states fight, we accept this bias. There are interesting questions about the deployment of covert force, but we appreciate former Department of Intelligence Director Wilson’s comment that “90% of intelligence comes from open sources. The other 10%, the clandestine work, is just the more dramatic. The real intelligence hero is Sherlock Holmes, not James Bond” ([Lands 2019](#)). It’s elementary—we leverage the data that are available and focus on research puzzles that interface with public attention and scrutiny.

Next, we spatially confined our sample to cases featuring the United States. Insofar as our impetus to collect these data is to examine how strong states fight, the US is a paragon. Its prolific intervention record also allows us to maximize observations in a pilot probe with one case. The US case is also deeply studied and data rich, better allowing us to test our theories and compare findings. Finally, we temporally confined our scope to start in 1989, when the US first used cruise missiles in conflict. With a special interest in the politics of emerging technologies, we aimed to collect a sample with temporal consistency in the military options, from ground troops to unmanned platforms, available to conflict decision-makers.

Of course, many of these scope limitations can and should be relaxed. Theoretical debates about how states fight would benefit from empirical information concerning pre-1989

interventions, non-US interventions, and a more expansive definition of military force (humanitarian aid, diplomatic evacuation, etc.). The dataset presented here is a proof of concept for the validity and value of using Wikipedia to identify and structure such information using comparatively feasible conditions. Future iterations of the project will expand in precisely these directions. In the following section, we describe the data collection process, with an emphasis on justifying our source, before demonstrating the data’s yields and potential.

Dataset Inputs

Collection - Scraping and Patching

For most readers, the Wiki interface is familiar. An executive summary sits under the page name, followed by a more detailed description of the entity, event, or concept. Most pages also contain an infobox containing a summary of attributes. For example, the “cosmic latte” page informs that this is the average hue of the universe based on spectral analysis of over 200,000 galaxies. The sections describe the discovery and naming of the color.⁴ These data, like most other sources available to scholars, are unstructured and require manual review and entry.

Unlike most other natural language sources, Wikipedia rests upon two Resource Description Format (RDF) databases: Wikidata and DBpedia ([Malyshev et al. 2018](#)). Wikidata contains an ontology for over 99 million items and their properties as “triples.” Each unique numeric identifier features a structured 1) subject, 2) predicate, and 3) object. For example, Germany (subject) has the capital (predicate) of Berlin (object). DBpedia maps information from each page’s infobox also as structured content that can be semantically queried. Since both databases exist in a uniform, machine-readable format, we were able to scrape a comprehensive list of Wikipedia pages describing a US military operation from 1989 onward

⁴Thank goodness “primordial clam chowder” didn’t win.

using SPARQL, an RDF query language. The resulting data contained nearly 500 entries and dozens of columns containing structured information from Wikidata and DBpedia infoboxes. After reasonable efforts to clean and process the queries as well as locate missing data from alternative sources, the neonate MONSTr had few holes.

Cleaning - Auditing and Validating

The primary contribution of our research method also constitutes its greatest burden: justifying Wikipedia itself. It carries a strong stigma in scholarly research that will be difficult to shake ([Becker 2015](#)).⁵ We submit that for certain types of information it is a useful source. Our efforts to audit and validate the data took two forms: 1) an examination of scholarly research on Wikipedia’s own data collection process and validation and 2) our own targeted and random audits of the Wikipedia pages that comprise MONSTr.

Although the use of Wikipedia in social science research remains relatively nascent, much has been done to evaluate its validity and reliability ([Greenstein and Zhu 2012](#)). In studies of over 10,000 randomly selected articles, scholars found Wikipedia’s accuracy on fact-based claims to be comparable to the Encyclopedia Britannica, the Dictionary of American History, and American National Biography Online using predictors such as page length, number of citations, external links, and edit history ([Rector 2008](#)).⁶ Specific to political science, a study on Wikipedia’s coverage of 246 major-party gubernatorial candidates contained no errors in candidate biographies and 242 out of the 246 (98.4%) reported election results within 0.2% of the actual outcome ([Brown 2011](#)). Validation exercises on party ideology data from Wikipedia similarly deem that it bears correct, consistent information ([Herrmann and Döring 2023](#)). Inconsistencies between (and within) Wikipedia and other sources largely

⁵Please go read the entertaining opening of this short article, which we concede is better than ours.

⁶Contrary to many course syllabi, Wikipedia vandalism is not as common as alleged: 81% of uncited claims are flagged by bots using automated vandalism detection within 24 hours and fixed within three hours ([Tramullas, Garrido-Picazo and Sánchez-Casabón 2016](#)).

pertain to frequently disputed figures (i.e., the precise size of armed forces in a battle) and more subjective interpretations (i.e., rationales and contested effects), datapoints that all sources struggle to pinpoint ([Oswald and Ohrenhofer 2022](#)).

In addition to its accuracy, scholars have also evaluated Wikipedia’s coverage. Wikipedia decides if an event should have its own page based on a criteria of “notability.”⁷ An event is considered sufficiently notable to have its own article if it “has lasting major consequences or affects a major geographical scope, or receives significant non-routine coverage that persists over a period of time. Coverage should be in multiple reliable sources with national or global scope.” Whether a page meets this criteria is decided by a combination of human editors and bots that evaluate the quality of new pages. Known biases in what meets these criteria have been uncovered, including a bias towards inclusion of recent events ([Royal and Kapila 2009](#)) and those relevant to the English-speaking and Western world ([Callahan and Herring 2011](#)). Comparisons can be drawn to media sources that have known biases in coverage of conflict ([Dietrich and Eck 2020](#); [Croicu and Eck 2022](#)), which can impact event datasets that scholars use by biasing their sample, detail, and accuracy ([Douglass, Scherer, Gannon, Gartzke, Lindsay, Carcelli, Wilkenfeld, Quinn, Aiken, Navarro, Lund, Murauskaite and Partridge 2022](#)). Attempts to compare Wikipedia’s coverage to news sources are complicated by their increased endogeneity, as news sources now use Wikipedia to decide what to cover ([Ren and Xu 2023](#)). Nonetheless, Wikipedia’s use of nonprofit, scholarly, and government sources as well as its ability to update pages based on new information means it is often more complete than media sources and also covers events that were only revealed or determined to be important well after they occurred ([Gildersleve, Lambiotte and Yasseri 2022](#)).

In addition to front-end and process audits, the authors manually checked variable values for a randomly selected sample of 6% of all observations prior to identifying the operation-

⁷Inclusion criteria differ by page type. All observations in MONSTr fall under “events,” for which inclusion criteria is detailed at [Wikipedia:Notability \(events\)](#).

level ones. Validation sources stemmed from three sources. First, after consulting with personnel from all four of the major military branches to identify reputable government sources containing information on military operations. Second, in cases where none of the selected government or military sources contained information about an operation, we turned to reputable think tank and NGO reports or news outlets. To avoid circularity in validation, we were mindful to select news outlets that were not themselves cited in the Wikipedia article, although that was not possible in a few instances. We also filed two Freedom of Information Act (FOIA) requests to obtain After Action Reports (AAR), Battle Damage Assessments (BDA), and ship deck logs for known cases where little open source evidence existed. With the exception of uncertainty about the precise end date of two observations and whether one operation was against al-Qaeda or al-Qaeda and the Taliban, Wikipedia’s coding for all other values could be independently verified. The manual audits also shed light on Wikipedia’s coverage and inclusion criteria. Small and uneventful military operations were more likely to be listed as a subheader on an existing page or in a table with a brief summary, resulting in their exclusion from MONSTr. Contrary to findings about Wikipedia’s Western and in-group bias, operations featuring high civilian casualties or misidentified targets usually appeared more detailed than those where the US accomplished its operational objective. A full description of the manual validation process is provided in the Appendix.

