

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

J Andrés Gannon*

Kerry Chávez†

Abstract

Research on strategies and force structures 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 interventions 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 nearly every post-1989 military intervention described in existing academic datasets plus 425 additional operations. The data we introduce offers unprecedented coverage and granularity that enables analysis of myriad factors associated with how, when, and where the United States conducts military interventions. We describe the data collection process, demonstrate its contents and validity, and discuss its potential

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

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

Previous versions of this paper were presented at APSA 2021 and 2022. For feedback on earlier drafts, the authors thank Jonathan Caverley, Rex Douglass, Erik Gartzke, Nadiya Kostyuk, Kendrick Kuo, Ashley Leeds, Erik Lin-Greenberg, Sara Plana, Thomas Scherer, Ryan Shandler, 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 Department of Defense Minerva Research Initiative, the APSA Centennial Center, and the University of California – San Diego 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.

applications to existing theories about force structure design and strategy in war.

Keywords: military intervention, use of force, dataset

United States (US) President Obama, a liberal law professor critical of the War on Terror, 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 manned 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 operations using a familiar, yet under-utilized data source: Wikipedia. Existing datasets about in military interventions remain limited because collecting such information manually is intensive and oft incomplete. Aware of the stigma attached to Wikipedia, we nonetheless note that what we gain is worth the tradeoffs. By scraping the Wikiverse for pages on US military interventions since 1989, our dataset identifies almost ten times as many events during this period as all existing

datasets combined. Using Wikidata and DBpedia, the structured bedrocks of Wikipedia, we can 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 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 breadth and depth, measures of the means of force for each intervention, 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 ([Arreguin-Toft 2001](#)). Military myopia arguments ([Gentry 2002](#); [Lyll and Wilson 2009](#)), capital-rich force structure narratives ([Gartzke 2001](#); [Caverley 2009](#)), and uncertainty in pre-war estimates of costs and success ([Sullivan 2007](#)) 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 approaches within, but typically focus on the possession of these platforms rather than their use, offering insight only on the latent potential for force.¹ 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 Machain 2017), 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 of military force not as isolated dichotomous decisions, but from a menu of substitutable (or combinable) 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 structure and strategy require it to be open and well-lit.

¹Recent exceptions that focus on use include Martinez Machain (2015), Post (2019), and Gannon (2022).

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 interventions and the idea of substitutability in mind, a concern about assumptions of independence across observations crops up. If we seek to explain why an intervention featured solely unpiloted systems, perhaps 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, say, a place like Iraq, perhaps one might expect interventions nested within such a surge campaign to feature a higher frequency of ground troops. Attempting to explain force structure choices outside the strategic contexts in which they occur might lead to bias in estimating independent causal impacts. Yet existing datasets largely list interventions within the same campaign as distinct observations.

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 more disaggregated list of US military interventions 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, we contend that the breadth, depth, and granularity 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 interventions from 1989 to January 2021. Insofar as our case selection is most consistent with the exemplary Military Interventions by Major Powers (MIPS) dataset’s definition, it merits quoting. An intervention is:

“a use of armed force that involves the official deployment of at least 500 regular military personnel (ground, air, or naval) to attain immediate term political objectives 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 should be excluded” (Sullivan and Koch 2009, 3).

Like MIPS, we exclude threats, training exercises or mobilizations that demonstrate rather than apply force, ebbs and flows of personnel and materiel at permanent military bases, diplomatic cover and evacuation, and humanitarian operations. We deviate from MIPS in that we do not set a size threshold for inclusion.

We also lean on the International Military Intervention (IMI) dataset’s definition, which clarifies that events must be purposeful (Pearson and Baumann 1993; Kisangani and Pickering 2008). 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 intervention 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 intervention. We also insist that political and military objectives cannot be delinked, disqualifying nonpolitical uses of force. Table A1 in the Appendix summarizes our nearness to MIPS and IMI across definitional fine points, and our departures from the broader approaches of RAND (Kavanagh et al. 2019) and Military Intervention Project (MIP) (Kushi and Toft 2022). Focusing on definitions of intervention, we omit other important datasets that take a broader approach in this section that we later consider.

Our dataset includes some paramilitary activities. Specifically, we capture some Central Intelligence Agency (CIA) drone strikes, but not all. We defend this on two fronts. First, drones are distinct as a means of force since some are operated by the military (in combat zones) and some by the CIA (in noncombat zones). In fact, extensive criticism has been levied at the CIA program for its lack of military control and transparency. Second, every dataset on military interventions exhibits nonrandom missingness on some drone cases due to their covert nature (even those devoted solely to drone strikes). MONSTr’s coverage is appealing in that the drone strikes it includes have met some threshold of public awareness to make it onto Wikipedia. Since we expect that public constraints amid the domestic politics of war are meaningful factors in how states fight, we accept this bias. In fact, given this variation there should be scholarly pursuits identifying what criteria determines publication on Wikipedia in the first place, a question our dataset allows us to pose and test.

Similar in logic, by nature of being collectible from the internet our cases are open source. 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. The dataset presented here is a proof of concept for the validity and value of using Wikipedia to identify and structure such information. 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.² These data, like most other sources available to scholars, are unstructured and require manual review and entry.

