Introducing the Joint Military Exercises Data Set

Abstract

The number of joint military exercises substantially increased over the last 30 years and they now form a major part of peacetime defense cooperation. Academic work that incorporates joint military exercises, however, has been largely limited to qualitative studies due to a lack of reliable, cross national time series data. This paper introduces the Joint Military Exercises Data Set to solve this problem. I describe joint military exercises and place them in context of the growing body of work looking at non-formal alliance forms of international cooperation and signaling and I describe the data collection process. Finally, I validate the usefulness of looking at joint military exercises by showing that dyads that exercise together are associated with a decreased likelihood of engaging in militarized disputes against each other, an increased likelihood of being on the same side of a dispute, and an increase in arms sales and transfers.

Word Count: 10,700

1 Introduction

Over the last 30 years, there has been a substantial increase in the number of joint military exercises[1] taking place around the world. There has also been a substantial increase in the number of countries participating in joint military exercises. This increase can be seen in almost every region of the world. Yet, until recently, there has been little scholarship looking at either the determinants of joint military exercises or their effect on other outcomes.

The following patterns, based on the data set introduced in this paper, attest to this development. Figure 1 shows the number of joint military exercises taking place in each region of the world, separated by the type of exercise. Since 1990, every region except South American has seen a substantial increase. While the majority of exercises practice combat activities in field training conditions, there has also been an increase in other types of exercises. Exercises focusing on non-combat activities such as delivering humanitarian assistance, unheard of before 1990, have proliferated.

Every great power has increased its exercise participation over time (see Figure 2). While the Soviet Union/Russia briefly decreased its exercise activity following the end of the Cold War, it now conducts many more exercises than during the 1980s. China only started widespread exercise activity in the mid-2000s, but now conducts dozens per year. The United States and its allies conduct more than the other great powers, and the NATO allies have conducted more and more multilateral exercises. Even the number of exercises in which no major power participates has been increasing.

This increase in joint military exercises is also marked by an increase in the number of participants that exercise together but do not share a formal military alliance. Figure 2 shows that in recent years a majority of bilateral exercises were between non-allies and that substantial fraction of multilateral exercises included at least one non-allied bilateral

1. I use the term *joint* military exercises to refer to multinational exercises. United States military doctrine uses the term *joint* to refer to exercises involving multiple service branches and instead uses the term *combined* to refer to multinational exercises. I prefer the term *joint* over *combined* because it is used much more frequently in both popular reporting and in the academic literature.

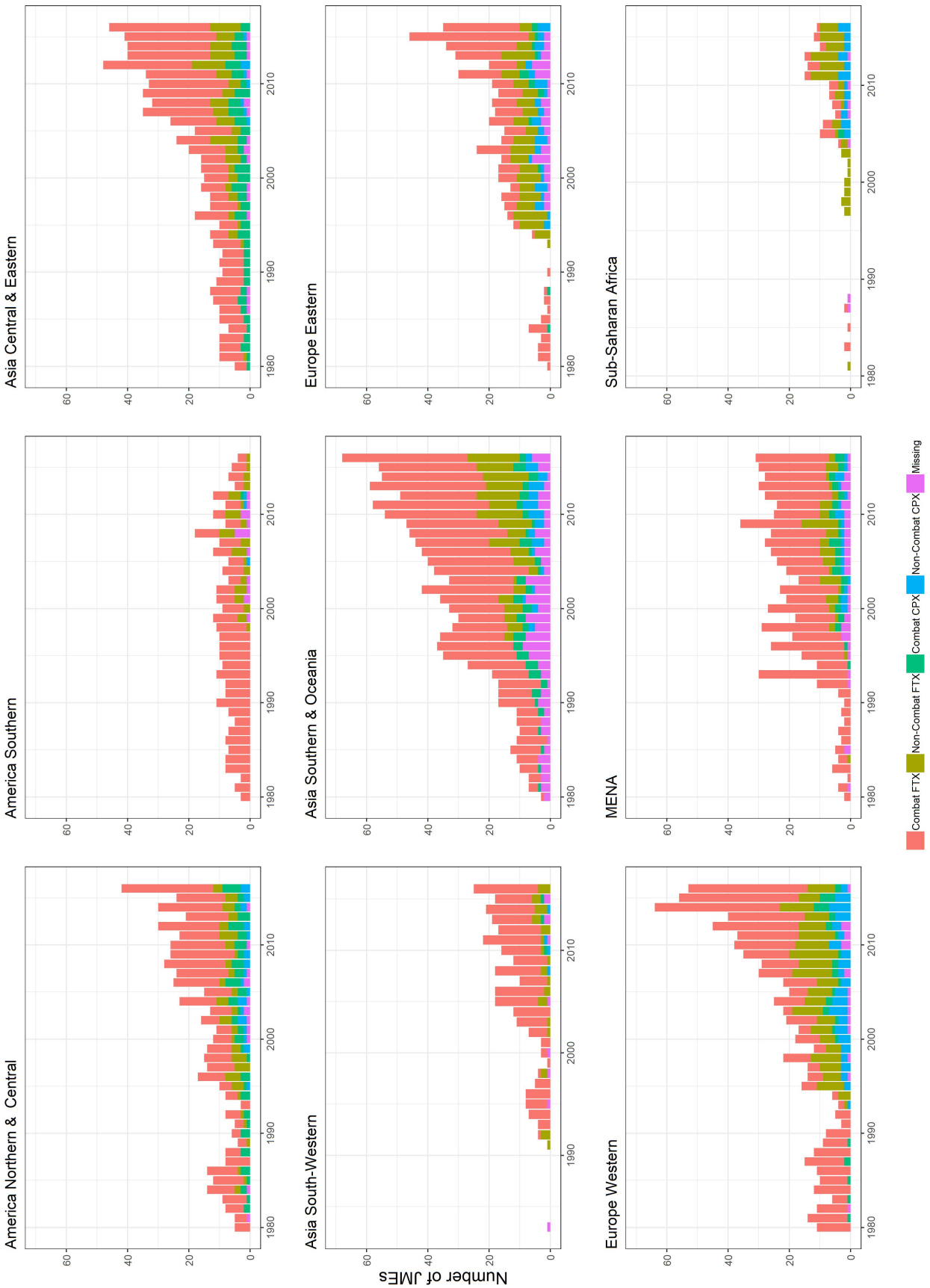

Figure 1: Joint Military Exercises Growth By Region Hosting Exercise and Exercise Type

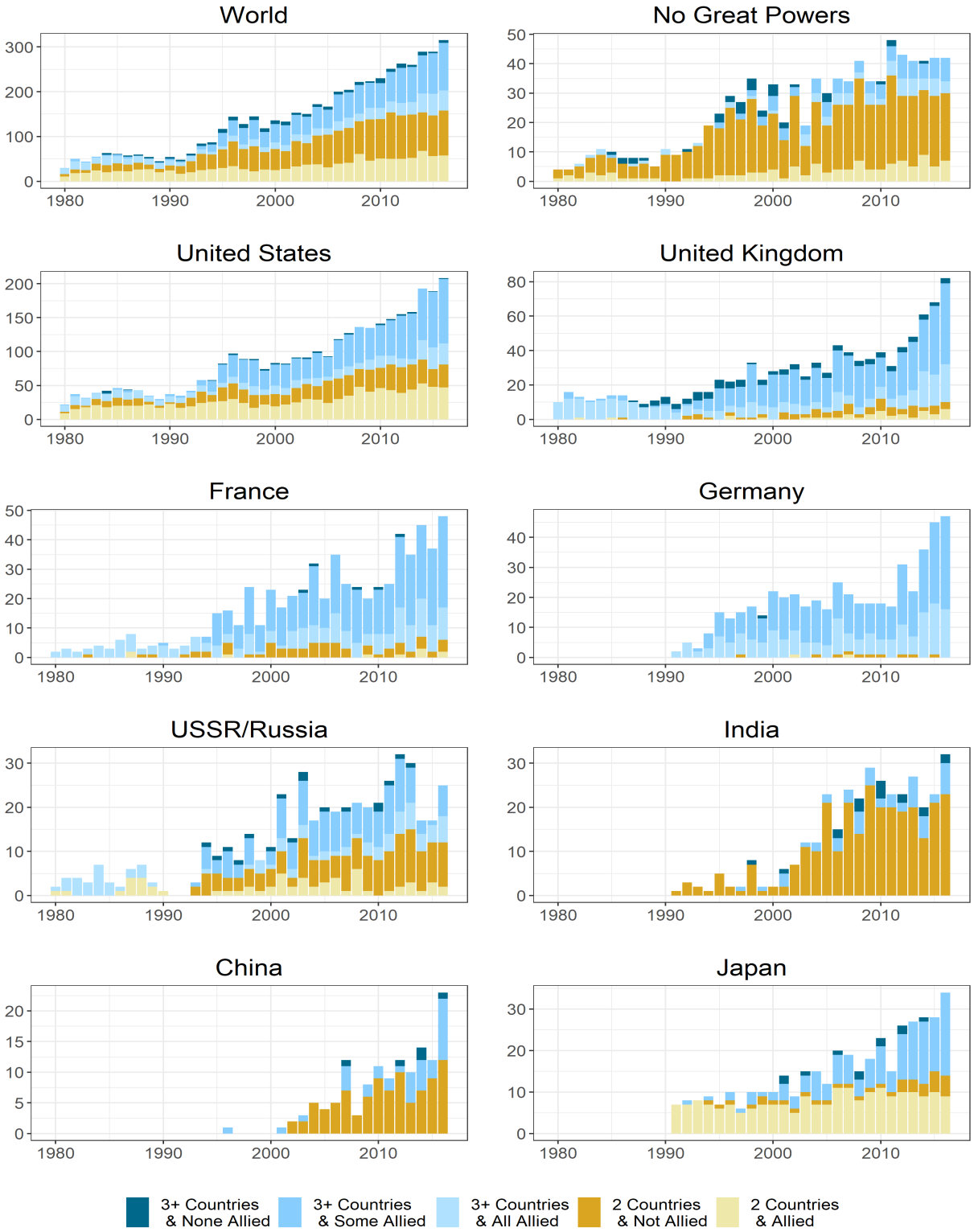

Figure 2: Joint military exercises per year by number and alliance status of participating countries. Germany, India, and Japan become great powers in 1992.