Clustering - Dependencies and Levels

Existing datasets on military intervention rarely record the relations between cases. This limitation can complicate the ability to analyze decision-making in the same dependent, dynamic context as political and military leaders. As an example, while MIPS, ICB, and MIDs have one observation for the two-decade War in Afghanistan, IMI considers the attacking and peacekeeping portions separately and RAND includes Operation Enduring Freedom (OEF), Operation Swift Freedom, International Security Assistance Force, Resolute Support and

Freedom’s Sentinel. In reality, Operation Swift Freedom was *part of* OEF - Afghanistan, Freedom’s Sentinel was a sub-campaign of Resolute Support, and of course all of these occurred under the umbrella of the Afghanistan War.

Since MONSTr delves into the operational level, we regarded a particular urgency to capture the higher levels in which each observation is situated. For each case, Wikipedia identifies a “part of” variable that indicates the broader event to which the former is a subset. The level of ancestors (or higher levels of aggregation) for each operation varies. Some simply have one: Operation Uphold Democracy is part of the 1991 Haitian coup d’état. Others have more: the December 2018 airstrikes in Gandarshe are part of the American intervention in Somalia, which is a part of OEF – Horn of Africa, which is part of OEF. This Wikipedia feature enables us to identify and code the campaign and strategic levels in which each operation ought to be nested. To accomplish this, we constructed a dendrogram of the entire dataset, the trunk being any US military intervention in our time period and each branch being an increasingly intricate disaggregation from strategic to operational to tactical levels. We then independently and manually audited all 491 lines to cull operations, paring events at the tactical level, ensuring definitional accuracy, and validating ancestors. This resolved ambiguity for observations that were a part of multiple events, as one could be identified as the immediate parent (War in Afghanistan) and the other a higher level ancestor (War on Terror).

Notably, Wikipedia page names are not always an indication of the event level. They sometimes represent military designations and sometimes public labels. Technically, wars contain campaigns, which contain operations, which contain battles. In Wiki terminology, some entries entitled “Operation” are not operations according to the DoD definition, some events bearing names like “Battle of” are operations, and at times operations are nested within campaigns that are labeled as battles. For instance, Operation Euphrates Shield is a broader campaign containing four operations featuring alternative labels, like the East

Aleppo Offensive (2017) described as a military operation to capture the Islamic State stronghold of Dayr Hafir and gain control of the city’s water source and airbase. As another example, the Battle of Panjwai is a two-phase campaign to withstand Taliban forces in the district composed of two operations—Zahara and Medusa. Similarly, the RDF labels for “instance of” abide by the same nomenclature inconsistencies and cannot be used to determine the case inclusion, which was instead done by the authors.⁸ During our operation identification exercise, we focused on the series of actions and levels of objectives rather than the Wiki names or sizes of events. For any discrepancies, we subjected cases to further scrutiny and debate until reaching consensus. The hierarchical relationships are a unique feature of our dataset that yields several advantages. It allows us to code, for example, how many days into OEF did the 2018 airstrikes occur. Furthermore, it facilitates hierarchical models accounting for structural dependencies among operations borne of the same strategic campaign and parent-intervention fixed effects.

Distinctive Outputs

Granularity of Coverage

As already acknowledged, there are several datasets on military interventions at the strategic level. We do not supplant, but supplement them by identifying the more granular operations within them. For instance, most existing intervention datasets have an observation for the 2003 Iraq War. Ours identifies military operations and codes covariates associated with the individual operations within that war, as shown in Table 1. By disaggregating the unit of analysis, MONSTr provides information on 313 operations representing the 65 unique US interventions covered in all post-1989 datasets combined. This dimensionality is imperative

⁸For example, Operation Vigilant Sentinel is an instance of “conflict” while Operation Northern Watch is an instance of “military operation.” For a more thorough evaluation of the scope and relationships of conflict labels in Wikipedia, see [Douglass, Scherer, Gannon and Gartzke \(2022\)](#).

for understanding how states fight. Reaffirming the connection between the study of strategy and the empirical operation we have sketched from other academic and military publications, US Army Field Manual 100-5 (1993) states that strategic objectives and military battles are harmonized by *operational* art. Like going to war with the Army you have rather than the one you want, scholars examining force employment have done the best they can with the data that exist. Armed with Wikipedia, we can now go to war against unanswered research questions with the data we want.

[Table 1 about here]

Depth of Datapoints

In addition to increasing the granularity of existing data by improving the precision of our unit of analysis, we were able to extract more granular variables characterizing operations. Shown in Table 2, MONSTR stacks up well against existing datasets on objective factors. It features start and end dates for each operation down to the day and a count of casualties (including civilians) for each side. The dataset also contains indicators for seven different types of force, an improvement over blunt domain codings collected at the intervention level or focuses on a single type when it comes to our research questions. Most operations include geocoded latitude and longitude coordinates, facilitating finer spatial analyses. It also identifies all state and nonstate actors active in the operation, both initiators and targets. Notably, MONSTR does not feature information on the aim and outcome of events. Treading lightly through data norms, we recognize that there is subjectivity afoot in discerning factors like military versus political objectives, *ex ante* and *ex post* objectives, and in interpreting outcomes. Although Wikipedia often includes these variables, we exclude them absent reliable evidence that it would be an improvement over the existing state of the art.

[Table 2 about here]

Means of Force

[Hart \(1967, 335\)](#) defines military strategy as “the art of distributing and applying *military means* to fulfill the ends of policy.” Although strategy involves more than simple force structure design and deployment, these are nonetheless crucial components. The dearth and incomparability of the means of force in existing datasets are a chief reason we turned to Wikipedia since we could extract this variable for each operation. After several iterations of coding, auditing, and refining, we observed a co-occurrence of seven main categories of force: ground troops, paramilitary, close air support, air to air combat, aerial bombing, cruise missiles, and drones. Although some intervention datasets identify naval events during this time period, those events are identified in MONSTr by how the target was attacked, rather than the launch point (ie cruise missiles launched from a destroyer are coded as cruise missiles and an island landing for a raid is coded as ground troops). We explain this choice and our divergence from all cases coded as naval elsewhere in detail in the Appendix, allowing scholars to recode relevant cases in MONSTr as naval if desired.

Each force structure has its tradeoffs. Boots on the ground are more versatile, discriminating in context, and effective at winning hearts and minds in counterinsurgent settings (when done well). At the same time, they are more directly exposed to enemy attack, implying higher casualties, and by numbers and interaction density alone more prone to natural human error. Manned aerial platforms reduce risks and a degree of error, but exchange this for context-specific and time-sensitive intelligence and population-centric social capital ([Lyall and Wilson 2009](#)). Unpiloted platforms (cruise missiles and drones) represent the lowest tactical risk to troops but the highest political risks of blowback ([Johnston and Sarbahi 2016](#)) and downsides to decapitation campaigns ([Price 2012](#)). In addition to risks and effects, each of these means differently signals resolve ([Zegart 2018](#)).

Figure 1 shows the cumulative count of the means of force across the time scope of the

dataset. Several confirming trends stand out. The Gulf War air campaign was potent and swift, ratcheting up operations in a steep blip. Given debates about whether air power drove victory or was overrated, having these fine-grained data will help scholars adjudicate its role. The 2007 Iraq surge, largely comprising ground troops, is readily apparent. The concentration of aerial bombing in the Syria campaign is also conspicuous, supporting notions that Obama was trying to signal resolve to the Assad regime. Overall, having measures of the means of force grants us great purchase in the study of military force employment and the temporality of war.⁹ Descriptive statistics for the types of force in the dataset and other select covariates feature in the online appendix.