Unlike most other sources, Wikipedia rests upon two Resource Description Format (RDF) databases: Wikidata and DBpedia ([Malyshev et al. 2018](#)). Wikidata contains 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

²Thank goodness “primordial clam chowder” didn’t win.

uniform, machine-readable format, we were able to scrape a comprehensive list of Wikipedia pages describing a US military intervention from 1989 onward using SPARQL, an RDF query language. The resulting data contained over 500 entries and dozens of columns containing structured information from Wikidata and DBpedia infoboxes. After reasonable efforts by research assistants to 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](#)).³ 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](#); [Mesgari et al. 2015](#); [Benjakob and Aviram 2018](#)). In studies of over 10,000 randomly selected articles, scholars found Wikipedia’s accuracy to be comparable to the Encyclopedia Britannica ([Blumenstock 2008](#)), the Dictionary of American History, and American National Biography Online ([Rector 2008](#)) using predictors such as page length, number of citations, external links, and edit history.⁴ 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](#)

³Please go read the entertaining opening of this short article, which we concede is better than ours.

⁴Contrary 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 ([Adler et al. 2011](#); [Tramullas, Garrido-Picazo and Sánchez-Casabón 2016](#)).

2011). Validation exercises on party ideology data from Wikipedia similarly deem that it bears correct, consistent information (Göbel and Munzert 2021; Herrmann and Döring 2021).

In addition to front-end and process audits, the authors manually checked variable values for a randomly selected sample of 6% of observations. With the exception of uncertainty about the precise end date of two observations, Wikipedia’s coding for all other values could be independently verified using external sources. Inconsistencies largely 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. On fact-based claims (such as when and where an event occurred), Wikipedia is as accurate and often more comprehensive (Rector 2008). Aware that readers might still be thinking we’re monsters, our data and code are publicly available.

Clustering - Dependencies and Levels

Existing datasets on military intervention exhibit a concerning assumption of independence across cases. All valuable, respected, and oft used—MIPS, IMI, International Crisis Behavior (ICB), Militarized Interstate Disputes (MIDs), Uppsala Conflict Data Program (UCDP) / Peace Research Institute Oslo (PRIO) —this assumption limits 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.

For each military intervention, Wikipedia identifies a “part of” variable that indicates the

broader event to which the former is a subset. The level of “parents” for each intervention 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 operational and strategic levels in which each intervention ought to be nested. 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. Since the unit of analysis is the military intervention “leaf” nodes (the most disaggregated level), it facilitates hierarchical models accounting for structural dependencies among interventions borne of the same strategic campaign and parent-intervention fixed effects.

Distinctive Outputs

Breadth of Coverage

As already acknowledged, there are several datasets that feature military interventions or specific platforms like the two excellent aerial bombing datasets by [Horowitz and Reiter \(2001\)](#) and [Allen and Machain \(2017\)](#). There is significant variation, however, in the cases they include. Of course, much of this is by design. Many were not devised to exclusively or completely cover military interventions. Some datasets take a broader approach in defining and identifying cases—international crises, militarized disputes—some narrower, examining a specific type of intervention.⁵ Nonetheless, some of the variation is due to constraints. When subset to our definition of US military interventions since 1989, the heterogeneity in coverage is stark. Only one case appears in all existing datasets—the 1991 Gulf War. Prior

⁵Some variation is also due to different publication dates. IMI’s release in 2008 should not be faulted for failing to include post-2008 military interventions.

to MIP’s recent debut (Kushi and Toft 2022), the combined coverage of all these datasets post-1989 yields only 65 unique interventions. Our dataset provides information on just under 500 US military interventions. Once appropriately nested, we offer 336 observations composed of “leaf” nodes. Even counting only leaf nodes and even confining ourselves to a quite narrowly defined family of uses of military force, our dataset considerably expands coverage during its time scope. Figure 1 compares MONSTr to other datasets to demonstrate the large variation and its improvement in coverage.

[Figure 1 about here]

Our dataset includes so many more observations partly because Wiki enables us to disaggregate the unit of analysis to a more granular level. Current datasets typically identify and code variables associated with wars, such as the 2003 Iraq War. Ours identifies military operations and codes covariates associated with each individual event, such as Operation Viking Hammer conducted within the Iraq War. This disaggregation is imperative for understanding how states fight. Hart (1967) defines military strategy as “the art of distributing and applying military means to fulfill the ends of policy” (335). US Army Field Manual 100-5 (1993) affirms that strategic objectives and military battles are harmonized by *operational* art. Consequently, we see operations as the appropriate level for analyzing military strategy. Many scholars agree. For instance, Allen and Machain (2017) 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” (549). Like going to war with the Army you have rather than the one you want, scholars do 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.

Depth of Datapoints

In addition to extending breadth by improving the precision of our unit of analysis, we were able to extract more granular variables characterizing interventions. Shown in Table 1, MONSTr stacks up well against existing datasets on objective factors. It features start and end dates for each intervention 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, a significant improvement over blunt domain codings or focus on a single type when it comes to our research questions. Most interventions include geocoded latitude and longitude coordinates, facilitating finer spatial analyses. It also identifies all state and nonstate actors active in the intervention, both intervenors and targets. Notably, MONSTr does not feature information on the aim and outcome of the military intervention. 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 1 about here]

Means of Force

To analyze military strategy, we remind that it is “the art of distributing and applying *military means* to fulfill the ends of policy” (Hart 1967, 335). 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 intervention. 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. Table 2 compares the means of force identified across existing datasets and ours. By our

definition of intervention, we do not observe any naval interventions within our time scope, and where naval vessels constitute the launching pad of cruise missiles or piloted aircraft they are coded as such. We do capture several nuanced means of land and air applications of force while other sources depict them in broad domains.