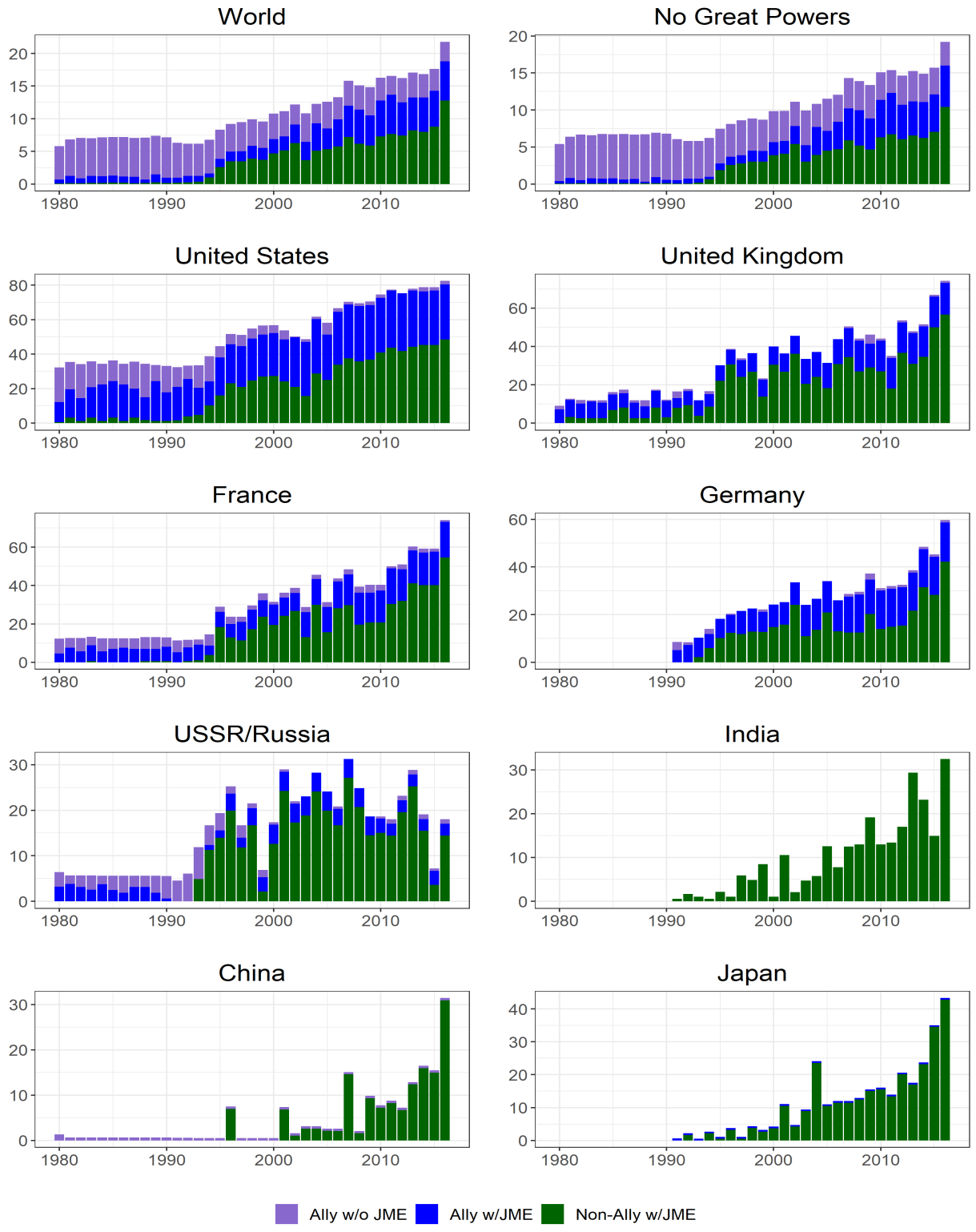

Figure 3: Percent of dyads participating in 1+ joint military exercise per year by dyad alliance status. Germany, India, and Japan become great powers in 1992.

relationship. Figure 3 shows that an increasing percentage of dyadic relationships conduct at least one exercise together each year. It also shows that much of the growth in the percentage of dyads conducting exercises together is concentrated in non-allied dyads.

Both the growth of joint military exercises and their absence from study are surprising. There has not been a substantial increase in the number of wars or disputes during this time period that would lead countries fighting with others to increase their cooperation. Nor has there been an increase in formal alliance treaties that could have led countries to initiate new joint military exercises. Yet the ubiquity of joint military exercises suggests that countries find them useful. However, there has not previously been high-quality data on joint military exercise activity and what data exists has been drawn from limited samples (Blackwill and Legro 1989; Caravelli 1983; Selden 2013) or based on data that is troublingly flawed (Blankenship and Kuo 2017; D'Orazio 2012; Kinne 2018; McManus and Nieman 2019; McManus and Yarhi-Milo 2017; Wolfley 2018, 2021).

In this paper, I introduce the joint military exercises data set, which includes information on over 5,000 joint military exercises between 1980 and 2016 and that involved 180 countries. The data set includes information about the exercise participants, the dates on which the exercise occurred, the location of the exercise, the method of training used, the type of training conducted, and estimates for the number of personnel and some types of equipment employed in the exercise. The next section of the paper defines joint military exercises and place them in the context of the growing body of work looking at forms of international military cooperation and signaling that take place outside the context of formal alliances. I then describe that data collection process, which involved using machine learning processes on publicly available news articles to identify information about exercises, as well as discussing both its advantage relative to prior research and its limitations and possible biases. Subsequently, I validate the data by showing that participating in joint military exercises is associated with a dyad being less likely to engage in militarized interstate disputes, more likely to join an existing dispute on the side of one's exercise partner, and engaging in more

arms sales and transfers. The final section concludes.

2 What Are Joint Military Exercises?

To understand what is meant by joint military exercises, it it is worth considering the full range of training activities in which militaries engage. We can then see how joint military exercises are a subset of this larger class. When recruits join, they are put through basic training that is designed to give them rudimentary skills and inculcate military values and discipline. They usually acquire a specialty that requires additional training. Officers undergo specialized training to prepare them to lead others and refine their skills and knowledge.

Training does not stop after being assigned to units. For example, in the United States Army, soldiers in brigade combat teams go through a suite of training workups as they prepare for their deployment windows.[2] After ensuring they are qualified on basic skills and techniques, they begin to participate in training operations involving small organizational units such as at the platoon level and then work their way up to conducting company level training exercises against simulated opposition provided by personnel from other units. Then they progress to larger exercises at the battalion level and then they often deploy to a Combat Training Center for a brigade level operational exercises. The Army has one of its centers in Hohenfels, Germany, and it regularly hosts training exercises that include non-United States personnel.

The content of training activities varies based on the size of the unit or units involved and the types of missions for which the unit is preparing. Units participating in Combat Training Center exercises tend to be preparing for specific deployments and therefore engage in combat exercises tailored to the theater to which they will be sent. For example, units preparing to deploy to Afghanistan tend to engage in exercises that mimic counterinsurgency

2. For information on United State Army training refer to Headquarters, Department of the Army (2010, 2016, 2017, 2019).

operations. The exercises opponents study the techniques of insurgents, and the Army builds environments that simulate Afghanistan. If, on the other hand, a unit is preparing to deploy to Europe, it is likely to engage in an exercise that simulates conventional combined arms operations and the opposing force is likely to have studied Russian doctrine. Throughout the training cycle, units will of course work on a variety of scenarios and will build training exercises for multiple contingencies.

The form of an exercise can generally be broken down on two dimensions. First, exercises can consist primarily of entirely simulated activities. They can employ only a small number of personnel at a single command level or they could employ thousands or tens of thousands of personnel across a number of locations and command levels. The primary distinguishing characteristic is that they use computer or other types of simulations and do not involve the use of large equipment such as armored vehicles or combat aircraft in the field. Because this type of exercise often happens at headquarters, I refer to these types of exercises as *command post exercises*. Second, exercises can employ personnel to directly practice carrying out a mission. This could involve small units practicing activities like a platoon managing a riot or whole divisions or corps engaged in massive maneuvers. The distinguishing characteristic is that the personnel are employed in the field with the equipment they would need to carry out the mission. These exercises could occur on a military base or even outside of a base, but it must involve employing equipment in the field and thus I term them *field training exercises*.

The missions for which an exercises trains also vary. Some exercises train *combat* operations. For ground forces, these could include the elements of combined arms maneuvers such as mounted and dismounted infantry and armor operations, artillery operations, amphibious landing operations, airborne operations, air defense, combat engineering, or even counterinsurgency operations. For air forces this can include both air-to-air combat and bombing operations. For naval forces, scenarios such as surface and (anti-) subsurface warfare, air defense, or de-mining in a combat environment would count. The unifying theme is that

these exercises prepare a unit for its primary warfighting mission.

Other exercises prepare for *non-combat* operations. A frequent scenario in this type of exercise involves preparing to respond to a humanitarian crisis or a natural disaster. Naval and air forces often conduct search and rescue exercises as well. Training for peacekeeping missions or law-enforcement-type activities such as interdiction or anti-piracy operations is another form of non-combat exercise. Finally, some exercises focus on communications and logistics without a broader scenario and thus would also count as non-combat activity.

The most well-known exercises, such as the *Autumn Forge* and *REFORGER* exercises by NATO countries during the Cold War or the Russian-led *Zapad* exercises tend to be field training exercises that involve training for combat operations.[3] These exercises employ thousands of individuals and their equipment, include operations over a wide geographic area, and usually involve the use of live ammunition. By their nature, these are highly visible exercises.

Other exercises might be less visible. The *Viking* series of exercises were command post exercises often involving thousands of personnel in command post settings using a peacekeeping scenario (Folke Bernadotte Academy 2017). The *Combined Endeavor* series of exercises has at times brought together over 2,000 personnel for a communications focused command post exercise (Lyle 2014). The *Tiger Balm* field training exercises between Singapore and the United States involve air combat scenarios and employ live ammunition, but tend to include fewer than 400 personnel. These exercises would be easily detectable by interested parties but may not attract as much attention from the media (Limon 2016). Because they only practice search and rescue operations and tend to involve fewer than 100 people, the *Baltic Bikini* field training exercises between Estonia, Latvia, and Lithuania are even less visible (The Baltic Times 2018). Smaller command post exercises such as the *Eternity* exercises between Azerbaijan, Georgia, and Turkey that focus on peacekeeping and law enforcement scenarios might be hard to detect if not for the participating nations publicizing their activity

3. Blackwill and Legro (1989) provide a good overview of the *Autumn Forge* and *REFORGER* series. D. Johnson (2017) provides an overview of recent *Zapad* exercises.

(Ministry of Defence of Georgia 2015).

To be included as a joint military exercises, an event requires four simultaneous characteristics. First, the event must include the personnel from at least two different countries in active roles. It cannot be an event in which one country participates and others merely observe. Second, the event must train the personnel of all of the participating countries. This means the event must include on all sides units with qualified personnel and not trainees that have yet to pass through basic qualifications. Third, the event must practice military operations, although these need not be combat operations. Fourth, the event must be practice or training, it cannot be an active operation or in direct support of an active operation.

Not all exercises are joint military exercises. Training with forces from other countries is usually not necessary to prepare a unit for deployment. Training in a foreign country may not count as an exercise either. The United States military stations numerous forces in foreign countries that conduct their own exercises without the participation of the host nation. Other examples include Singapore conducting training exercises in the United States or Australia but without the participation of United States or Australian military personnel (Government of Australia Department of Defense, n.d.; Embassy of the United States of America Singapore 2019). Some unilateral training exercises are observed by foreign military attachés or other international dignitaries, but that too would not make the event a joint military exercise.