[Figure 1 about here]

Nested Interdependencies

Researchers have long expressed concern over un-modeled interdependencies between conflict events. While frequently acknowledged in theories, the empirical mismatch due to limitations in available data problematize the important inferences scholars can make. Leveraging the structure of Wikipedia’s internal architecture, we identified and coded the nested relationships for each operation. Figure 2 illustrates the high degree of interdependence among operations in our dataset. Moving to the right disaggregates each node into the operations that are part of it. Figure 3 homes in on the 1991 Gulf War to provide a more focused example. The Gulf War can be decomposed into three broad approaches: the no-fly zones, liberation of Kuwait, and air campaigns. Each features several operations in support of its objective. Looking at the distribution of the means of force along the right side of the figure, the importance of controlling for strategic campaigns becomes apparent. Intuitively, the aerial-focused ones feature higher frequencies of air power while the liberation campaign

⁹The amount of missiles fired or troops deployed was of initial interest, but accurate data, especially concerning ordnance, proved challenging.

involves more ground troops, often in tandem with close air support. Although Wikipedia’s nested dependencies are not a perfect mapping relative to military or political dependencies, they will help scholars to better model this vital trait.

[Figures 2 and 3 about here]

Geographic Information

Nearly all operations in MONSTr are geocoded with GPS coordinates or a local municipality indicating where the event occurred.¹⁰ For spatial analysts and geographic information system (GIS) users, this opens a wealth of inquiries. For instance, it allows for the coding of terrain, which might influence the force structure deployed (Gleditsch and Weidmann 2012; Shaver, Carter and Shawa 2019). It also allows analysis of urbanicity, proximity to strategic assets, adaptation amid shifting combat fronts, power projection, contagion, and more (limited only by GIS layers and imagination). In Figure 4, we plot the locations of the US operations in the dataset and identify one potential application – the association between urban/non-urban targets and the type of military force used. Although merely descriptive, the types of force used in operations in Syria, Iraq, and Afghanistan raise interesting questions. Non-urban operations in Iraq appear to primarily involve ground troops except in the mountainous north, where aerial bombing dominates. The majority of US operations in Afghanistan occur in non-urban areas, with those crossing the Pakistan border constituting aerial operations typically associated with US drone strikes during the Obama administration. This descriptive evidence suggests ways in which these data could be used to test theories about the relationship between geography, target type, and the means of applications of military force.

[Figure 4 about here]

¹⁰In cases where the exact location was unknown, additional sources were consulted to identify the location with the most precisely reliable location. Those sources and methods are documented in the dataset.

As a GIS-capable team, we are enthusiastic about these prospects. Still, as an intellectually honest team, we caution interpretations about the geography of US military operations from these data. First, Wikipedia does not distinguish between geometry type, which matters for events that occurred in multiple districts or towns. In these cases, we deferred to the coordinates provided by the DBpedia infoboxes to ensure a consistent protocol and to enable future researchers to compare this to other sources. In addition, we addressed non-random missingness using Google Earth data which also raises questions about the consequences of geographic coding using multiple gazetteers ([Douglass and Harkness 2018](#)).

Actors Involved

The identity, traits, and connections of actors in conflict have a dramatic impact on how they fight. Scholars have long recognized and researched this ([Huth 1998](#)). Powerful states, more likely to have economic assets to protect and military ones to project power, intervene more frequently ([Sullivan and Koch 2009](#)). Democracies might deploy force for different reasons, in different ways, and for different durations than autocracies ([Caverley 2014](#); [Talmadge 2016](#)). Some states are primarily compelled by security, some economy, some to contain spillover of violence or refugees, some drawn in by allies or against rivals, and some by humanitarian impulses.¹¹ Considering intervenors alone, multilateral campaigns—distributing costs and risks across partners yet involving larger coordination challenges—require different calculus than unilateral ones. For example, in Serbia 1996 US generals realized that to keep the 19 North Atlantic Treaty Organization (NATO) allies on board, their preference of a saturated, decisive air campaign was off the table ([Cooper 2001](#)). Speaking only of targets, states might behave differently against nonstate combatants relative to fellow states. And in the same way that multilateral state coalitions change the fighting calculus, nonstate actors form complex and shifting coalitions too. The Syria Civil War, with its makeshift operations rooms

¹¹The many good works exploring each determinant are too many to list here.

and mercurial alliances, illustrates this well ([Cranmer and Menninga 2018](#)). In short, the consequences of intervention hinge on collective actor traits on both sides and interactions across ([Carment and Rowlands 1998](#); [Caverley 2009](#)).

While we know these factors matter, much analysis oversimplifies by looking at the primary intervenor and target. This masks important variation in participants, especially those occurring over years or even decades. The roster comprising the coalition throughout the Afghanistan War, for instance, oscillated over its life. Even if it remained constant, not all members participated in each operation, to the same degree, or for the same reason ([Gannon and Kent 2021](#)). Nor did the targets remain constant over the course of the conflict. At times, the US unilaterally deployed special forces against anti-coalition insurgents (Operation Red Wings). At other times, it pitched in beside Dutch pilots to provide close air support for Aussies and their Afghan mentees ambushed by the Taliban (First Battle of Kakarak). Unsurprisingly, the force employments selected for these events were quite different. Who fights, whom they are fighting alongside, and whom they are fighting against are important for explaining comportment in conflict ([Cappella Zielinski and Grauer 2022](#)).

Our data capture this variation in actor interaction as showcased in Figure 5. These three network plots of the US and its partners (side A) feature state allies in green and nonstate actors fighting alongside the US in orange. The edges are weighted by the number of operations shared between dyads. Maintaining focus on the Afghanistan War, the top right panel portrays the many actors participating throughout the 20-year war. Intuitively, the US most densely interacted with the state of Afghanistan itself after ousting the Taliban. While key Western partners situate strongly in the network, direct US involvement with the United Kingdom (UK) was most prominent. Other interesting traits emerge in comparison to the 1991 Gulf War (top left) and the 2015 Syria Civil War (bottom left). First, the composition of US partners changed dramatically across these conflicts. All allies in the Gulf War were states. In Afghanistan, state partners remain the norm but plenty of nonstate actors joined

the fray. In the Syria War, nonstate partners outnumber states two-to-one. Edges shared with the UK are thick throughout, but in Syria US cooperation with a nonstate actor eclipses it by a small margin. Interaction density between actors varies across conflicts as well, being straightforward and balanced in the Gulf, layered and feathery in Afghanistan, and sprawling and asymmetric in Syria. Even with only allies in only three strategic level interventions, we see intriguing patterns that spark questions. MONSTr affords the ability to analyze them in significantly more detail.

[Figure 5 about here]

Demonstration

As a proof of concept demonstrating the validity and value of these data, this section presents the results of statistical models based on existing theories about why states deploy particular military means. The goal of these models is to see how the new, more disaggregated data from a novel source compares to existing findings. While the more fine-grained unit of analysis makes an exact replication impossible, the results of these models should nonetheless provide evidence that is either consistent or inconsistent with existing research. Existing research analyzes the determinants of ground forces, air power, and a rare few examine the conditions of their combination. We emulate this to compare general patterns and findings.

Specification

The unit of analysis is the US military operation. For the dependent variables, we take advantage of the disaggregated types of force available in our dataset. Thus, each model in Table 3 features coefficients for a separate binary dependent variable indicating the use of a key means of force: GROUND, AIR AND GROUND in conjunction, AIR, and subsets of air

power to explore the use of PILOTED and UNPILOTED platforms.¹² The dependent variables are binary and the data are nested, calling for a logit estimator in hierarchical models that account for the strategic dependencies of operations within each broader war.