[Table 2 about here]

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 (Lyll 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). Overall, having measures of the means of force grants us great purchase in the study of military strategy.⁶ Descriptive statistics for the types of force in the dataset and other select covariates feature in the online appendix.

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

⁶The amount of missiles fired or troops deployed was of initial interest, but accurate data, especially concerning ordnance, proved challenging.

the structure of Wikipedia’s internal architecture, we identified and coded the nested relationships for each intervention. Figure 2 illustrates the high degree of interdependence among interventions in our dataset. Moving to the right disaggregates each node into the interventions that are a 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 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 interventions in MONSTr are geocoded with GPS coordinates or a local municipality indicating where the intervention occurred.⁷ 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, intervention 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 interventions in the dataset and identify one potential application – the association between urban/non-urban targets and the type of military force used. Although

⁷In 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.

merely descriptive, the type of force used in interventions in Syria, Iraq, and Afghanistan raises 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 intervention.

[Figure 4 about here]

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 interventions from these data. First, Wikipedia does not distinguish between geometry type, which matters for interventions 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.⁸ 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 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 structures 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 captures this variation in actor interaction as showcased in Figure 5. These

⁸The many good works exploring each determinant are too many to list here.

three network plots of the US and its alliance partners (side A) label state partners 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 campaign. 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 campaigns 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 events, 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 intervene using 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 intervention.⁹ 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.¹⁰ The dependent variables are binary and the data are nested, calling for a logit estimator in a hierarchical model.

We also build upon previously theorized explanators for state force structures. Past studies have examined factors such as target type, coalition effort, 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 (Arreguin-Toft 2001). Second, we include a variable indicating whether the intervention was MULTILATERAL, recognizing that cost- and burden-sharing affect the coordination of operations (Cappella Zielinski and Grauer 2022). Third, 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 might change over the life of a conflict. We include a DURATION variable and a measure cap-

⁹As described above, the analyses are conducted on only the most disaggregated “leaf” interventions, leaving covariates about the broader interventions (like the Gulf War) as model parameters.

¹⁰Since 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.

turing how many DAYS INTO THE PARENT conflict the intervention 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 interventions, it appears that those with nonstate combatants are much less likely to feature air power than interventions where there are none. This finding does not hold for unpiloted aerial platforms, however, suggesting that most of this relationship is driven by the negative association between *piloted* aerial bombing and nonstate targets. Multilateral interventions are positively associated with the use of US ground forces, affirming the idea that sharing risks, costs, and potentially blame across partners enables militaries to put more boots on the ground. Conversely, in military interventions where the US engages in aerial bombing, it tends to do so unilaterally.

[Table 3 about here]

When it comes to geography, interventions taking place in mountainous terrain are associated with fewer ground interventions and more aerial bombing, suggesting a reluctance by the US to use boots on the ground when it is insurmountable, such as in the infamous 72-hour bombing campaign of the Tora Bora mountain caves targeting Osama bin Laden in late 2001. Urban targets are associated with a lower frequency of unpiloted aerial attacks, which runs counter to theories that their 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 interventions are associated with an increased frequency of piloted air power, yet the deeper it is into a broader parent campaign the more likely it will feature ground forces. These temporal variations warrant further consideration and theorizing. An interesting null finding is that none of these factors appears to impact combined air and ground forces. [Martinez Machain \(2015\)](#) demonstrates that ground troops affect the duration of aerial campaigns divergently depending on how discriminate the latter’s approach. The shorter durations from less selective campaigns intermixed with the longer durations from more discriminate ones might wash out significant findings looking along the opposite causal arrow. Overall, the results mostly affirm conventional wisdom, but also point to new puzzles that the dataset will enable scholars to examine.

Conclusion and Future Directions

With the advent of automated extraction, and the audacity to apply it to a well-known yet long-taboo source, we were able to collect a first cut of data in a relatively brief, low-labor project. The Wiki-based MONSTr dataset features objective correlates of nearly all US military interventions from 1989 to January 2021 in a nested structure reflecting observation interdependencies. Its breadth is more expansive, its depth more granular, 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 intervention traits. Many findings accord with existing theory – like the use of ground troops in multilateral operations further into the conflict and 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. Theo-

retically, we wonder about the determinants of force structures, both retesting conventional wisdom and expanding it. In particular, we expect that domestic public factors constrain and compel the force structures 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 how operations change across the life of a parent conflict. Our demonstrative models suggest that the means of force shift the longer the intervention and the deeper into the parent campaign. Conversely, scholars can examine the impact of force structures 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 structure 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 Python, 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 inhibits inquiry into the full 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 assumption of independence across observations in existing datasets separates interventions 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” ([Stoker 1992](#)).

Table 1: Comparison of Intervention-level Variables

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 2: Comparison of Datasets Identifying Military Strategy

	<i>MONSTR</i>	<i>MIPS</i>	<i>IMI</i>	<i>MIP</i>	<i>Allen</i>	<i>Horowitz & Reiter</i>
Sea						
All naval	-	-	9	34	-	-
Land						
All land	-	7	-	60	3	-
Ground troops	174	-	-	-	-	-
Paramilitary	92	-	-	-	-	-
Air						
All air	-	3	13	34	4	5
Aerial bombing	149	-	-	-	-	-
Air-to-air	7	-	-	-	-	-
Close air support	110	-	-	-	-	-
Cruise missiles	16	-	-	-	-	-
Drones	31	-	-	-	-	-

Note: For direct comparison across datasets, observation counts refer only to US military interventions post-1989.