Nor are all interactions between military forces from different countries joint military exercises. Militaries engage in a variety of interactions that would fall short of counting as an exercise. Militaries engage in routine joint planning or conduct joint operations. Militaries also have exchange programs to expose the personnel from partner nations to their own doctrine. One of the most prominent of these programs is the United States' International Military Education and Training (IMET) program. There also many programs that fall under banners such as Building Partner Capacity or Security Force Assistance that seek to build up the basic capabilities of a partner nation. These activities can involve training, but

they are distinguished from joint military exercises in that one side is furnishing the training to a lower-level unit rather than engaging in its own training. Often this activity involves the provision of basic military training.[4]

3 Putting Joint Military Exercises in Context

While there are extensive literatures on military alliances, coalitions, and military assistance, there is little existing work that looks specifically at joint military exercises. This section seeks to situate joint military exercises into these literatures. I discuss how alliances are thought to influence deterrence and military operations and how joint military exercises might contribute to these outcomes. Countries also transfer military equipment or provide basic training for others, and joint military exercises could be another form of such assistance. I review the determinants of these other forms of assistance as well as whether they provide the hoped for benefits and whether joint military exercises might share similar causes or effects.

3.1 Joint Military Exercises, Formal Alliances, and Deterrence

How countries deter aggression against themselves and others has been one of the central questions of international relations. Of particular interest when examining joint military exercises is the work on extended deterrence, in which a state attempts to dissuade aggression against a third party. Most previous work on how countries can credibly signal their willingness to fight in extended deterrence situations has focused on formal alliances. Smith (1995, 1998) argues that it is costly to sign an alliance agreement and then to subsequently not honor the commitment. While Smith models formal alliances, he notes that, in theory, any action a state could take that would increase the costs of abandonment would be sufficient, which might include joint military exercises. Morrow (1994, 2000) argues that alliances also

4. On different forms of military-to-military cooperation, see Joint Chiefs of Staff (2017).

operate through two other channels. If alliances are costly to maintain, then only states that have close enough interests to make fighting together worthwhile would sign them. In addition, if states that have formed an alliance prior to a crisis are better at fighting together than *ad hoc* coalitions of the same states that form during a crisis (Weitsman 2013), an alliance might shift the military balance sufficiently to maintain general deterrence. Preparation can increase military capabilities and make threats and promises more credible (Slantchev 2005).

From a signalling perspective, joint military exercises should primarily be considered a type of sunk cost signal (Fearon 1997).[5] Carrying out an exercise requires states to mobilize personnel and resources before a disagreement escalates to a crisis (general deterrence failure) or a crisis escalates to war (immediate deterrence failure).[6] While the literature on sunk cost signaling has not explicitly evaluated joint military exercises, the same sorts of incentives that lead to other forms of sunk cost signaling ought to apply.

Even states that can generate audience costs and thus engage in tying hands signaling might turn to joint military exercises if such exercises increase their combined military effectiveness. A state engaged in general deterrence ought to be less likely to be challenged when the military balance is more favorable.[7] Once a crisis begins, a joint military exercise that shifts the balance of power in the defenders' favor or that reveals favorable new information about the defenders' military strength ought to also make the defenders' threat more likely to succeed (Huth 1988; Huth and Russett 1984, 1988). If joint military exercises can build joint military capabilities or reveal private information about such capabilities, we would expect states engaged in both general and immediate extended deterrence to employ them.

5. While a joint military exercise might also incorporate a tying hands element if the decision to hold the exercise brings the leader's credibility or the nation's reputation into play, it will always include a sunk cost element because the costs of the exercise are borne regardless of whether the leader follows through or backs down in the end.

6. States have carried out joint exercises during wartime, e.g. the series of exercises before the Normandy invasion in World War II (Yung 2006), but in this case, the purpose seems to be less one of costly signalling and more of improving joint military effectiveness.

7. See Fearon (1994). Technically this claim depends on the interests of the defender. A highly capable defender is more likely to be challenged on a peripheral issue and a weak defender is more likely to be challenged on a core issue. But conditional on the value the defender places on the issue, stronger defenders are more likely to succeed at general deterrence.

There is some evidence that states engaged in general extended deterrence are more likely than other states to employ joint military exercises, at least with respect to exercises that focus on conventional combat scenarios. During the Cold War, after the United States transferred forces from Germany to Vietnam, it began conducting the *REFORGER* series of fall exercises with NATO allies. These exercises involved the large-scale transfer of forces from the continental United States to their fighting positions in Europe and they would participate in joint military exercises with the forces with which they linked up in Europe. Some years also involved significant British reinforcement operations (Blackwill and Legro 1989). Soviet-led joint exercises with Warsaw Pact members often served primarily to cow domestic opposition in it's satellites, but doing so also increased the credibility of external-facing threats (Blackwill and Legro 1989; Caravelli 1983).

The post-Cold War period has also seen exercises used apparently as a tool for enhancing deterrence. Bernhardt and Sukin (2020) suggests that South Korean joint military exercises could be sending signals of its resolve and military capabilities to North Korea. Selden (2013) points to joint military exercises as one way that states threatened by an emerging Russia or China have tried to grow closer to the United States. These threatened states tried to demonstrate their commitment to alliances with the United States, contributed to United States-led coalitions, and/or demonstrate their usefulness to the United States. As part of this last element, states have bound their procurement and training strategies up with the United States, including by conducting joint military exercises. Although Selden does not discuss it, Russia and China also conduct joint military exercises, mostly but not exclusively with nearby states, which perhaps suggests they too are looking at exercises as a way of winning over their peripheries and establishing general conventional deterrence against the United States.

There is hardly any previous work on the use of joint military exercises specifically in cases of immediate deterrence. The anecdotal record, however, does provide some support. During Operation Desert Shield before the 1991 Gulf War, United Nations forces participated in

joint military exercises and other training designed to improve and/or demonstrate coalition military capabilities. There were also frequent (mostly United States-led) joint military exercises in Kuwait and the surrounding region during the 1990s, and President Clinton increased the scale and frequency of joint exercises around the 1998 Iraq crisis. South Korea engages in frequent joint military exercises, but it is probably best to characterize most of these exercises as part of a strategy of general deterrence. However, South Korea and the United States did initiate a number of additional and unusually large exercises in the wake of the 2010 sinking of the ROKS *Choenan*, which might be interpreted as part of a strategy of immediate deterrence.

The case study literature on military exercises suggests that they are generally large events whose primary purpose is to further deterrence. Blackwill and Legro (1989) study NATO and Warsaw Pact joint military operations during the Cold War and they focus on exercises that approach or surpass the notification requirements for European exercises established in the 1975 Helsinki Accords. These large exercises ensured that NATO could conduct rapid reinforcement and provided a signal by the United States that it would support its European allies. They argue that even smaller exercises were designed with essentially the same purpose. On the Warsaw Pact side, exercises served both to deter NATO and to deter defection by non-Soviet Pact members (Caravelli 1983). Farrell (2009) discusses the largest South Korea-United States joint exercises, *Team Spirit*, and therefore also focuses on exercises as a tool of deterrence.

Wolfley (2018, 2021) provides an exception to the case study approach that focuses on larger exercises and joint military exercises as a tool for deterrence. He notes the expansion of smaller exercises and exercises between non-formal allies. He argues that these exercises were not designed to enhance general deterrence, but rather to recruit or evaluate potential allies, to influence civil-military relations, build nascent military capabilities that would allow weaker states to maintain internal security, or as confidence building measures between antagonistic states. Wolfley terms these non-deterrence focused exercises "shaping exercises"

and argues that major powers are using these types of exercises to manage emerging security threats.

While I have focused on the possibility that joint military exercises could be used to make extended deterrence more credible, there is also the possibility that they bind reputations and capabilities together in a way that increases the risk of entrapment. Snyder (1961, 1997) argues that while formal alliances can be used to deter, they may also lead states into wars they would prefer to avoid. If a state believes that it will enjoy the support of its alliance partners, it may feel emboldened to resist reasonable concessions or even to initiate a crisis. The worry in this scenario is that one state has bound itself so closely to another that it feels it cannot abandon its partner, even if the partner pursues policies the first state opposes. While there is little evidence that entrapment is frequent (Beckley 2015; Benson 2012; Crawford 2003; Kim 2011), Blankenship and Kuo (2017) posit that the fear of entrapment helps structure states' decisions to conduct joint military exercises.

Joint military exercises can relate to alliances, but much of the growth in joint military exercises over the last 30 years has occurred outside of formal alliance structures. Bernhardt (2020, Chapter 2) argues that the growth in joint military exercises, especially between non-allies, is due to a combination of factors. The impending decrease in the risk of great power war with the end of the Cold War led militaries to search for new missions and purposes. Military leaders argued, successfully, that less traditional security concerns such as responding to foreign civil wars or aiding people hit by natural disasters, demanded military-led solutions. Joint military exercises, including ones that train for these non-conventional combat missions, proliferated as a result. Moreover, these types of interventions are generally done multilaterally and outside of traditional alliance frameworks, therefore spurring new forms of training with new sets of partners.

3.2 Joint Military Exercises and the Foreign Policy Toolbox

Besides the work on alliances, there is a burgeoning literature on a variety of other tools states might be using to improve military cooperation and signal their interests. Kinne (2018, 2019) points to the growth of defense cooperation agreements. These agreements help structure a variety of defense interactions, including things like arms sales and basic training assistance. These agreements vary on what they cover, and they are much more widespread than formal alliance treaties. Some of the parameters of a joint military exercise, such as the legal basis by which foreign soldiers will come to a country, can be governed by these agreements.

The tool that is most similar to joint military exercises is the provision of more basic forms of military training and officer training programs. A number of studies have examined the use of Security Force Assistance, which largely comprises basic military training and advice provided by special forces personnel and can also include providing military equipment (Biddle, Macdonald, and Baker 2018; Department of Defense Cost Assessment and Program Evaluation Simulation and Analysis Center 2010; Karlin 2018; McNerney et al. 2014; Paul et al. 2013; Watts et al. 2012). These studies all, however, look at the effectiveness of security force assistance on recipient country military capabilities and military effectiveness rather than theorizing which countries receive such assistance and in what forms. The cases these studies examine are largely ones in which the recipient country possessed relatively rudimentary capabilities prior to the assistance. Participating in joint military exercises, however, requires countries to have at least some units with more than basic capabilities, at least in the eyes of the force's own country. Joint military exercises might be viewed as being a rung up the latter of military assistance.

In addition to the types of basic aid provided as part of Security Force Assistance programs, many countries also invite the officers of other countries to participate in their professional military education programs. Examples of such education include the programs offered by the various war colleges in the United States or by the United Kingdom at Sand-

hurst. A number of studies have looked at the effect of training in the United States through the IMET program. Atkinson (2006, 2010) has found that having more officers participate in IMET is associated with a higher probability of a country democratizing, which she argues is due to the officers being socialized to prefer a model of civil-military relations similar to that in the United States, but Savage and Caverley (2017) find a nearly opposite result in that higher rates of IMET participation is associated with a higher probability of a military launching a coup attempt, which they argue is due to the officers gaining important organization skills through their training. Joint military exercises might have similar effects by exposing militaries to other models of civil-military relations or by inculcating improved organization.