We also build upon previously theorized explanators for state force employment. Past studies have examined factors such as target type, geography, and timing. First, we include an indicator for whether side B features a NONSTATE TARGET, which tend to use different strategies to deflect the brute force of stronger state actors (Arreguín-Toft 2001). Second, we consider two geographic features that likely affect the feasibility of different force structures. MOUNTAINOUS terrain is more difficult for foot soldiers to traverse, especially if combatants are advantageously embedded throughout. URBAN settings are complex operating environments, dense with structures, sensitive infrastructure, and civilians, so might be less amenable to higher-tech means of force. Finally, strategies and force employment design might change over the life of a conflict. We include a DURATION variable and a measure capturing how many DAYS INTO THE PARENT conflict the operation occurred, both logged to normalize their distributions. All models include president fixed effects to control for leader idiosyncrasy and cubic polynomials of the number of years since the dataset’s start to control for nonlinear time dependence (Carter and Signorino 2010).

Results

Model results are reported in Table 3. Concerning the target of US operations, it appears that those with nonstate combatants are much less likely to feature combined ground and air tactics or air power pooled. When it comes to geography, operations taking place in mountainous terrain are associated with more aerial bombing, bringing to mind the infamous 72-hour bombing campaign of the Tora Bora mountain caves targeting Osama bin Laden

¹²Since 1) the dependent variables span the same range and 2) models stem from the same sample and feature the same explanators and specifications, we are on safe statistical ground to compare across models.

in late 2001. No clear association exists with urban targets, which runs counter to theories that drones’ more precise and discriminating nature makes them well-suited to operations in complex environments dense with civilians. Instead, the US may avoid drone strikes in urban areas because of the domestic backlash to civilian deaths.

Intriguingly, longer operations are associated with an increased frequency of piloted air power, yet air operations do not seem to feature earlier in a conflict. These temporal variations warrant further consideration and theorizing. [Martinez Machain \(2015\)](#) demonstrates that ground troops affect the duration of aerial campaigns divergently depending on how discriminate the latter’s approach. While we don’t observe that the days into the parent intervention affect force employment choices, suggesting that this hinges more on conflict dynamics than mere time, we do note that longer interventions tend to feature more ground, combined arms, and manned air power. Overall, the results mostly affirm conventional wisdom, but also point to new puzzles that the dataset will enable scholars to examine.

[[Table 3](#) about here]

Conclusion and Future Directions

With the advent of automated extraction, and the audacity to apply it to a well-known yet long-avoided source, we were able to collect a first cut of data in a relatively easy and well identified sample as a proof of concept. The Wiki-based MONSTr dataset features objective correlates of the operations within nearly all US military interventions from 1989 to January 2021 in a nested structure reflecting observation interdependencies. Its unit of analysis is more granular, its depth more detailed, and its structure more realistic than existing sources. It also features the means of force used in each operation, opening avenues to explore fresh puzzles about the prosecution of conflict. We have scantily demonstrated the potential applications of these data, showing preliminary patterns of force using actor

and event traits. Many findings accord with existing theory – like the use of aerial bombing in mountainous environments – while others provoke new ideas – like an aversion to piloted bombing against nonstate targets and fewer drone strikes against urban targets.

We observe several avenues for future research, both theoretical and empirical. Theoretically, we wonder about the determinants of force employment, both retesting conventional wisdom and expanding it. In particular, we expect that domestic public factors constrain and compel the force employment of strong, democratic states. We also promote explorations of actor dynamics on both sides and across. As conflicts increasingly feature nonstate actors as partners and targets, future research can unpack how states fight alongside or against them given the asymmetry in strength and unconventional strategies. We also see potential to explore the temporality of war, how operations change across the life of a parent conflict. Our demonstrative models suggest that the means of force shift the longer the intervention yet perhaps not depending on depth into the parent campaign. Conversely, scholars can examine the impact of force employment on domestic political moods, strategic leader behavior, partner configurations, duration, and more. We see value in analyzing the combinations of military means in tailored force employment packages. For instance, ground forces and close air support no doubt co-occur and unmanned air power likely flies alone more often than not. Empirically, despite the considerable expansion in coverage and granularity, our data are bounded in a brief time period for one nation. Data stretching further back and abroad are there, awaiting extraction. Given the machine-readable format, it is exceptionally easier to approach an endeavor of this scope than with manual efforts (once you learn SPARQL and Wiki’s anatomy).

There remain many research questions on how states fight. Scholars have done impressive work with limited data on the means of force, yet their siloed, specific nature and high levels of aggregation inhibit inquiry into the full, operational spectrum of military power. Especially since means of force are substitutable and combinable, understanding their interaction

is critical to advance knowledge. On top of that, the practical assumption of independence across observations in existing datasets separates events that are strategically linked. Producing a dataset that overcomes these challenges and continues to offer the same depth and precision (in fact, more) seems Herculean. We are well aware and in great admiration of the painstaking, long-term efforts to stitch together data on military interventions from dozens of sources. Although we acknowledge Wikipedia’s drawbacks, we maintain that, with transparency about these traits, we can learn a great deal about the conduct of war that we would otherwise not know. We have no expectation or hope that a Wiki-based dataset will replace or compete with others. We do, however, expect and hope that it will expand in coverage and use. Combined with more rigorous qualitative and quantitative approaches, we believe our data make a meaningful contribution to the social scientific and practitioner communities aiming to accumulate knowledge. We hope MONSTr grows. Many will choose not to use it but we hope in the end some will profess, “I love you too much to condemn you” ([Coppola 1992](#)).

Table 1: Sample of Iraq War (2003) operations

Operation	Means				Location		Time				Casualties		
	Aerial bombing	CAS	Ground troops	Paramilitary	Lat	Lon	Start	End	Duration	Days into war	Side A	Side B	Civilian
Operation Augurs of Prosperity	0	0	1	0	33.77335	45.14945	2008-07-29	2008-08-11	13	1958	58	15	70
Operation Planet X	0	0	1	0	31.99662	44.31640	2003-05-15	2003-05-15	0	56	NA	NA	NA
Operation Bayonet Lightning	0	0	1	0	35.32500	43.77222	2003-12-02	2003-12-02	0	257	NA	NA	NA
Operation Spear	1	0	1	0	34.38431	41.01832	2005-06-17	2005-06-22	5	820	NA	NA	4
Operation Imposing Law	0	1	1	0	33.31524	44.36607	2007-02-14	2007-11-24	283	1427	872	1219	7482
Operation Marne Courageous	0	1	1	0	33.31524	44.36607	2007-11-16	2007-11-16	0	1702	0	0	0
Operation Guardian Tiger IV	0	0	1	0	34.13972	42.37806	2006-08-01	2006-09-30	60	1230	0	NA	NA
Operation Together Forward	0	0	1	0	33.31524	44.36607	2006-07-09	2006-10-24	107	1207	298	400	NA

Table 2: Comparison of Variable Granularity

Dataset → Variable ↓	<i>MONSTR</i>	<i>MIPS</i>	<i>IMI</i>	<i>ICB</i>	<i>MIDs</i>	<i>MIP</i>	<i>UCDP/ PRIO</i>	<i>RAND</i>	<i>Allen</i>	<i>Horowitz & Reiter</i>
<i>Scope</i>	1989-2021	1945-2001	1946-2005	1918-2015	1816-2010	1776-2019	1946-2018	1900-2015	1914-2003	1917-1999
<i>N (US, all dates)</i>	336	35	109	71	325	392	18	145	49	17
<i>N (US, post-1989)</i>	336	16	38	20	72	115	17	64	4	5
<i>Date</i>	day	day	day	day	day	day	day	year	year	year
<i>Casualties</i>	count	count	count	ordinal	ordinal	count	-	-	-	-
<i>Military Means</i>	all	ordinal	domain	-	-	domain	-	-	air, land	air
<i>Geography</i>	geocoded	state	state	subcontinent	-	state	territory	state	-	-
<i>Intervenors</i>	all	primary	primary	all	primary	primary	primary	primary	primary	all
<i>Targets</i>	all	all	primary	primary	primary	all	primary	primary	primary	all
<i>Target Type</i>	-	✓	-	-	-	-	✓	-	✓	-
<i>Objective</i>	-	✓	✓	✓	-	✓	-	-	✓	✓
<i>Outcome</i>	-	✓	✓	✓	✓	✓	-	-	-	✓