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

	Ground	Ground & Air	Air (All)	Air (Piloted)	Air (Unpiloted)
Nonstate Target	0.803 (0.192)	−0.800 (0.100)	−1.451** (0.036)	−1.240** (0.026)	−0.052 (0.947)
Multilateral	0.618* (0.079)	0.114 (0.696)	−0.654* (0.057)	−0.171 (0.573)	−0.613 (0.181)
Mountainous	−0.891* (0.059)	0.293 (0.333)	0.833* (0.082)	0.479 (0.225)	0.386 (0.535)
Urban	−0.017 (0.964)	0.060 (0.841)	0.276 (0.447)	0.379 (0.245)	−1.100* (0.061)
Duration (log)	0.091 (0.292)	0.117 (0.102)	0.071 (0.409)	0.157** (0.042)	0.007 (0.948)
Day into Parent (log)	0.249* (0.081)	0.122 (0.283)	−0.142 (0.359)	−0.098 (0.417)	−0.074 (0.688)
<i>N</i>	326	326	326	326	326
<i>R</i> ² (marg)	0.261	0.137	0.225	0.149	0.175
AIC	339.6	424.0	367.9	411.1	224.8
BIC	396.4	480.8	424.7	467.9	281.6
RMSE	0.38	0.46	0.39	0.43	0.27

Note: All models include president fixed effects and year cubic polynomials.