Arms sales and transfers are another tool states might employ to signal their interests or build partner capabilities. Arms sales have been associated with increased influence by both suppliers or recipients over the others (Caverley 2007; Sislin 1994), similarly to how some have argued joint military exercises present an entrapment risk (Blankenship and Kuo 2017). With respect to the association between providing and receiving arms, studies have found a positive association when potential providers and recipients share a pre-exisiting close relationship or if the recipient is in a strategically advantageous geographic location (Erickson 2013; Grinberg, n.d. R. A. Johnson 2015; McManus and Yarhi-Milo 2017; Perkins and Neumayer 2010), indeterminate (Blanton 2000, 2005; Erickson 2013; Johnson and Willardson 2018; McManus and Yarhi-Milo 2017; Perkins and Neumayer 2010) or negative associations between a recipient's respect for human rights or democratic norms and receiving arms from Western suppliers (Akerman and Seim 2014; De Soysa and Midford 2012; Platte and Leuffen 2016), and that transfers are less likely when recipients have rivals different from the provider's, when recipients have a reputation for corruption, and when recipient countries are ruled by military regimes (Grinberg, n.d.). Mapping supplier-recipient pairs has also been shown to be a good measure of which bloc a country belonged to during the Cold War (Fearon and Hansen 2017). These studies suggest that arms sales and transfers are largely a

function of geostrategic variables related to international rivalries, which may only partially apply to joint military exercises.

There has also been a proliferation of research on overseas basing. Permanently stationing forces overseas can send important political signals and help countries tie their hands. Schelling (1966) famously described the role of American forces stationed in Berlin during the Cold War as a political trip wire that would ensure the president had to respond forcefully to Soviet aggression. More recent research continues to tie the presence of forward deployed forces to alliance credibility and the ability to reassure friends and partners (Hunzeker and Lanoszka 2016). Basing forces overseas can also influence the host state's foreign policy, potentially leading it to become more aggressive or adjust its defense spending (Allen, VanDusky-Allen, and Flynn 2016; Machain and Morgan 2013). Stationing soldiers on foreign territory has also been found to have profound effects on the host country's domestic politics (Cooley 2008). Joint military exercises, while not requiring permanent basing rights, usually require inviting foreign forces into one's own country and could act as a trip wire or commitment device in a similar manner to permanently deployed personnel.[8]

Visits by heads of state or important ministers might also share much in common with joint military exercises. They are a scarce resource because of the time commitment and the number of possible countries to which a dignitary might travel. They are also a sunk-cost signal. Like joint military exercises, it is possible to calibrate the content of the signal more directly than with formal alliances as countries can choose the level of the dignitary, the length of the visit, and potentially the forum in which the interaction occurs. Visits by heads of state have been connected to the success or failure of extended deterrence (McManus 2018). As for why some states are the subject of meetings between key dignitaries, McManus and Yarhi-Milo (2017) argue that the United States is most likely to use a highly visible signal

8. Some joint military exercises take place in international waters and thus would not lead to any foreign military presence in a country, although even in these cases there is often an in-port phase to the exercise. Some air exercises could also take place without requiring any forces to be on the ground in a foreign country. In principle, command post exercises might involve geographically disparate forces using communication links to simulate troop movements and thus not require any foreign troop presence.

of commitment like a presidential meeting when the target state is also a democracy and Lebovic and Saunders (2016) argue that strategic factors, such as whether an interlocutor is an ally or a key adversary, as well as domestic political concerns can influence the number of leadership visits and the destination of those visits. This research suggests that personal interactions help leaders determine each other's credibility and genuineness and develop trust, factors that could also extend to the bonds formed by military personnel during joint military exercises (Hall and Yarhi-Milo 2012; Holmes 2013).[9]

There have been attempts to combine measures like these to build a measure of foreign policy support. McManus and Nieman (2019) employed a latent variable approach to construct a level of support from major powers for smaller powers. They include measures such as visits and statements from country leaders, formal alliances, nuclear and troop deployments, arms sales, and joint military exercises.[10] Their approach allows them to let the data determine the importance of the different types of foreign policy signals. They find that all of the signal types can be important, and that after nuclear deployments, conducting joint military exercises is one of the richest signals for the United States, France, and the United Kingdom. They also show that their measure of derived support correlates with a willingness to militarily support a protégé.

4 Collecting the Joint Military Exercises data set

Joint military exercises are an important way countries signal their foreign policy interests and build joint capabilities, but we lack reliable data to use them in quantitative analyses. D'Orazio (2013) collected information on joint military exercises, but an analysis of these data reveals a number of flaws. The data include exercises that likely did not take place, such

9. Conversations with U.S. military officers suggest that one of the primary benefits of joint military exercises are the bonds officers form with their counterparts. These bonds help create common understandings and support interoperability in military operations. Interview subjects preferred not to be identified by name or rank. They included members of the Army, Navy, Air Force, Marines, and Coast Guard. All interviews were conducted under Stanford University Internal Review Board Protocol 51151.

10. They use the problematic D'Orazio (2013) joint military exercise data without addressing any of the issues I discuss in the following section.

as an exercise recorded as being between the United States and North Korea,[11] and misses some important exercises that did occur. In some instances, there are duplicate exercise entries or some publicly disclosed iterations of an exercise series are missing. In other cases, the data lack the necessary information for creating dyadic data sets or the listed countries appear incorrect. For example, for exercises between 1980 and 2010, there are five cases in which there are not at least two countries listed as participating in the exercise, 34 examples of a named exercise appearing more than one time in the same year,[12] and there are at least 188 additional entries that show strong indications of being duplicates.[13] This amounts to at least 13% of the relevant entries. There are also cases of prominent exercise series that occur on a regular basis but for which certain years are missing. For example, the *Ulchi Focus Lens/Ulchi Freedom Guardian* series of exercises is inconsistently included in the data. This exercise has occurred annually without exception since before 1980 through 2016, such that there should be more than 30 entries. The data, however, include only 13 entries, and two of those (1997 and 1999) are duplicates. These data issues could induce significant measurement error for analyses that use either the exercise as the unit of analysis or construct binary or count measures at the dyad-year level of joint military exercise activity, in addition to any concerns over whether the collection process yielded a representative sample of joint military exercises.

Although these flaws attest in part to the difficulty in collecting information about joint military exercises, it should be possible to collect reliable data on these activities. The existence of most joint military exercises is public knowledge. Transporting personnel from one country to another and having foreign military personnel in country makes it difficult to conceal the presence of an exercise. When soldiers are permanently based abroad it may be

11. Exercise 730 in D'Orazio (2013) data records an exercise between the United States and North Korea in September 2010.

12. Some exercise series take place multiple times per year, but the duplicates in this case have the same estimated dates rather than being different iterations within the same year.

13. These exercises do not share the same name but started in the same year and month, included the same total number of countries and at least two of the same countries, and started either within two days of each other or were in the same location. A full analysis of the duplicates is impossible due to missing data and inconsistency in naming conventions.

easier to shield exercise activity and it is certainly possible to conceal certain aspects of an exercise. When multiple exercises occur simultaneously, it may be possible to only publicize some while maintaining secrecy for the others.[14]

In many cases, militaries broadcast to friends and adversaries alike the presence of a joint military exercise. A secret exercise might build up joint capabilities and show trust to a friend, but if it truly is unobserved by a foe then it will not send a signal of any kind. Partly because countries often want others to know they are exercising, they take steps to prepare public relations strategies for the exercise. In other cases, international agreements or conventions lead countries to announce their intention to hold an exercise and invite observers from non-participants. Advance warning and foreign observers help build confidence that an exercise is not actually a prelude to a military attack and inviting observers can increase the signalling content of an exercise by showing interested parties the combined capabilities the exercise participants possess.

Despite exercise participants taking steps to publicize their activity, there is little in the way of systematic reporting of joint military exercises. Countries in the Organization for Security and Cooperation in Europe (OSCE) agreed in the Helsinki Final Act in 1975 to give limited advance warning to each other if their exercises exceeded certain size thresholds (including for unilateral exercises) and since the Stockholm Document in 1986 have agreed to file expected exercise calendars for the forthcoming year as well. However, exercises that fall below the size thresholds do not need to be reported, exercises outside of Europe do not need to be reported, the extent to which the relevant countries comply with the agreements is suspect, and the OSCE does little to publish information it receives for the general public.[15]

Because there are no official sources that systematically collect reliable information on

14. One prominent example of this was the *Abel Archer* exercise in 1983. *Abel Archer* took place during the *Autumn Forge* program of exercises but was a classified exercise since it involved nuclear war scenarios. Shielding the existence of an exercise from a foreign intelligence service is likely harder than merely classifying it, as the crisis that resulted from *Abel Archer* shows.

15. NATO countries claim that many Russian-led exercises, including some joint exercises, exceed the reporting thresholds despite Russian claims to the contrary. Russia has scheduled simultaneous maneuvers that appear to relate to each other and potentially falsified the size of the forces involved in some maneuvers in an attempt to stay below the notification thresholds. See, for example, Quinlivan and Oliker (2011).

joint military exercises, I rely on news media reporting. If militaries publicize their activity, then it should be discoverable in open source media. The problem is how to find the reports militaries make to the news about their activities and then to systematize the information contained in those articles. I used a technique based on the process by which the Militarized Interstate Dispute v. 4.0 (MID4) data set was built to collect information on joint military exercises (Palmer et al. 2015). The process involved collecting potentially relevant news articles, classifying them into relevant and irrelevant categories, and then reading through the relevant articles to extract the information about joint military exercises.

I collected news articles using LexisNexis, using a search query that sought to identify all articles that mention joint military exercises. The query consisted of three parts. First, the articles need to mention a word related to militaries or military organizations. Second, the articles needed to mention words related to joint activity.[16] Finally, within 25 words of the joint term, the article needed to mention words related to exercise or training. The exact terms used in the querry are showin in Table 1. Articles also had to be between 150 and 5,000 words, long enough to exclude mere headlines that would confuse a classifier and short enough to exclude major reports that are not actually news articles.

LexisNexis has hundreds of different sources from which to search. The MID4 project found that for their purposes, general news sources with global coverage were sufficient for identifying disputes. They determined this by calibrating different categories of news sources against the previous versions of the MID data. Because there is not a previous version of joint military exercise data against which I can calibrate sources, I erred on the side of including more types of sources. Moreover, while militarized disputes may be prominent enough to garner coverage in sources with a global scope, joint military exercises may be small enough that only sources with more tailored geographic coverage would bother to report on them. I therefore used two pre-defined categories of sources, Major World Newspapers and Wire Service Stories, as well as the BBC Monitoring Service's International Reports. The BBC

16. I used both words drawn from United States military publications, my reading of popular news media, and the names of prominent international organizations or alliances.

Military Category	Joint Category	Exercise Category
(militar! OR war OR wars OR combat! OR defen! OR maritime OR navy OR navies OR naval OR air force! OR army OR armies OR marines OR marine corps OR armed forces)	(multi-nation! OR multination! OR joint OR combine! OR bilatera! OR multi-lateral! OR multilateral! OR international OR (nato OR north atlantic treaty organi!) OR (warsaw pact OR warsaw treaty) OR (seato OR south east asian treaty organi!) OR (cento OR central treaty organi!) OR (anzus) OR (csto OR collective security treaty organi!) OR (ecomog OR ecomil OR economic community of west african states monitoring group) OR (rio pact OR rio treaty OR TIAR OR inter american treaty of reciprocal assistance) OR (sco OR shanghai cooperation organi!) OR (rss OR regional security system) OR (peninsula shield force OR gulf cooperation council OR gcc) OR (asean OR association of southeast asian nations) OR (unasur OR council of south american defen! OR south american defen! council)OR (common security and defen! policy OR csdp OR european security and defen! policy OR esdp) OR (commonwealth of independent states OR CAPS(cis)) OR (CAPS(guam) PRE/1 (organi! OR treaty))OR (islamic military alliance OR CAPS(ima) OR imaft OR islamic military counter terrorism coalition) OR (organi! of islamic cooperation OR CAPS(oic)) OR (south atlantic peace and cooperation zone OR zpcas OR zopcas) OR (inter american defen! board OR iadb))	(exercis! OR train! OR simulat! OR (war PRE/1 game!) OR (military PRE/1 game!) OR drill! OR operation! OR (manoeuvr! OR maneuver!))