Table 3: Exploratory Models of Determinants of Means of Military Force

	Ground	Ground and Air	Air (all)	Air (Piloted)	Air (Unpiloted)
Non-state target	0.542 (0.477)	-1.251** (0.037)	-1.787** (0.017)	-1.501** (0.016)	-0.336 (0.694)
Mountainous target	-0.591 (0.207)	0.354 (0.301)	0.788* (0.064)	0.509 (0.183)	0.225 (0.737)
Urban target	0.339 (0.412)	0.300 (0.373)	0.156 (0.675)	0.314 (0.356)	-0.620 (0.289)
Duration (log)	0.305*** (0.001)	0.238*** (0.002)	0.001 (0.992)	0.126* (0.090)	0.015 (0.895)
Day into Parent (log)	0.164 (0.267)	0.137 (0.289)	-0.037 (0.784)	0.016 (0.892)	-0.087 (0.632)
Num.Obs.	260	260	260	260	260
R2 Marg.	0.511	0.430	0.459	0.401	0.324
R2 Cond.	0.565		0.491	0.435	0.578
AIC	274.8	328.7	300.2	340.1	196.5
BIC	328.2	382.1	353.6	393.5	249.9
ICC	0.1		0.1	0.1	0.4
RMSE	0.38	0.44	0.41	0.44	0.28

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

All models include president fixed effects and cubic spline polynomials.

Figure 1: Timeline of US Military Operations

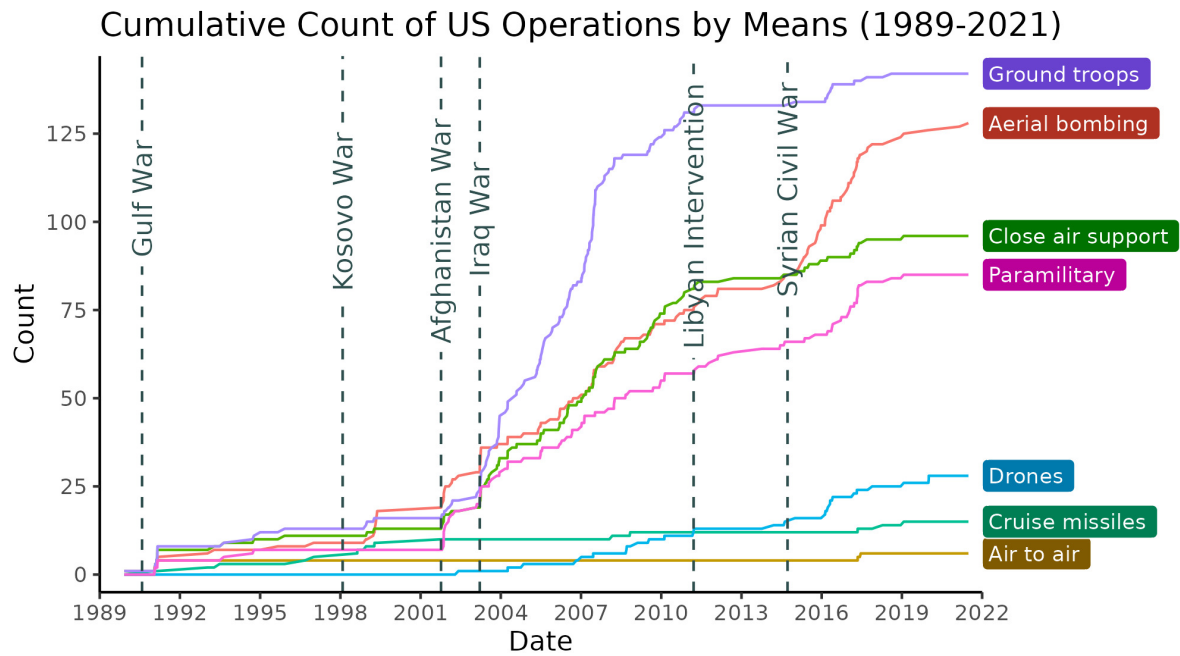


Figure 2: Nested Dependencies of US Military Operations

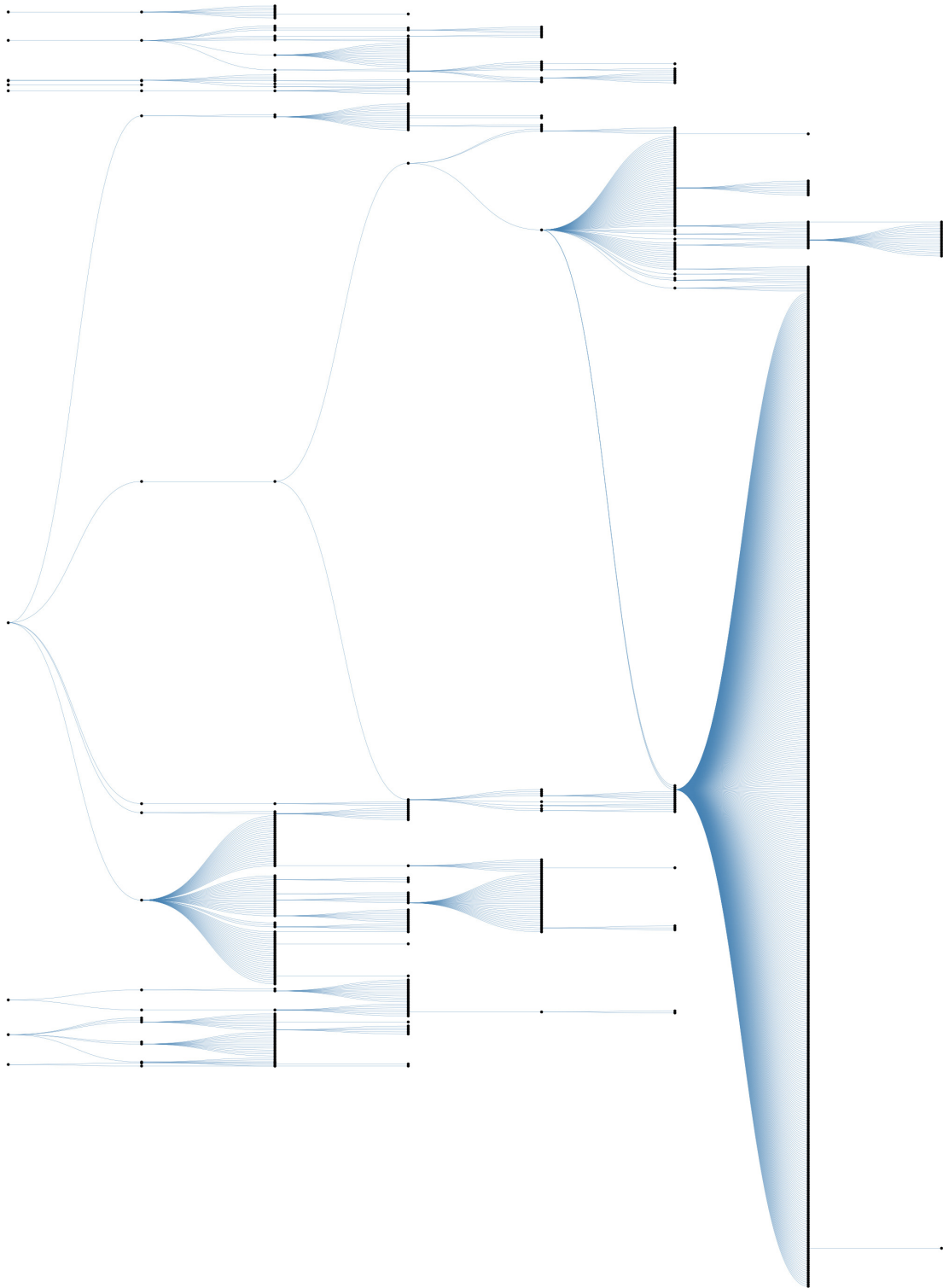


Figure 3: Nested Structure of the Gulf War (1991)