* p<0.10, ** p<0.05, *** p<0.01

Figure 1: Comparison of Dataset Coverage

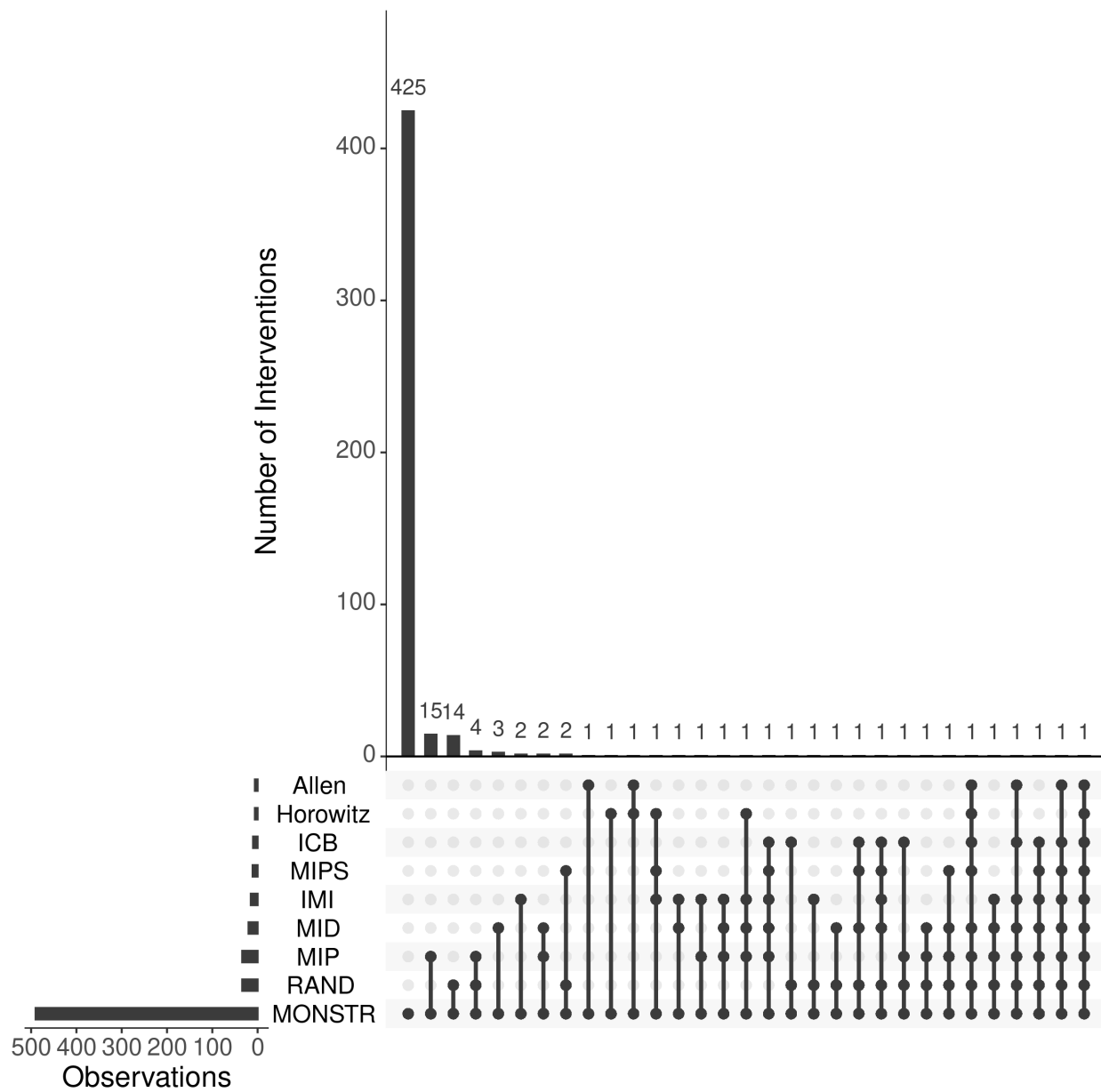


Figure 2: Nested Dependencies of US Military Interventions

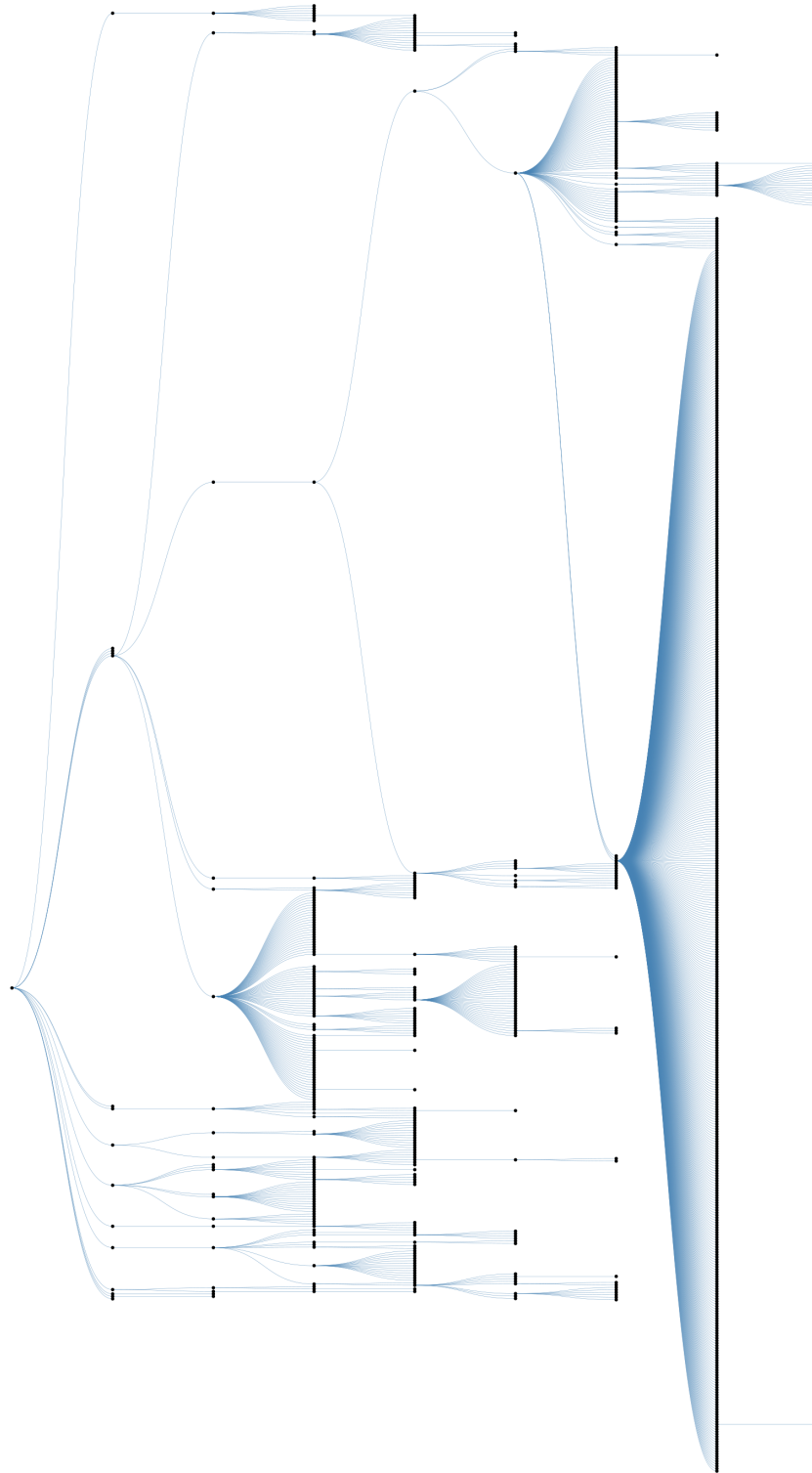


Figure 3: Nested Structure of the Gulf War (1991)