Table 1: LexisNexis Search Terms

Monitoring Service translates to English articles from foreign languages and so helps to compensate for the English-language constraint I face in building the classifier.

After downloading all of the articles from the period 1977-2016, I conducted two further steps to pare down the number of articles. First, I require that articles mention either a major international organization or alliance or at least two different countries in the body of the article.[17] Second, because of the large number and types of sources from which I drew articles, there were many articles that contained nearly the same information or were in fact duplicates of each other. To remove these duplicates and near-duplicates, I searched for similar articles published within two days on either side of an article and removed the duplicates.[18] This led to a corpus of 751,854 news articles, still far too many to read manually.

To separate the articles into those that deserved close reading and those that were unlikely to contain information about joint military exercises, I employed a support vector machine. Support vector machines are a type of supervised machine learning that is well suited to classifying documents due to its ability to handle high-dimensional data (Joachims 1997, 1998, 1999, 2001; Kolari et al. 2006). While there are a host of different classifiers, I used the support vector machine because it was successful for the MID4 project. Following standard procedures, I converted each article into its unigrams, the words in the article, stemmed them using the Porter (1980) stemmer, and removed common stop words.[19] This process results in a term frequency matrix in which each row corresponds to an article, each column to a unique unigram, and the cells are the number of times that the unigram appears in the article. I further reduced dimensionality by only keeping the top 1% most frequently occurring unigrams. I then followed standard practice by weighting by the inverse document frequency.

17. I use the country names from Schroedt (n.d.), which includes common synonyms and alternative names, capital cities, abbreviations, and common spelling errors.

18. After stemming and removing stop words, I created a matrix of all trigrams for the articles on the given day and two days on either side of the given day and then found the Euclidean distance between articles. Based on an analysis of articles, I determined that articles with a distance of less than 12 were highly likely to be duplicates.

19. I also removed the article meta data, frequently appearing titles such as "Mr." or "Mrs." and month names so as not to bias particular time periods.

As a supervised learning method, support vector machines require data with known categories. This presents a challenge, as there is no existing data set of articles where it is known whether the article contains information about a joint military exercise. Moreover, there are no generalizable rules of thumb for deciding how many entries are needed to build an effective classifier. To fill this gap, I skimmed 2% of the articles to record whether they were about joint military exercises. By using 2% of the articles, I could build a classifier using nearly as many articles as terms and still have a substantial test set with which to evaluate the model.[20]

I trained the support vector machine on around 12,500 articles and reserved 2,500 articles for validation. Support vector machines allow the user to choose a number of parameters that influence the fit of the model. Choosing different kernels, cost parameters, and flexibility parameters can lead to drastically different performances. I employed 5-fold cross-validation across a number of possible specifications to determine the optimal choices. After choosing the best model on the training data, I evaluated its performance on the validation set.[21] Summary statistics of the model's performance are shown in Table 2. They show that the model performed quite well, with a precision of 0.703 and a recall of 0.731. In the evaluation data, the model predicts that around 11.2% of articles are about joint military exercises compared to the null rate of 10.7%, but the precision represents a seven-fold increase in classification accuracy over randomly assigning categories.

	Predicted No	Predicted Yes
True No	2,149	83
True Yes	72	196

Table 2: Confusion Matrix for Optimal Support Vector Machine

After building the model using the training data, I applied it to the whole corpus to

20. How many entries one needs for training data depends on the structure of the underlying data. The answer depends on the number of predictors, the number of categories to sort, the strength of the predictors, and the sparseness of the document-term matrix (Grimmer and Stewart 2013). Hopkins and King (2010) suggest a "rule of 500" and other sources suggest significantly more (Collingwood and Wilkerson 2012).

21. The model that performed best employed a RBF kernel, a cost of 50, and gamma of 1^{-5}. All models were built in Python using the `scikit-learn` package.

obtain the predicted class of all 751,854 articles. This results in 82,040 articles classified as containing information about joint military exercises, in line with the base rate at which such articles appear in the training data. This number, however, is still far too many for me to read closely. To further narrow the number of articles to read, I looked at the 5% of articles that the model was most confident were about joint military exercises.[22] This results in 4,102 articles to read closely, a large but manageable task. Moreover, the precision in this group is 0.922, meaning the vast majority of articles yield relevant information about a joint military exercise.[23]

D'Orazio (2013) used a similar approach, but with a few important differences.[24] First, I began with a significantly wider set of source articles. Instead of relying only on a small handful of the largest English-language international wire services, I included more regionally focused wires, regional newspapers, and articles translated from local, foreign-language sources. While militaries publish information on their exercise activity, it often is not noteworthy enough to rise to the inclusion of a major international story and thus may not receive an article in an outlet such as the Associated Press. I also started with a set of search terms that I believe was better tailored to including any article containing information about joint military exercises. Second, the way in which my machine learning process functioned was considerably more successful. This could be due to my creation of a specialized training data set or because of different model choice decisions. Finally, I read through significantly more articles that the machine learning process judged as being about joint military exercises, and thus was likely able to identify far more different exercises as well as gathering more information about each exercise.

22. With support vector machines, the model calculates a separating hyperplane to divide the entries into each category. The greater the distance an entry is from the separating hyperplane, the more confident the model is that it has been classified correctly.

23. The precision for the MID4 project was reported as around 0.083, although this represents the precision for the full set of articles, not for the 5% most confidently classified articles (Palmer et al. 2015, p. 226).

24. D'Orazio (2013) followed a similar order of operations. He began by generating a list of search terms, collected articles from LexisNexis, built a classifier using a support vector machine, and then had research assistants read a subset of the positively identified articles to extract exercise information. It is unclear how he generated training data for the support vector machine.

I read through the articles in the most likely set to extract information about joint military exercises. Many articles contained information about more than one exercise[25] and there were often multiple articles about a particular exercise. For each exercise identified in the articles, I attempted to record the name, if any, of the exercise; the countries that were active participants in the exercise; the dates of the exercise;[26] the country and/or body of water in which the exercise primarily occurred; whether the exercise included field training and/or command post training elements; whether it included the use of live-fire training; the number of aircraft and ships of various types from each country as well as for the exercise as a whole; the number of armored vehicles employed by each country and in the exercise as a whole; the number of personnel participating for each country and for the exercise as a whole; the service branches each country sent to the exercise; the type of combat and/or non-combat training conducted; an alliance organization, if any, that helped organize the exercise; whether there were international observers; and whether the exercise was part of a recurring series.[27] I then conducted additional research of these specific exercises to fill in as much information about the exercise as is publicly available.[28] This process yielded 5,198 joint military exercises from 1977-2016. The review process ensures that I do not have duplicate exercises, implausible exercises, and incomplete inclusion of major exercise series, all problems evident in the D'Orazio (2013) data effort.

The new data set is much more comprehensive than previous efforts. The top panel of Figure 4 shows the percentage of dyads for which at least one joint military exercise was

25. This was especially true with respect to exercises that are part of a named series that repeats on a regular basis.

26. Where an article was unclear about the dates of an exercise but mentioned it starting or ending at the beginning or start of month I used the 1st; if it says middle of the month, the 15th (14th for February); and the end of month was the last day of month.

27. To count as part of a recurring series, it must be the second or subsequent iteration of an exercise.

28. Where sources disagreed about exercise details, I attempted to determine the most credible information. I preferred information from major international wire services such as the AP, AFP, and UPI and then from lesser known and regional outlets and major newspapers. When there was a conflict between information originally in English and originally in a foreign language I used the information from the original English source. When recording estimates of country or exercise material or personnel contributions, I used the most conservative available estimates. This means that if an article says "thousands" of personnel participated, I record 1,000 personnel and prefer the smaller number when sources conflict.

recorded in both my new data and the D'Orazio (2013) data for the temporal period in which they overlap. In nearly all cases, the new data includes many more dyads that would previously have been missed. There are few cases in which the older data adds dyads missed by the new data, and the years in which there are the largest divergences, it appears to be in part because the D'Orazio (2013) included as joint military exercises some large multilateral workshops that would not meet my definition of an exercise. Moreover, in in the years in which there are more dyads included in the older data, there are still substantially more dyads that relying on the older data would have left out.

The bottom panel uses the number of exercises in which a dyad participated rather than a binary measure. It looks at the difference between the number of exercises in a year a dyad recorded in the new data and the number in the D'Orazio (2013) data.[29] The density plots show the average within dyad difference for three periods: the Cold War, between the end of the Cold War and the September 11, 2001 attacks, and post-September 11 to 2010. It shows that in all periods, the new data records, in most instances, more exercises per dyad than the older data. This difference is increasingly large in the most recent period.

4.1 Possible Biases

There are three major areas of possible bias about which one should potentially be concerned. First, and most importantly, it is possible that the increasing availability of digitized news articles leads to an increase in discoverable joint military exercises such that any observed increase in recorded exercises is really just an artifact of the increase in available news. Second, relying on English language news articles could induce bias toward exercises involving English speaking countries and lead to observing an increase in exercises primarily involving such countries rather than a more generalizable trend. Relatedly, apart from any bias induced by the language of the news articles, there could be a bias toward particular geographic regions of the world if some regions are more likely to report on joint military exercises than

29. I only include dyads for which either data set recorded at least one joint military exercise in any year. I have trimmed the positive side x-axis of the figure. The full density extends out above 15.