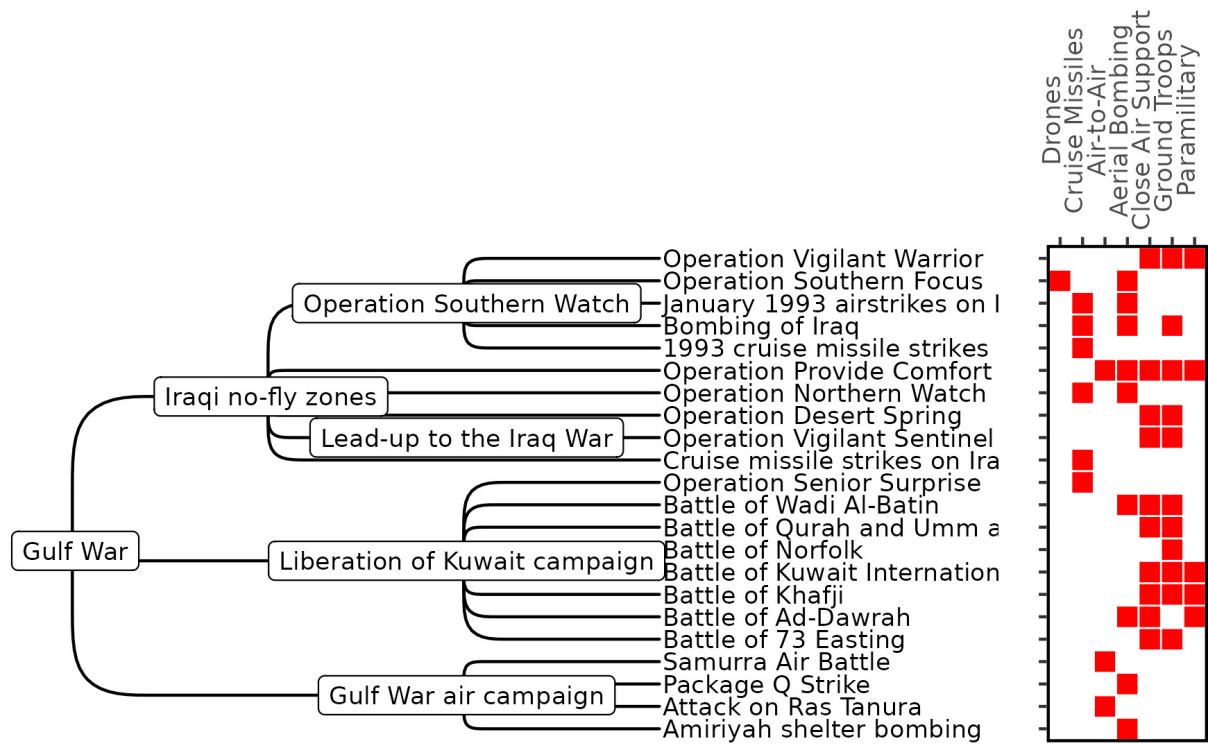


Figure 4: Locations of US Military Operations (1989-2021)

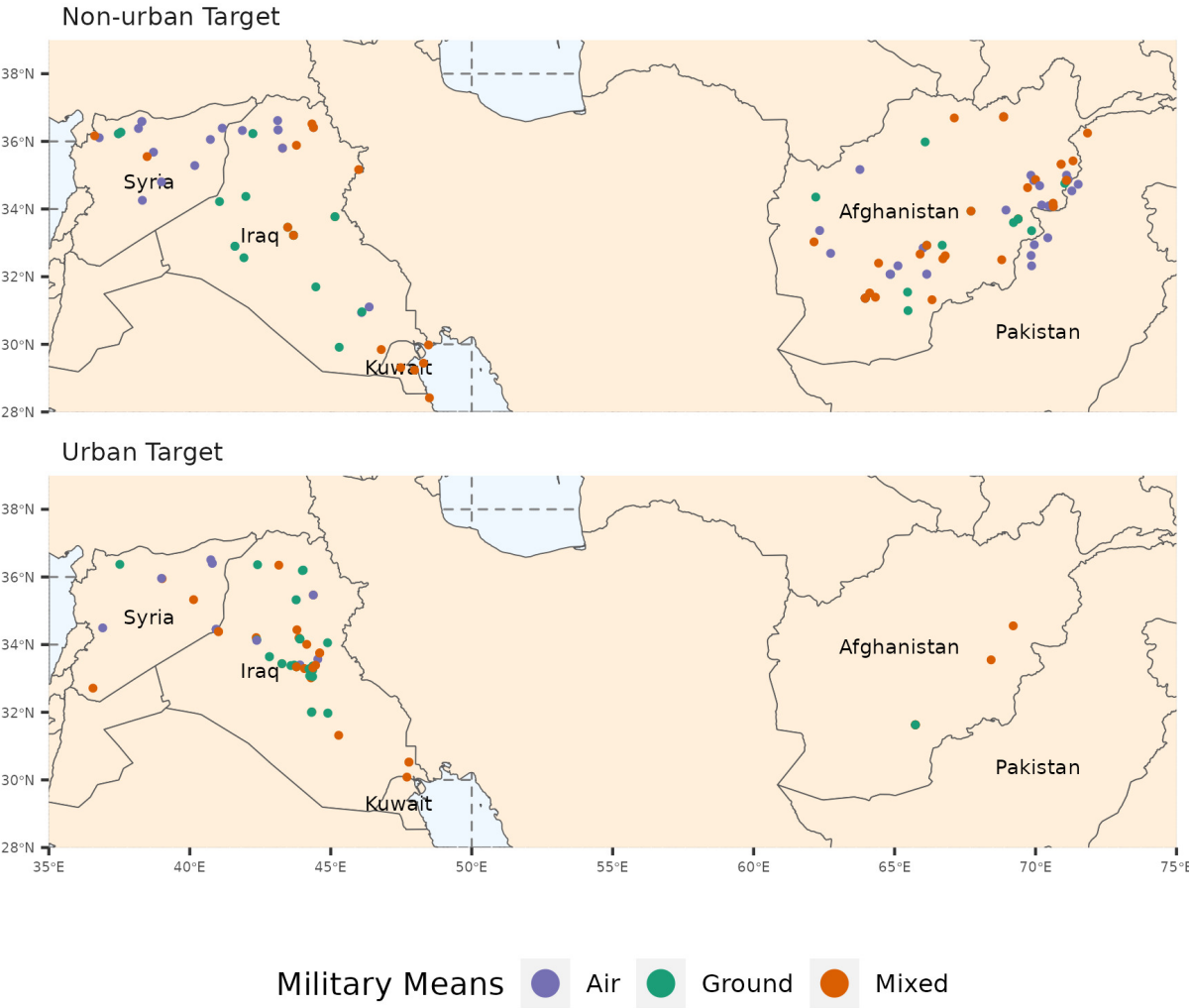
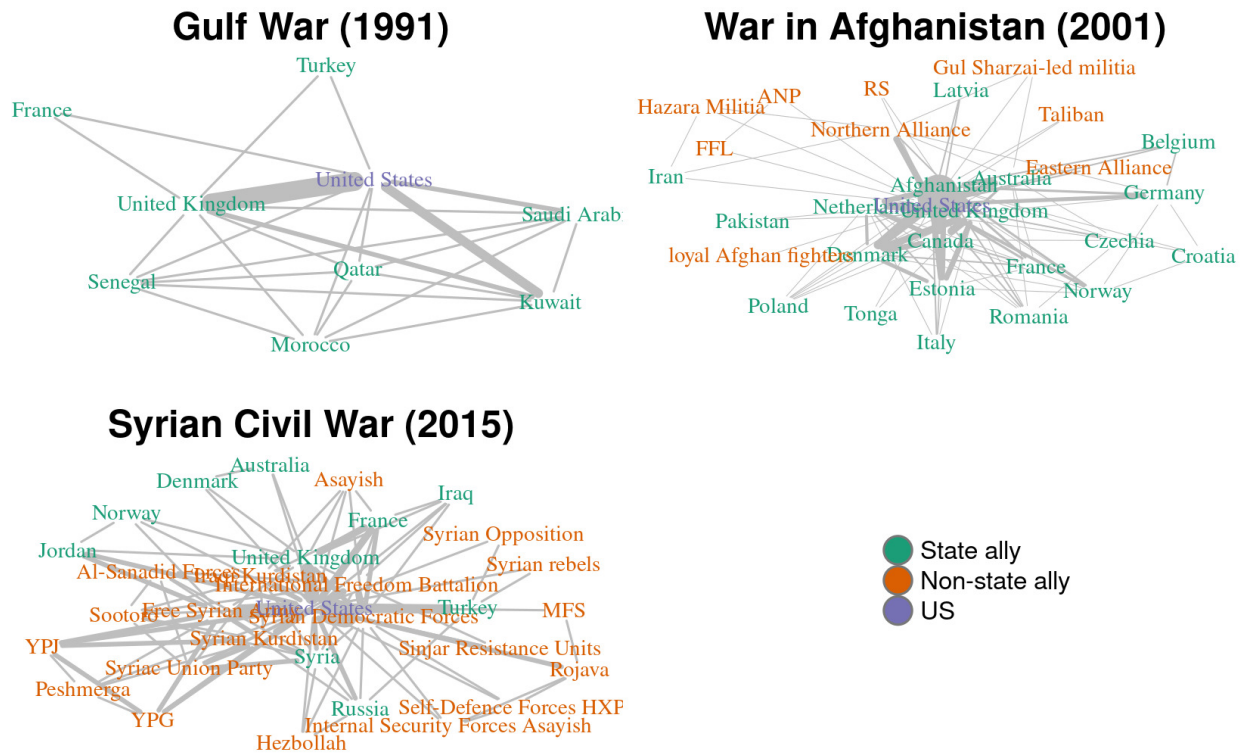