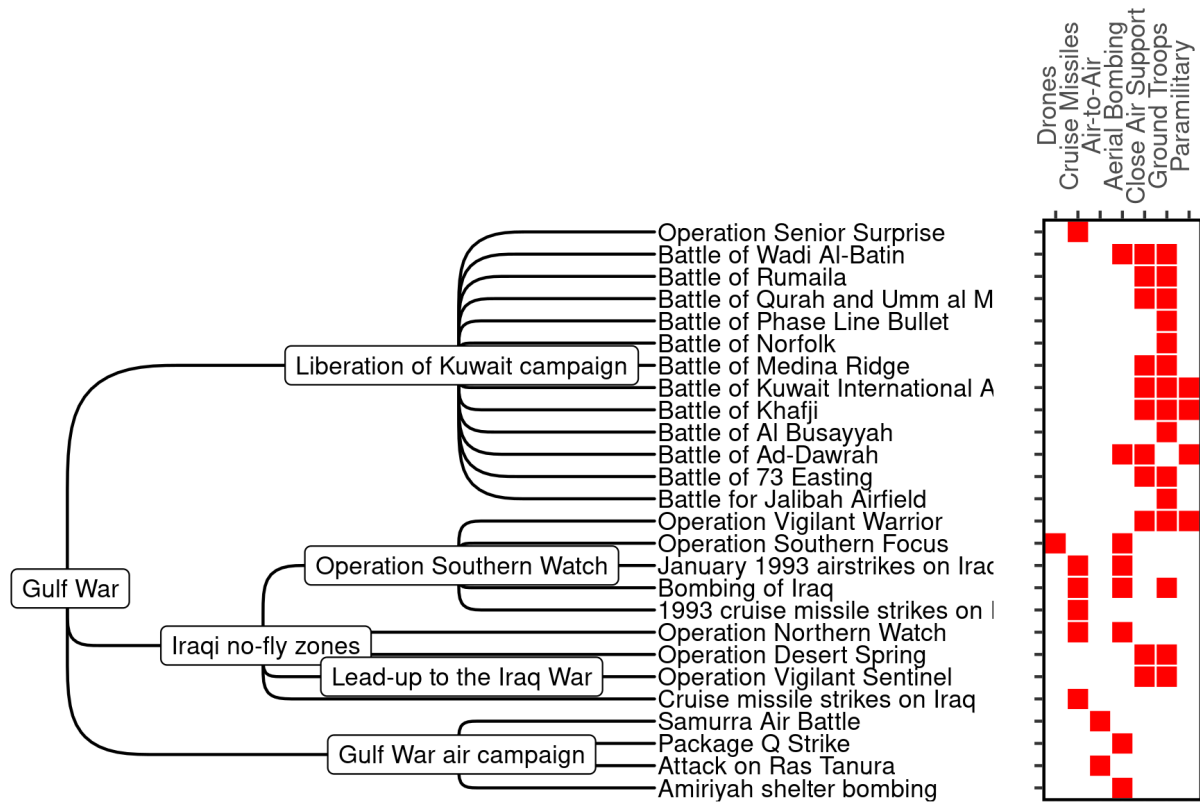


Figure 4: Locations of US Military Interventions (1989-2021)

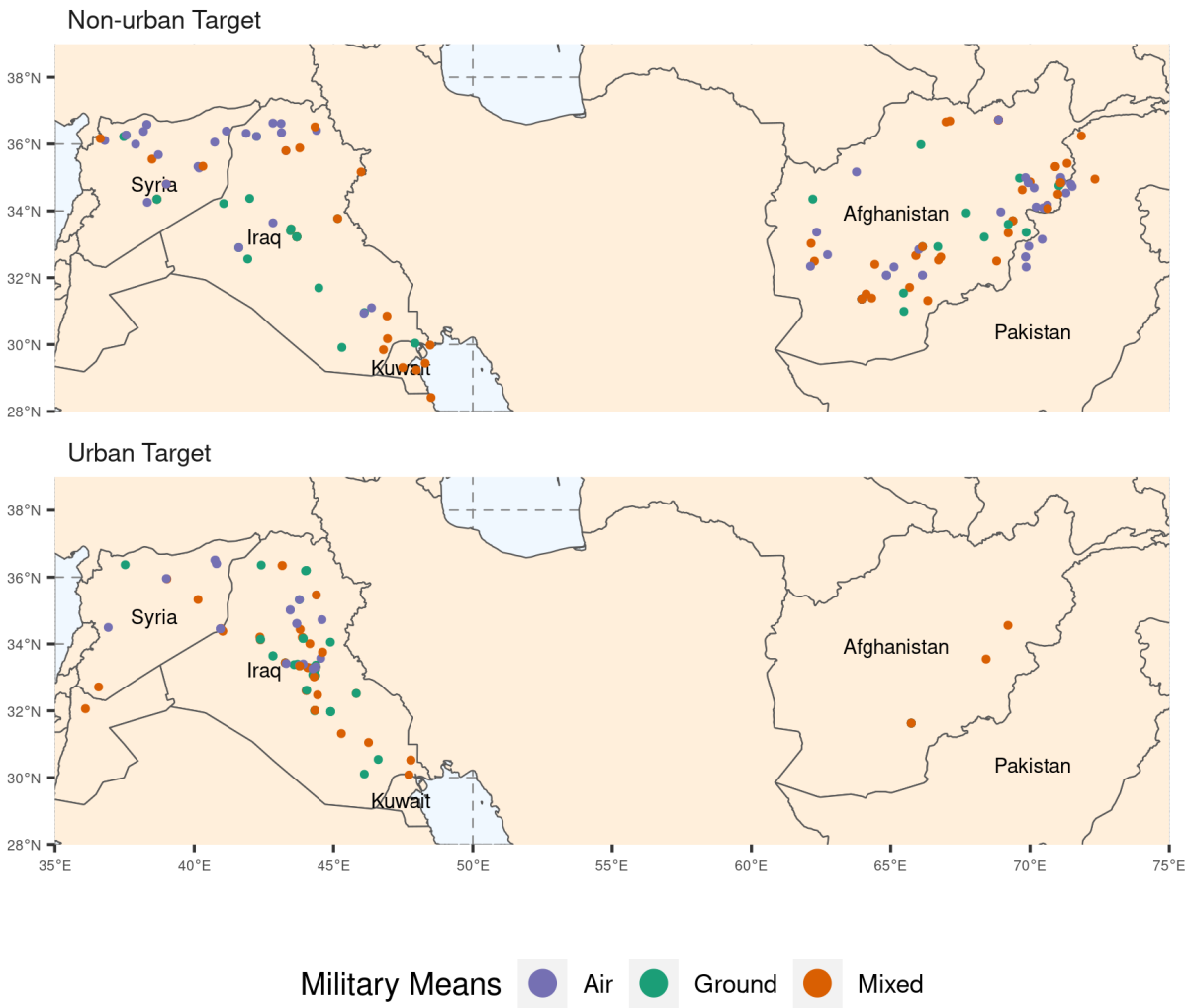
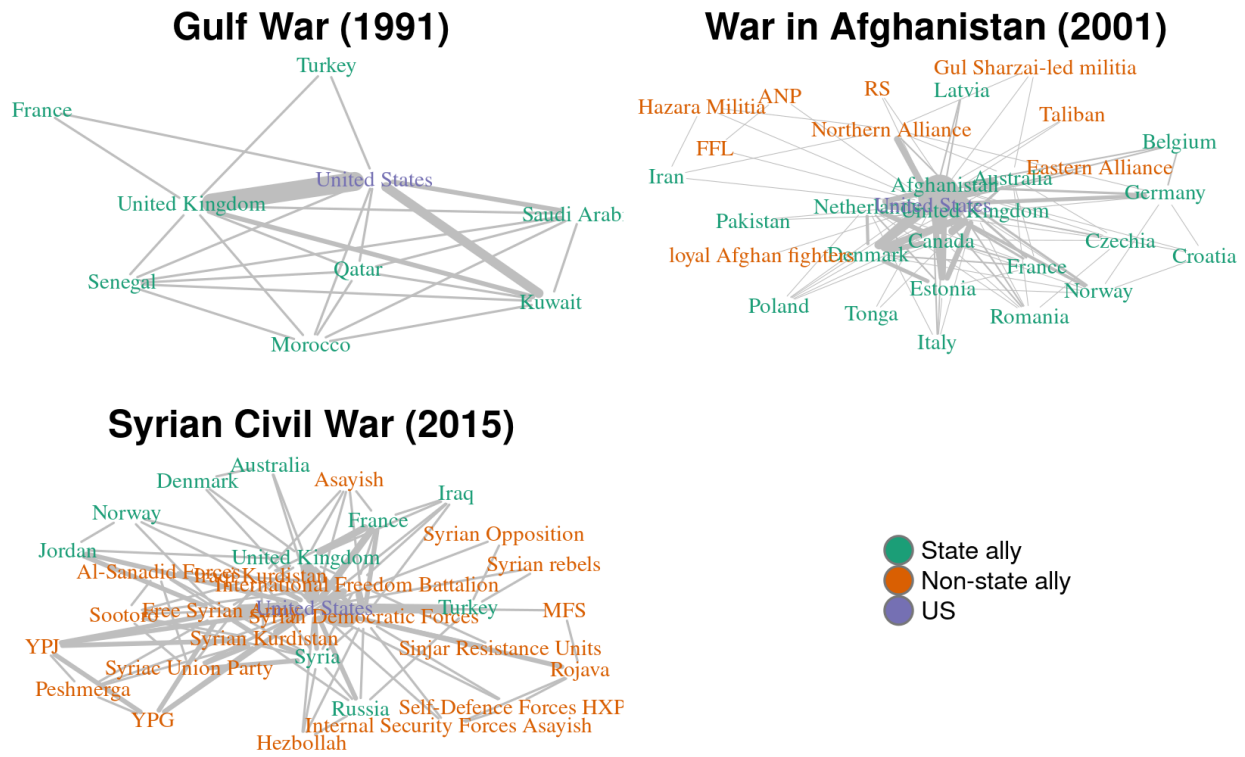