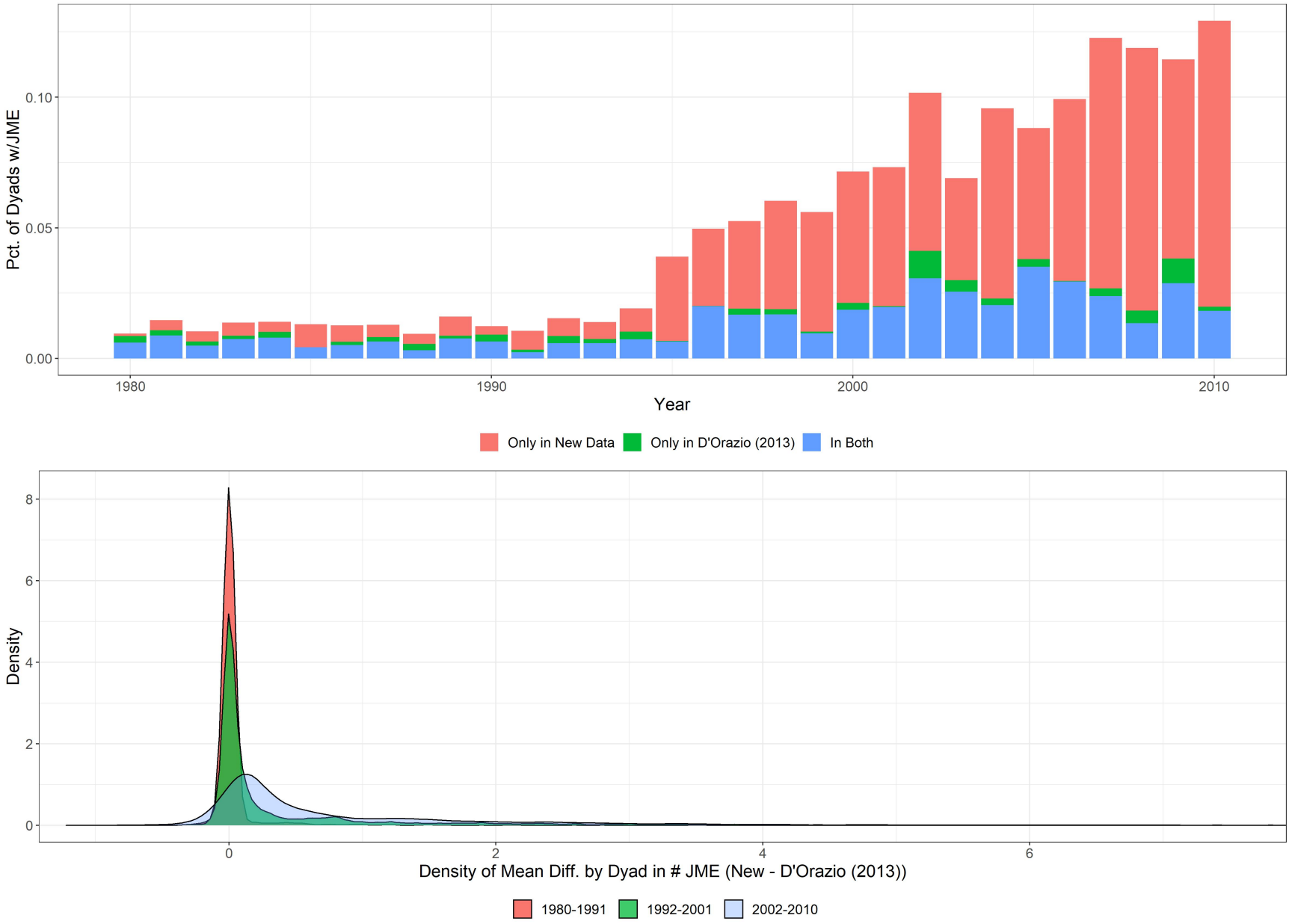

Figure 4: Comparison of New Data to D'Orazio (2013) Data

others. Third, the decision to only read the most likely set of articles to collect the original list of joint military exercises could lead to concerns about missing data that reading a larger set of articles would avoid. Fortunately, there are no reasons to believe that these first two issues has significantly biased the data and the experience of coding the exercises and doing the subsequent review leads me to discount the severity of the third.

With respect to temporal bias from the availability of news articles, it is true that there is substantially more news available electronically as time goes on. Figure 5 shows a number of plots of the trends in news articles from which I code exercises. Panels A, B, and C, respectively, show the number of news articles in my original corpus, the number of articles the classifier believes are about joint military exercises, and the number of articles in the group of the top 5% about which the classifier is most confident and that were read closely to make the data. All three panels show a substantial increase over time in the number of articles. Panels D, E, and F show panels A, B, and C divided by the number of joint military exercises recorded in each year. The concern presented by panels A, B, and C is that there could be a large number of unrecorded exercises in earlier years when there were relatively few available news articles from which to extract information. However, panels D, E, and F suggest that there would have also been a very large shift in the way the news media report. They show that there are fewer articles per observed joint military exercise in earlier years. To believe there are a large number of missing exercises only from those years, it would have to be true that this ratio would be even further depressed relative to more recent years. Such a substantial change in reporting norms seems unlikely, which suggests that the increase in observed exercises is real.

It is more difficult to directly address the possible bias arising from varying news availability due to geographic and language differences. By using a wide variety of sources, including some that were not originally in English, I have attempted to mitigate this issue. The Wire Service category on LexisNexis includes services from all over the world, including many from non-English speaking countries but that produce English language content. The Major

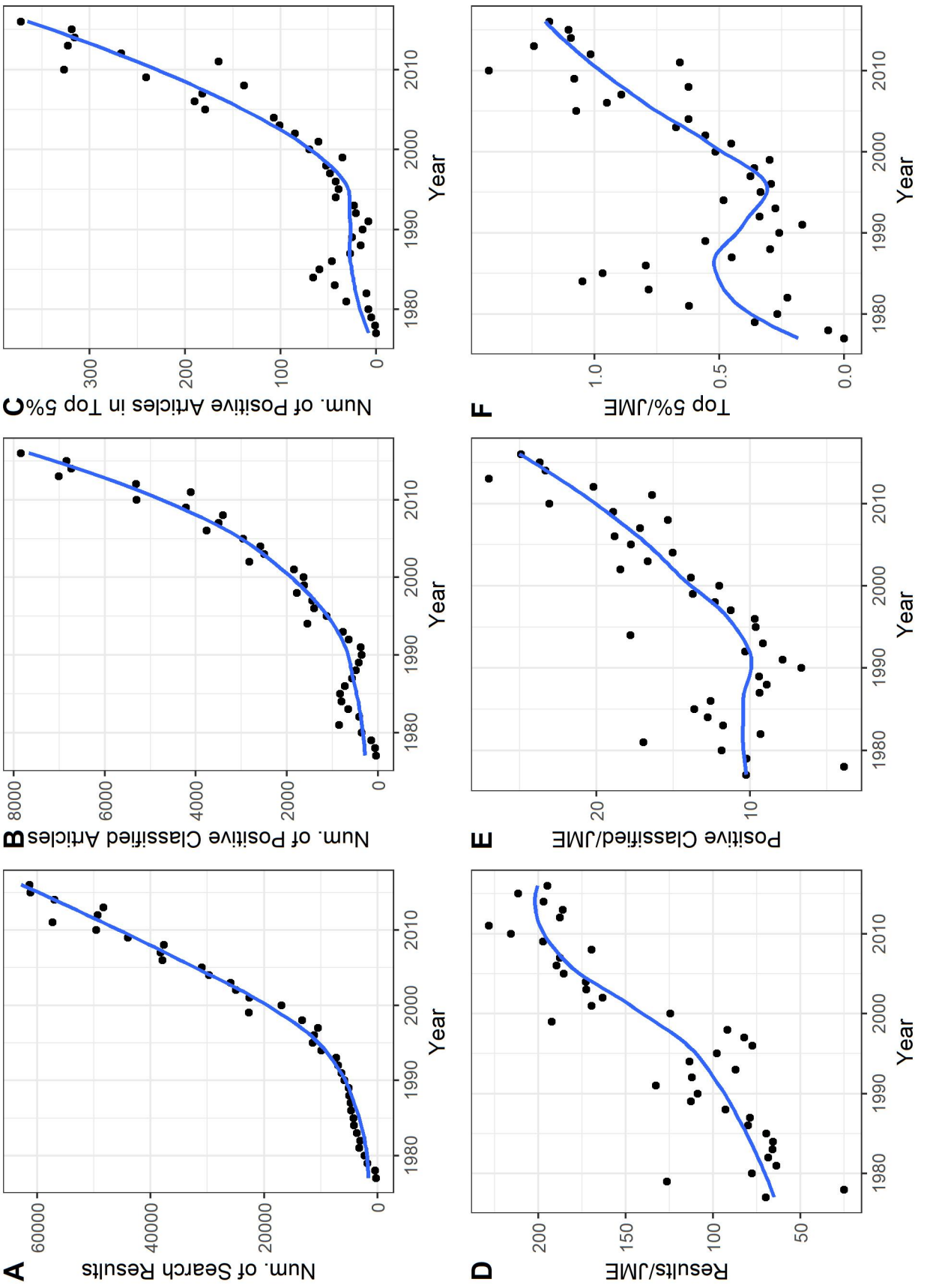

Figure 5: Change in News Articles

World Newspapers category and the larger wire services tend to be based in larger and richer countries. Over time, however, LexisNexis has added substantially more wire services from smaller or poorer countries and regions. If news media from larger and richer countries and regions are more likely to report on joint military exercises involving larger or richer countries and regions, then the major potential bias would be missing exercises that only involve smaller or poorer countries and regions, especially in earlier years. However, while there are fewer exercises that do not include any of the major powers—a reasonable proxy for where the majority of the largest media sources are based—than there are exercises that include at least one major power, the trend is not toward an increasing share of joint military exercises that do not involve a major power. This suggests that there is not substantial geographic or language induced bias.

Finally, the decision to code exercises only from those articles the classifier deemed most likely to be about joint military exercises but not from the set of articles it classified as about exercises but about which it showed less confidence could lead to an under-counting of joint military exercises. Indeed, there are certainly more exercises that have taken place than are included in the data. However, when reading through the articles, I found there were many examples of a single exercise being identified by multiple articles and the unread articles could plausibly contain largely additional information about already identified exercises. Second, when an article identified an exercise as part of a recurring series, I recorded all occurrences of that exercise, even if the original article gave no details about those previous iterations. This meant that a single article could potentially identify significantly more than a single exercise. My review process filled in the missing information about these exercises. Lastly, I reviewed a fraction of additional articles the classifier believed were about joint military exercises. I can infer how confident the support vector machine was about whether an article contained information on joint military exercises by looking at the distance the article falls from the separating hyperplane generated by the algorithm. Whereas the data set is based on reading the 5% of articles furthest from this hyperplane (and thus the articles about

which the classifier is most confident in its decision), here I sampled from articles slightly closer to the hyperplane as well as from a set closest to the hyperplane (articles about which the algorithm had less confidence in its decision but for which it classified as being about a military exercise). In the group further from the hyperplane, all of the articles were classified correctly, and all the exercises they mentioned had already been included in the data. In the group closest to the hyperplane, only 44% of the articles were classified correctly (in that they included information about joint military exercises), but again, all of the articles that did contain information on exercises mentioned exercises that were already in the data. This suggests that the risk of substantial under-counting of exercises is low. It also suggests that these low-confidence articles may contain mostly passing references to exercises discussed more thoroughly in other articles.

5 Validating the Joint Military Exercises Data

If joint military exercises are an important part of how states conduct military diplomacy and signalling, then they ought to correlate with other measures of foreign policy activity. Interestingly, while having a formal alliance does positively correlate with conducting joint military exercises, as noted previously, much of the growth in exercises has occurred between countries that do not share an alliance. It is important that joint military exercises not merely replicate alliances. Instead, they provide an additional, rich set of information about foreign policy interests.

In this section, I test the correlation between a dyad conducting joint military exercises together and whether it engages in militarized disputes, whether conducting joint military exercises is associated with a higher probability of joining a militarized interstate dispute, and whether joint military exercises are correlated with trading weapons with each other. Joint military exercises can reveal information about a country's military capabilities to its exercise partners, which would potentially be risky if either country anticipated engaging

in a military conflict with the other.[30] They could also be an important part of military strategy or signalling, such that those that conduct exercises are more likely to join their exercise partners in disputes. Similarly, if countries engage in arms trades with each other, they likely anticipate that the other is a useful partner. They might find that joint training assists the recipient of the arms in using the new weapon system. Thus, we should expect that joint military exercises negatively correlate with engaging in militarized disputes and positively with arms trades.