Figure 5: Social Network Plots of Select Conflicts



References

- Allen, Susan Hannah and Carla Martinez Machain. 2019. “Understanding the Impact of Air Power.” *Conflict Management and Peace Science* 36(5):545–558.
- Arreguín-Toft, Ivan. 2001. “How the Weak Win Wars: A Theory of Asymmetric Conflict.” *International Security* 26(1):93–128.
- Becker, Bernd W. 2015. “Research Faux Pas: The Stigma of Wikipedia.” *Behavioral & Social Sciences Librarian* 34(3):165–169.
- Becker, Jo and Scott Shane. 2012. “Secret ‘Kill List’ Proves a Test of Obama’s Principles and Will.” *The New York Times* .
- Brown, Adam R. 2011. “Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage.” *PS: Political Science & Politics* 44(2):339–343.
- Callahan, Ewa S. and Susan C. Herring. 2011. “Cultural Bias in Wikipedia Content on Famous Persons.” *Journal of the American Society for Information Science and Technology* 62(10):1899–1915.
- Cappella Zielinski, Rosella and Ryan Grauer. 2022. “Understanding Battlefield Coalitions.” *Journal of Strategic Studies* 45(2):177–185.
- Carment, David and Dane Rowlands. 1998. “Three’s Company: Evaluating Third-Party Intervention in Intrastate Conflict.” *The Journal of Conflict Resolution* 42(5):572–599.

- Carter, David B. and Curtis S. Signorino. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18(3):271–292.
- Caverley, Jonathan D. 2009. "The Myth of Military Myopia: Democracy, Small Wars, and Vietnam." *International Security* 34(3):119–157.
- Caverley, Jonathan D. 2014. *Democratic Militarism: Voting, Wealth, and War*. United Kingdom: Cambridge University Press.
- Chairman of the Joint Chiefs of Staff. 2018. Joint Publication 3-0 Joint Operations. Joint Publication 3-0 The Joint Staff Washington, DC: .
- Chairman of the Joint Chiefs of Staff. 2021. "DOD Dictionary of Military and Associated Terms." .
- Cooper, Scott A. 2001. "The Politics of Airstrikes." *Policy Review; Washington* (107):55–66.
- Coppola, Francis Ford. 1992. "Dracula." .
- Cranmer, Skyler and Elizabeth Menninga. 2018. "Coalition Quality and Multinational Dispute Outcomes." *International Interactions* 44(2):217–243.
- Crisher, Brian and Mark Souva. 2014. "Power at Sea: A Naval Power Dataset, 1865-2011." *International Interactions* 40(4):602–629.
- Croicu, Mihai and Kristine Eck. 2022. "Reporting of Non-Fatal Conflict Events." *International Interactions* 48(3):450–470.

- Dietrich, Nick and Kristine Eck. 2020. “Known Unknowns: Media Bias in the Reporting of Political Violence.” *International Interactions* 46(6):1043–1060.
- Douglass, Rex W. and Kristen A. Harkness. 2018. “Measuring the Landscape of Civil War: Evaluating Geographic Coding Decisions with Historic Data from the Mau Mau Rebellion.” *Journal of Peace Research* 55(2):190–205.
- Douglass, Rex W., Thomas Leo Scherer, J Andrés Gannon and Erik A. Gartzke. 2022. “Connecting Conflict Data through Wikimedia.”.
- Douglass, Rex W., Thomas Leo Scherer, J Andrés Gannon, Erik Gartzke, Jon Lindsay, Shannon Carcelli, Jonathan Wilkenfeld, David M. Quinn, Catherine Aiken, Jose Miguel Cabezas Navarro, Neil Lund, Egle Murauskaite and Diana Partridge. 2022. “Introducing the ICBe Dataset: Very High Recall and Precision Event Extraction from Narratives about International Crises.”.
- Early, Bryan Robert, Nolan Fahrenkopf, Michael C. Horowitz and James Igoe Walsh. 2022. “Climbing the Ladder: Explaining the Vertical Proliferation of Cruise Missiles.” *Journal of Conflict Resolution* 66(6):955–982.
- Fuhrmann, Matthew and Michael C. Horowitz. 2017. “Droning On: Explaining the Proliferation of Unmanned Aerial Vehicles.” *International Organization* 71(2):397–418.
- Gannon, J Andrés. 2021. Planes, Trains, and Armored Mobiles: Introducing a Dataset of the Global Distribution of Military Capabilities (rDMC). SSRN Scholarly Paper ID 3930390 Social Science Research Network Rochester, NY: .

- Gannon, J Andrés. 2022. “One If by Land, and Two If by Sea: Cross-domain Contests and the Escalation of International Crises.” *International Studies Quarterly* 66(4).
- Gannon, J Andrés and Daniel Kent. 2021. “Keeping Your Friends Close, but Acquaintances Closer: Why Weakly Allied States Make Committed Coalition Partners.” *Journal of Conflict Resolution* 65(5):889–918.
- Gartzke, Erik A. 2001. “Democracy and the Preparation for War: Does Regime Type Affect States’ Anticipation of Casualties?” *International Studies Quarterly* 45(3):467–484.
- Gartzke, Erik A., Jeffrey M. Kaplow and Rupal N. Mehta. 2014. “The Determinants of Nuclear Force Structure.” *Journal of Conflict Resolution* 58(3):481–508.
- Gentry, John A. 2002. Doomed to Fail: America’s Blind Faith in Military Technology. Technical report Defense Technical Information Center Fort Belvoir, VA: .
- Gildersleve, Patrick, Renaud Lambiotte and Taha Yasseri. 2022. “Between News and History: Identifying Networked Topics of Collective Attention on Wikipedia.”.
- Gleditsch, Kristian Skrede and Nils B. Weidmann. 2012. “Richardson in the Information Age: Geographic Information Systems and Spatial Data in International Studies.” *Annual Review of Political Science* 15(1):461–481.
- Greenstein, Shane and Feng Zhu. 2012. “Is Wikipedia Biased?” *American Economic Review* 102(3):343–348.