Figure 5: Social Network Plots of Select Conflicts



References

- Adler, B Thomas, Luca de Alfaro, Santiago M Mola-Velasco, Pablo Rosso and Andrew G West. 2011. Wikipedia Vandalism Detection: Combining Natural Language, Metadata, and Reputation Features. In *Computational Linguistics and Intelligent Text Processing*. Lecture Notes in Computer Science Berlin, Germany: Springer pp. 277–288.
- Allen, Susan H and Carla Martinez Machain. 2017. “Understanding the Impact of Air Power.” *Conflict Management and Peace Science* 36(5):545–558.
- Arreguin-Toft, Ivan. 2001. “How the Weak Win Wars: A Theory of Asymmetric Conflict.” *International Security* 26(2):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.” New York Times. <https://www.nytimes.com/2012/05/29/world/obamas-leadership-in-war-on-al-qaeda.html>.
- Benjakob, Omer and Rona Aviram. 2018. “A Clockwork Wikipedia: From a Broad Perspective to a Case Study.” *Journal of Biological Rhythms* 33(3):233–244.
- Blumenstock, Joshua E. 2008. “Automatically Assessing the Quality of Wikipedia Articles.” *UC Berkeley* p. published online.

- 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.
- 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." *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*. New York, NY: Cambridge University Press.
- Cooper, Scott. 2001. "The Politics of Airstrikes." *Policy Review* 107:55–66.
- Cranmer, Skyler and Elizabeth Menninga. 2018. "Coalition Quality and Multinational Dispute Outcomes." *International Interactions* 44(2):217–243.
- Crisher, Brian Benjamin and Mark Souva. 2014. "Power at Sea: A Naval Power Dataset, 1865-2011." *International Interactions* 40(4):602–629.
- Douglass, Rex W and Kristen Harkness. 2018. "Measuring the Landscape of Civil War: Eval-

- uating Geographic Coding Decisions with Historic Data from the Mau Mau Rebellion.” *Journal of Peace Research* 55(2):190–205.
- 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. 2022. “One If by Land, and Two If by Sea: Cross-domain Contests and the Escalation of International Crises.” *International Studies Quarterly* 66(4):published online.
- 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. 2001. “Democracy and the Preparation for War: Does Regime Type Affect States’ Anticipation of Casualties?” *International Studies Quarterly* 45(3):467–484.
- Gartzke, Erik, 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.” *Parameters* 32(4):88–103.