In the analysis that follows, I operationalize joint military exercises in a number of ways. The simplest way is as a binary variable for whether a dyad conducted any exercises in a given year. This variable is less likely to suffer from bias compared to a count of the number of exercises a dyad conducted because it does not rely on having identified every exercise, just at least one exercise. However, I also test specifications that use the count of the number of exercises. Not all exercises are the same, so I also disaggregate exercises based on whether the exercises purely focused on non-combat activities and whether it was a field training exercise that practiced combat maneuvers. This latter type of exercise is the most directly applicable to conventional war fighting and the most likely to send a deterrent signal to potential adversaries.

I include a number of control variables to account for other factors that might influence whether a dyad engages in disputes or trades weapons. In the analysis looking at the propensity to engage in a MID and to engage in arms sales, I include a measure of the share of combined dyadic national capabilities for the larger state as well as a logged measure of the national capabilities of the smaller state (Singer, Bremer, and Stuckey 1972).[31] I also control for whether the dyad includes at least one major power (Correlates of War Project 2017). To

30. As an exercise can reveal weaknesses, they could also reveal strengths. If an exercise led a potential adversary to revise upward its estimate of an opponents military capabilities, it would improve the bargaining position of the newly credible stronger country. A country could invite a foreign observer if it was mostly concerned with revealing private information rather than inviting the potential adversary to actively participate in an exercise.

31. I extend the national material capabilities data following the same protocol and sources through 2015 for the arms trade analysis. See Chapter 2 for details.

account for the role regime type might play, I include controls for whether the countries in the dyad are both democracies or both non-democracies (Marshall, Gurr, and Jaggers 2018).[32] I also include two controls for geography, the minimum distance between the borders of the two countries (Weidmann and Gleditsch 2010) and whether the countries share a border (Stinnett et al. 2002). Lastly, I include controls for the underlying relationship between the members of the dyad, including whether they have a formal offensive or defensive alliance (Leeds et al. 2002), the log of the number of shared memberships in intergovernmental organizations (Pevehouse et al. 2020), and the log of the total dyadic trade (Barbieri, Keshk, and Pollins 2009; Barbieri and Keshk 2016). I interact the joint military exercise measure with the alliance measure to see how the effect of joint military exercises might differ when conducted within a formal alliance structure.

I begin by looking at the likelihood that countries engage in militarized disputes. Version 4.3 of the Militarized Interstate Dispute data (Palmer et al. 2015) records disputes through 2010, so I am not able to use the more recent years of joint military exercises. The dependent variable is whether a dyad engaged in any disputes in a given year. Because the data is binary, I use a series of logit models. To control for the temporal dynamics of militarized disputes, I include the raw number, square, and cube of the number of years since a dyad had a dispute (Carter and Signorino 2010). The analysis uses all dyads in the international system from 1980-2010.

The results are in Table 3. Under all specifications, the direct effect of conducting a joint military exercise is negative and statistically significant. Interestingly, the direct effect of a formal alliance is to increase the probability of a dispute, in most cases significantly so. The interaction effect between conducting an exercise and being in an alliance is also positive, although it is not statistically significant in all specifications. This is surprising, since an alliance ought to reduce the probability of a militarized dispute, and one would think that allies that also conduct joint exercises together would be even more likely to not engage in

32. I count a country as having a democratic government if it has a combined polity score of 6 or greater on a -10 to 10 scale where -10 is the most autocratic government and 10 is the most democratic.

militarized disputes with each other. This result deserves further research.

The unit of observation for the analysis looking at joining a dispute is different than for engaging in a dispute or trading arms, and thus the way the variables enter the model is different. Rather than using a sample of dyads in the world, I start with each state that was an originator of a dispute and create dyad-year data that pairs each originator with every other country in the world (except the originator(s) on the other side of the dispute). The dependent variable is whether a candidate state joined the dispute on the side of the originator in the dyad. A candidate state remains in the sample until it either joins the dispute or the dispute ends.

The main independent variables enter the model identically to the analysis looking at the propensity to engage in a dispute. I measure the presence or number of joint military exercises of particular types that the candidate state and the originator engaged in. I also include controls for the log of the number of shared international organization memberships, the log of the amount of trade the dyad conducts, whether the countries are both democracies or both non-democracies, and whether the originator in question of the candidate joiner is a major power. Material capabilities enters in two ways. First, I include a control for the share of power amongst the originators that the side of the originator possessed when the dispute began. Second, I include the log of the capabilities possessed by the candidate joiner. I also control for geography using a variable for whether the candidate joiner bordered one of the originators that was not on the initiating side. I control for temporal aspects of the dispute by including a cubic specification of the number of years since the start of the dispute. I use a logit model and cluster the standard errors on both the dispute level and the country level because in multi-year disputes I am looking at multiple observations per dyad and in all disputes each candidate country appears as a possible joiner on both sides of the dispute.

Table 4 shows that engaging in joint military exercises with an originator of a dispute is associated with statistically significantly higher probabilities of joining that originator in the dispute. Interestingly, the direct effect of a formal military alliance is negative, although

Table 3: JMEs and Engaging in MIDs

	Dependent variable:					
	anyMid					
	(1)	(2)	(3)	(4)	(5)	(6)
Binary JME	−0.362**					
	(0.144)					
Binary Combat FTX JME		−0.465**				
		(0.211)				
Binary Non-Combat JME			−0.321**			
			(0.153)			
Count JME				−0.149***		
				(0.051)		
Count Combat FTX JME					−0.318**	
					(0.135)	
Count Non-Combat JME						−0.148**
						(0.067)
Alliance	0.126	0.204**	0.110	0.176**	0.232***	0.151*
	(0.092)	(0.087)	(0.088)	(0.086)	(0.084)	(0.085)
JME*Alliance	0.454**	0.334	0.730***	0.116**	0.195	0.196**
	(0.191)	(0.256)	(0.216)	(0.054)	(0.141)	(0.078)
log(Shared IGO Memberships)	0.566***	0.546***	0.554***	0.562***	0.548***	0.550***
	(0.095)	(0.094)	(0.095)	(0.094)	(0.094)	(0.094)
log(Dyadic Trade)	−0.038**	−0.037**	−0.041***	−0.035**	−0.034**	−0.040***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Joint Democracy	−1.095***	−1.079***	−1.119***	−1.028***	−1.042***	−1.088***
	(0.115)	(0.114)	(0.115)	(0.115)	(0.112)	(0.116)
Joint Non-Democracy	−0.365***	−0.374***	−0.365***	−0.375***	−0.383***	−0.369***
	(0.096)	(0.096)	(0.096)	(0.096)	(0.096)	(0.096)
Balance of Power	1.329***	1.350***	1.362***	1.330***	1.335***	1.361***
	(0.275)	(0.275)	(0.275)	(0.275)	(0.275)	(0.274)
log(Power (Smaller))	0.509***	0.516***	0.515***	0.511***	0.517***	0.513***
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
Border	2.055***	2.050***	2.050***	2.058***	2.050***	2.052***
	(0.090)	(0.090)	(0.090)	(0.090)	(0.090)	(0.090)
Min. Distance	−0.159***	−0.157***	−0.159***	−0.158***	−0.155***	−0.159***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Major Power	0.489***	0.479***	0.473***	0.488***	0.488***	0.474***
	(0.100)	(0.100)	(0.100)	(0.100)	(0.100)	(0.100)
Peace	−0.420***	−0.420***	−0.421***	−0.419***	−0.420***	−0.420***
	(0.015)	(0.015)	(0.015)	(0.014)	(0.014)	(0.015)
Peace^2	0.012***	0.012***	0.012***	0.012***	0.012***	0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Peace^3	−0.0001***	−0.0001***	−0.0001***	−0.0001***	−0.0001***	−0.0001***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Constant	−2.335***	−2.265***	−2.270***	−2.336***	−2.272***	−2.280***
	(0.408)	(0.404)	(0.408)	(0.406)	(0.404)	(0.406)
Observations	306,362	306,362	306,362	306,362	306,362	306,362
Log Likelihood	−4,899	−4,900	−4,897	−4,897	−4,897	−4,899
Akaike Inf. Crit.	9,830	9,832	9,826	9,826	9,825	9,830

Note: All variables except Peace, Min. Distance. and Border are lagged one year. *p<0.1; **p<0.05; ***p<0.01

Table 4: Propensity to Join MIDs

	Dependent variable:					
	Potential Joiner Joined on Side of Originator					
	(1)	(2)	(3)	(4)	(5)	(6)
Binary JME	1.9856***					
	(0.3731)					
Binary Combat FTX JME		0.7832**				
		(0.3301)				
Binary Non-Combat JME			1.8015***			
			(0.4042)			
Count JME				0.2217***		
				(0.0583)		
Count Combat FTX JME					0.2631***	
					(0.0904)	
Count Non-Combat JME						0.2932***
						(0.0808)
Alliance	−0.3699	−0.9386	−0.1827	−1.0278	−1.1978	−0.8034
	(1.0770)	(1.1010)	(0.9941)	(1.2369)	(1.2752)	(1.1646)
JME*Alliance	−1.2153**	−0.4177	−1.5212**	−0.1183*	−0.0965	−0.1603
	(0.5684)	(0.5001)	(0.7478)	(0.0684)	(0.1030)	(0.1299)
log(Shared IGO Membership)	1.9292	2.2278	1.9300	2.3308*	2.3612*	2.2414*
	(1.3739)	(1.3700)	(1.3066)	(1.3752)	(1.3623)	(1.3623)
log(Dyadic Trade)	0.0433	0.1127*	0.0796	0.0905	0.1057	0.1112
	(0.0729)	(0.0652)	(0.0747)	(0.0703)	(0.0683)	(0.0712)
Joint Democracy	0.2054	0.4059	0.2260	0.2979	0.4121	0.2859
	(0.3269)	(0.3495)	(0.3176)	(0.3470)	(0.3505)	(0.3383)
Joint Non-Democracy	−0.9032***	−0.9501***	−0.9426***	−0.9379***	−0.9513***	−0.9526***
	(0.3019)	(0.3000)	(0.3037)	(0.3039)	(0.3014)	(0.3050)
Balance of Power	4.2761	4.2474	4.2576	4.2558	4.2440	4.2149
	(3.1778)	(3.2369)	(3.1478)	(3.2574)	(3.2376)	(3.2078)
Joiner Power	0.0809	0.0217	0.0876	0.0310	0.0150	0.0423
	(0.1614)	(0.1755)	(0.1506)	(0.1691)	(0.1749)	(0.1663)
Joiner Borders MID	0.4946	0.7327	0.5282	0.6882	0.7854	0.6771
	(0.7113)	(0.6622)	(0.7085)	(0.6969)	(0.6650)	(0.6823)
Major Power	−0.8532	−0.7085	−0.8374	−0.7780	−0.7177	−0.7588
	(0.6095)	(0.5796)	(0.6339)	(0.6261)	(0.5798)	(0.6213)
MID Duration	−1.2825	−1.3062	−1.2882	−1.2850	−1.3087	−1.2751
	(0.8805)	(0.9237)	(0.8854)	(0.9074)	(0.9306)	(0.8976)
MID Duration^2	0.2598	0.2791	0.2579	0.2660	0.2802	0.2598
	(0.2377)	(0.2536)	(0.2374)	(0.2429)	(0.2556)	(0.2391)
MID Duration^3	−0.0150	−0.0169	−0.0147	−0.0157	−0.0170	−0.0152
	(0.0167)	(0.0185)	(0.0167)	(0.0175)	(0.0188)	(0.0171)
Constant	−16.1853**	−17.6396**	−16.1761**	−17.8420**	−18.0996**	−17.5069**
	(7.3213)	(7.4674)	(6.9531)	(7.4656)	(7.4638)	(7.3222)
Observations	343,366	343,366	343,366	343,366	343,366	343,366
Log Likelihood	−3,258	−3,396	−3,289	−3,361	−3,392	−3,365
Akaike Inf. Crit.	6,545	6,822	6,608	6,752	6,813	6,759

Note: All variables except MID duration and border lagged one year. *p<0.1; **p<0.05; ***p<0.01

statistically insignificant, once one includes controls for the presence of joint military exercises and other factors. The interaction effect of an alliance and conducting joint military exercises is negative and in some cases statistically significant, suggesting that an alliance, after controlling for other factors, attenuates the effect of holding exercises.