- Harkness, Kristen A. and Michael Hunzeker. 2015. "Military Maladaptation: Counterinsurgency and the Politics of Failure." *Journal of Strategic Studies* 38(6):777–800.
- Hart, Sir Basil Henry Liddell. 1967. *Strategy: The Indirect Approach*. London: Faber & Faber.
- Herrmann, Michael and Holger Döring. 2023. "Party Positions from Wikipedia Classifications of Party Ideology." *Political Analysis* 31(1):22–41.
- Horowitz, Michael and Dan Reiter. 2001. "When Does Aerial Bombing Work?: Quantitative Empirical Tests, 1917-1999." *The Journal of Conflict Resolution* 45(2):147–173.
- Huth, Paul K. 1998. "Major Power Intervention in International Crises, 1918-1988." *The Journal of Conflict Resolution* 42(6):744–770.
- Johnston, Patrick B. and Anoop K. Sarbahi. 2016. "The Impact of US Drone Strikes on Terrorism in Pakistan." *International Studies Quarterly* 60(2):203–219.
- Kavanagh, Jennifer, Bryan Frederick, Alexandra Stark, Nathan Chandler, Meagan Smith, Matthew Povlock, Lynn Davis and Edward Geist. 2019. Characteristics of Successful U.S. Military Interventions. Technical report RAND Corporation Santa Monica, CA: .
- Kilcullen, David and Andres McDonald Exum. 2009. "Death From Above, Outrage Down Below." *The New York Times* .
- Kollars, Nina A. 2015. "War's Horizon: Soldier-Led Adaptation in Iraq and Vietnam." *Journal of Strategic Studies* 38(4):529–553.

- Kushi, Sidita and Monica Duffy Toft. 2023. “Introducing the Military Intervention Project: A New Dataset on US Military Interventions, 1776–2019.” *Journal of Conflict Resolution* 67(4):752–779.
- Lands, David R. 2019. “Publicly Available Information: The Secret to Unclassified Data, Part I.” <https://overthehorizonmdos.wpcomstaging.com/2019/04/08/publicly-available-information-the-secret-to-unclassified-data-part-i/>.
- Luttwak, Edward N. 1981. “The Operational Level of War.” *International Security* 5(3):61–79.
- Lyall, Jason and Isaiah Wilson. 2009. “Rage Against the Machines: Explaining Outcomes in Counterinsurgency Wars.” *International Organization* 63(1):67–106.
- Malyshev, Stanislav, Markus Krötzsch, Larry González, Julius Gonsior and Adrian Bielefeldt. 2018. Getting the Most Out of Wikidata: Semantic Technology Usage in Wikipedia’s Knowledge Graph. In *The Semantic Web – ISWC 2018*, ed. Denny Vrandečić, Kalina Bontcheva, Mari Carmen Suárez-Figueroa, Valentina Presutti, Irene Celino, Marta Sabou, Lucie-Aimée Kaffee and Elena Simperl. Lecture Notes in Computer Science Cham: Springer International Publishing pp. 376–394.
- Martinez Machain, Carla. 2015. “Air Campaign Duration and the Interaction of Air and Ground Forces.” *International Interactions* 41(3):539–564.
- McMaster, H. R. 1998. *Dereliction of Duty: Johnson, McNamara, the Joint Chiefs of Staff, and the Lies That Led to Vietnam*. Harper Collins.

- Millett, Allan R. and Williamson Murray. 1988. "Lessons of War." *The National Interest* (14):83–95.
- Millett, Allan R., Williamson Murray and Kenneth H. Watman. 1986. "The Effectiveness of Military Organizations." *International Security* 11(1):37–71.
- Morgan, T. Clifton and Glenn Palmer. 2000. "A Model of Foreign Policy Substitutability: Selecting the Right Tools for the Job(s)." *Journal of Conflict Resolution* 44(1):11–32.
- Oswald, Christian and Daniel Ohrenhofer. 2022. "Click, Click Boom: Using Wikipedia Data to Predict Changes in Battle-Related Deaths." *International Interactions* 48(4):678–696.
- Pape, Robert A. 1996. *Bombing to Win: Air Power and Coercion in War*. United Kingdom: Cornell University Press.
- Pearson, Frederic S. and Robert A. Baumann. 1988. "International Military Interventions: Identification and Classification." *International Interactions* 14(2):173–180.
- Pickering, Jeffrey and Emizet F. Kisangani. 2009. "The International Military Intervention Dataset: An Updated Resource for Conflict Scholars." *Journal of Peace Research* 46(4):589–599.
- Post, Abigail. 2019. "Flying to Fail: Costly Signals and Air Power in Crisis Bargaining." *Journal of Conflict Resolution* 63(4):869–895.
- Price, Bryan C. 2012. "Targeting Top Terrorists: How Leadership Decapitation Contributes to Counterterrorism." *International Security* 36(4):9–46.

- Rector, Lucy Holman. 2008. "Comparison of Wikipedia and Other Encyclopedias for Accuracy, Breadth, and Depth in Historical Articles." *Reference Services Review* 36(1):7–22.
- Reiter, Dan and Simon A. Mettler. 2013. "Ballistic Missiles and International Conflict." *Journal of Conflict Resolution* 57(5):854–880.
- Ren, Ruqin and Jian Xu. 2023. "It's Not an Encyclopedia, It's a Market of Agendas: Decentralized Agenda Networks between Wikipedia and Global News Media from 2015 to 2020." *New Media & Society* .
- Royal, Cindy and Deepina Kapila. 2009. "What's on Wikipedia, and What's Not...?: Assessing Completeness of Information." *Social Science Computer Review* 27(1):138–148.
- Shaver, Andrew, David B. Carter and Tsering Wangyal Shawa. 2019. "Terrain Ruggedness and Land Cover: Improved Data for Most Research Designs." *Conflict Management and Peace Science* 36(2):191–218.
- Sullivan, Patricia L. and Michael T. Koch. 2009. "Military Intervention by Powerful States, 1945—2003." *Journal of Peace Research* 46(5):707–718.
- Talmadge, Caitlin. 2016. "Different Threats, Different Militaries: Explaining Organizational Practices in Authoritarian Armies." *Security Studies* 25(1):111–141.
- Tecott Metz, Rachel and Andrew Halterman. 2021. "The Case for Campaign Analysis: A Method for Studying Military Operations." *International Security* 45(4):44–83.
- Tramullas, Jesús, Piedad Garrido-Picazo and Ana I. Sánchez-Casabón. 2016. Research on

- Wikipedia Vandalism: A Brief Literature Review. In *Proceedings of the 4th Spanish Conference on Information Retrieval*. CERI '16 New York, NY, USA: ACM pp. 15:1–15:4.
- von Clausewitz, Carl. 1976. *On War*. Princeton, N.J: Princeton University Press.
- Wallace, Geoffrey P. R. 2008. “Alliances, Institutional Design, and the Determinants of Military Strategy.” *Conflict Management and Peace Science* 25(3):224–243.
- Zabecki, David T. 2006. *The German 1918 Offensives: A Case Study in The Operational Level of War*. New York: Routledge.
- Zegart, Amy. 2018. “Cheap Fights, Credible Threats: The Future of Armed Drones and Coercion.” *Journal of Strategic Studies* 43(1):6–46.