- 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.
- Göbel, Sascha and Simon Munzert. 2021. "The Comparative Legislators Database." *British Journal of Political Science* 52(3):1398–1408.
- Hart, Sir Basil Henry Liddell. 1967. *Strategy: The Indirect Approach*. London, UK: Faber & Faber.
- Herrmann, Michael and Holger Döring. 2021. "Party Positions from Wikipedia Classifications of Party Ideology." *Political Analysis* p. published online.
- Horowitz, Michael and Dan Reiter. 2001. "When Does Aerial Bombing Work?: Quantitative Empirical Tests, 1917-1999." *Journal of Conflict Resolution* 45(2):147–173.
- Huth, Paul K. 1998. "Major Power Intervention in International Crises, 1918-1988." *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 Andrew McDonald Exum. 2009. “Death From Above, Outrage From Below.” New York Times. <https://www.nytimes.com/2009/05/17/opinion/17exum.html>.
- Kisangani, Emizet F and Jeffrey Pickering. 2008. *International Military Intervention, 1989-2005*. University of Michigan, Ann Arbor: University Consortium for Political and Social Research, Data Collection No 21282.
- Kushi, Sidita and Monica Duffy Toft. 2022. “Introducing the Military Intervention Project: A New Dataset on US Military Interventions, 1776-2019.” *Journal of Conflict Resolution* p. published online.
- Lands, David R. 2019. “Publicly Available Information: The Secret to Unclassified Data, Part I.” *Over the Horizon Journal* p. published online.
- 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 Switzerland: 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.
- Mesgari, Mostafa, Chitu Okoli, Mohamad Mehdi, Finn Årup Nielsen and Arto Lanamäki. 2015. ““The Sum of All Human Knowledge”: A Systematic Review of Scholarly Research on the Content of Wikipedia.” *Journal of the Association for Information Science and Technology* 66(2):219–245.
- 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.
- Pape, Robert A. 1996. *Bombing to Win: Air Power and Coercion in War*. Ithaca, NY: Cornell University Press.
- Pearson, Frederic S and Robert A Baumann. 1993. *International Military Intervention, 1946-1988*. University of Michigan, Ann Arbor: University Consortium for Political and Social Research, Data Collection No 6035.
- 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):7–2.
- Reiter, Dan and Simon A Mettler. 2013. "Ballistic Missiles and International Conflict." *Journal of Conflict Resolution* 57(5):854–880.
- 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.
- Stoker, Bram. 1992. *Dracula*. Ware, England: Wordsworth Editions.
- Sullivan, Patricia L. 2007. "War Aims and War Outcomes: Why Powerful States Lose Limited Wars." *Journal of Conflict Resolution* 51(3):496–524.
- 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.
- 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: ACM pp. 15:1–15:4.
- Wallace, Geoffrey PR. 2008. "Alliances, Institutional Design, and the Determinants of Military Strategy." *Conflict Management and Peace Science* 25(3):224–243.

Zegart, Amy. 2018. "Cheap Fights, Credible Threats: The Future of Armed Drones and Coercion." *Journal of Strategic Studies* 43(1):6–46.

Appendix

Table A1: Definitional Boundaries Across Datasets Capturing Military Interventions

Dataset → Criterion ↓	<i>MONSTr</i>	<i>MIPS</i>	<i>IMI</i>	<i>MIP</i>	<i>RAND</i>
<i>Political</i>	✓	✓	✓	✗	✗
<i>Size Threshold</i>	none	500	none	none	1000
<i>Target</i>	any foreign	any foreign	any foreign	state actor	any foreign
<i>Threats</i>	✗	✗	✗	✓	✓
<i>Training Exercises</i>	✗	✗	✗	✓	✓
<i>Troop Movement / Mobilization</i>	✗	✗	✗	✓	✓
<i>Diplomatic Cover / Evacuation</i>	✗	✗	✓	✓	✓
<i>Humanitarian Force</i>	✗	✗	✓	✓	✓
<i>Unintentional Contingencies</i>	✗	✗	✗	✓	✓
<i>Airspace Violations</i>	✗	✗	✗	✓	✓
<i>Sanctions Enforcement</i>	✗	✗	✗	✓	✓
<i>Irregular Troops</i>	*	✗	✗	✓	✓

Table 2: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Drones	336	0.09	0.28	0	1
Air to air	336	0.02	0.15	0	1
Cruise missiles	336	0.04	0.20	0	1
Aerial bombing	336	0.43	0.50	0	1
Close air support	336	0.32	0.47	0	1
Ground troops	336	0.49	0.50	0	1
Paramilitary	336	0.26	0.44	0	1
Duration (days)	333	50.67	181.76	0	2,151
Days into parent (days)	336	1,625.26	1,475.81	-4	9,596
Ally count	336	1.57	2.35	0	32
Adversary count (state)	336	0.32	0.63	0	5
Adversary count (non-state)	336	1.49	2.52	0	39

All military means in the original dataset shown.