The curious finding with respect to alliances can likely be explained by a selection effect mechanism (Fearon 2002). Countries originating a dispute can estimate, using more factors than can be accounted for in the model, the likelihood that they will be joined in the dispute. Similarly, because a dispute requires two sides, an opponent can estimate the probability that it will face multiple adversaries. If a country believes it is very likely to face multiple adversaries (and thus be more outmatched), it has an incentive to back down before a dispute begins. The disputes we do observe should be the ones in which either one party has a particularly strong attachment to the object of the dispute or in which the observable factors included in the model fail to convey an accurate assessment of the likelihood of facing multiple opponents. Thus, for credible alliances, disputes are unlikely to occur, but alliances that exist solely on paper can act as a mirage. Likewise, when one expects to be joined in a dispute because one has a strong alliance, one's opponent can observe this and will back down in advance. This should bias the estimates down from what the actual effect would be.

Interestingly, the same logic that mitigates against finding an effect for alliances should also function for joint military exercises. The stronger the signal sent by a joint exercise, the less likely that a possible opponent will be willing to engage in a dispute rather than backing down. If the strongest, most credible alliances are the ones that conduct joint military exercises, then the presence of both exercises and an alliance should enhance this selection effect logic, thus explaining the negative coefficient on the interaction between exercises and alliances. That there is still a positive and significant direct effect of joint military exercises suggests they do carry important information about which states are likely to join disputes together.

Table 5 shows the results looking at arms trades. SIPRI reports a proprietary measure of the value of weapons (excluding small arms) sold between countries ("SIPRI Arms Transfers Database" 2020). I use the log of the sum of the total of the value of arms deliveries between the pair of countries in the dyad. The measure is continuous, so I use ordinary least squares. I include as a control a one-year lag of the dependent variable to account for the propensity of a dyad to continue trading arms. The results again are in line with expectations. In most specifications, conducting joint military exercises is associated with a statistically significant increase in the value of arms traded within a dyad. The two statistically insignificant results are when I use only non-combat exercises, which suggests that perhaps countries focused on such activities are not as interested in the kinetic side of military might that relates to weapons sales. Because the models include the lag of the dependent variable, they suggest that arms sales grow after conducting exercises.

Unlike with militarized interstate disputes, here the conventional thinking on alliances is borne out, with the direct effect of being in an alliance being positive and statistically significant. While statistically significant, the direct effect of an alliance is much smaller than that of conducting joint military exercises in most specifications. The interaction effect of joint military exercises and an alliance is positive in all cases and statistically significant in most, which suggests that the alliances that also exercise are focused the most on interoperability both at the operational level and the equipment level. The magnitude of the interaction effect is generally quite large relative to the direct effects.

The three sets of results indicate that the joint military exercise variable captures a useful measure of security cooperation. It is associated with a decreased risk of militarized disputes, in line with expectations that countries that anticipate engaging in conflict with each other ought to be unlikely to practice fighting together. It is associated with a statistically significant increase in the probability of joining a dispute on the side of a preexisting exercise partner. In most cases, exercises are associated with statistically significant increases in arms sales in a dyad. Together, these results suggest that joint exercises are a useful method of

Table 5: JMEs and Arms Sales

	Dependent variable:					
	log(Arms Sales)					
	(1)	(2)	(3)	(4)	(5)	(6)
Binary JME	0.029***					
	(0.006)					
Binary Combat FTX JME		0.106***				
		(0.016)				
Binary Non-Combat JME			−0.002			
			(0.005)			
Count JME				0.016***		
				(0.003)		
Count Combat FTX JME					0.074***	
					(0.011)	
Count Non-Combat JME						0.003
						(0.002)
Alliance	0.016***	0.014***	0.032***	0.009*	0.015***	0.023***
	(0.004)	(0.004)	(0.006)	(0.005)	(0.005)	(0.006)
JME*Alliance	0.064***	0.069**	0.044***	0.020***	0.003	0.037***
	(0.016)	(0.029)	(0.011)	(0.005)	(0.014)	(0.007)
log(Shared IGO Memberships)	−0.017***	−0.016***	−0.015***	−0.017***	−0.016***	−0.015***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
log(Dyadic Trade)	0.010***	0.009***	0.011***	0.009***	0.009***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Joint Democracy	−0.003	−0.004*	0.001	−0.006**	−0.004	−0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Joint Non-Democracy	−0.018***	−0.017***	−0.018***	−0.017***	−0.017***	−0.018***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Balance of Power	0.002	0.002	−0.0001	0.003	0.005	0.0004
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
log(Power (Smaller))	0.010***	0.010***	0.009***	0.010***	0.010***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Border	0.0004	−0.004	0.002	−0.005	−0.008	0.0004
	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)
Min. Distance	0.001***	0.001***	0.001***	0.001***	0.0004*	0.001***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Major Power	0.099***	0.097***	0.100***	0.096***	0.094***	0.099***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
log(Arm Sales$_{t-1}$)	0.717***	0.713***	0.719***	0.710***	0.705***	0.717***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Constant	0.119***	0.116***	0.106***	0.120***	0.118***	0.109***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Observations	361,738	361,738	361,738	361,738	361,738	361,738
R^2	0.585	0.586	0.584	0.587	0.588	0.585
Adjusted R^2	0.585	0.586	0.584	0.587	0.588	0.585
F Statistic	39,208***	39,386***	39,132***	39,485***	39,688***	39,213***

Note: Control variables lagged one year except border and minimum distance. *p<0.1; **p<0.05; ***p<0.01

signalling intentions and increasing combined capabilities.

6 Conclusion

The joint military exercise data set contains information on over 5,000 joint military exercises from 1977-2016. These exercises occurred around the world and nearly every state has participated in at least one exercise. The data contain information on the participating states, the types of activities they practiced, the manner of the training, and, in many cases, the size of the exercise and participating service branches. The richness of the data will allow researchers to probe hypotheses related not just to the presence of an exercise, but to its content.

The joint military exercise data set is the most comprehensive, inclusive, and accurate data set of its kind. The limited amount of previous work that has used joint military exercises has been based on data rife with errors and likely plagued by unfortunate biases. The data introduced here, however, has been reviewed and corrected to avoid these types of problems. It not only extends further than previous data, but it includes many more joint military exercises. Moreover, the collection process appears to have successfully avoided temporal or geographic biases as well.

The joint military exercise data set will help spur further research on security cooperation outside of looking exclusively at formal alliances. The recent expansion of work on security cooperation that uses measures other than formal military alliances is a welcome development. Alliances are a useful measure of state activity and intentions, but, at least in recent years, change infrequently and fail to encompass the full range of foreign policy tools states have at their disposal. Joint military exercises are one of the ways states build cooperative relationships, signal their foreign policy intentions, and develop key military capabilities. Conducting exercises together is associated with a decrease in the likelihood of engaging in a militarized interstate dispute and an increase in the value of arms traded between states.

The number of exercises and the variety of participants has steadily increased over the last 30 years. This pattern of growth, moreover, is not merely replicating existing alliance patterns but instead likely reflects changes to the underlying security environment.

References

Akerman, Anders, and Anna Larsson Seim. 2014. "The global arms trade network 1950-2007." *Journal of Comparative Economics* 42 (3): 535–551.

Allen, Michael A., Julie VanDusky-Allen, and Michael E. Flynn. 2016. "The Localized and Spatial Effects of US Troop Deployments on Host-State Defense Spending." *Foreign Policy Analysis* 12 (4): 674–694.

Atkinson, Carol. 2006. "Constructivist Implications of Material Power: Military Engagement and the Socialization of States, 1972-2000." *International Studies Quarterly* 50 (3): 509–537.

———. 2010. "Does Soft Power Matter? A Comparative Analysis of Student Exchange Programs 1980–2006." *Foreign Policy Analysis* 6 (1): 1–22.

Barbieri, Katherine, and Omar M. G. Keshk. 2016. "Correlates of War Project Trade Data Set Codebook, Version 4.0." Correlates of War Project. Accessed June 25, 2020. http://correlatesofwar.org.

Barbieri, Katherine, Omar M. G. Keshk, and Brian Pollins. 2009. "TRADING DATA: Evaluating our Assumptions and Coding Rules." *Conflict Management and Peace Science* 26 (5): 471=491.

Beckley, Michael. 2015. "The Myth of Entangling Alliances: Reassessing the Security Risks of U.S. Defense Pacts." *International Security* 39 (4): 7–48.

Benson, Brett V. 2012. *Constructing International Security: Alliances, Deterrence, and Moral Hazard.* New York: Cambridge University Press.

Bernhardt, Jordan. 2020. "The Causes and Consequences of Joint Military Exercises." PhD diss., Stanford University.

Bernhardt, Jordan, and Lauren Sukin. 2020. "Joint Military Exercises and Crisis Dynamics on the Korean Peninsula." *Journal of Conflict Resolution.*

Biddle, Stephen, Julia Macdonald, and Ryan Baker. 2018. "Small Footprint, Small Payoff: The Military Effectiveness of Security Force Assistance." *The Journal of Strategic Studies* 41 (1-2): 89–142.

Blackwill, Robert D., and Jeffery W. Legro. 1989. "Constraining Ground Force Exercises of NATO and the Warsaw Pact." *International Security* 114 (3): 68–98.

Blankenship, Brian, and Raymond Kuo. 2017. "Excising Threats or Exhibiting Resolve? The Causes and Consequences of Joint Military Exercises." *Paper presented at Midwest Political Science Association Annual Conference.*

Blanton, Shannon Lindsey. 2000. "Promoting Human Rights and Democracy in the Developing World: U.S. Rhetoric versus U.S. Arms Exports." *American Journal of Political Science* 44 (1): 123–131.

———. 2005. "Foreign Policy in Transition? Human Rights, Democracy, and U.S. Arms Exports." *International Studies Quarterly* 49 (4): 647–667.

Caravelli, John M. 1983. "Soviet and Joint Warsaw Pact Exercises: Functions and Utility." *Armed Forces & Society* 9 (393): 393–426.

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. 2007. "United States Hegemony and the New Economics of Defense." *Security Studies* 16 (4): 598–614.

Collingwood, Loren, and John Wilkerson. 2012. "Tradeoffs in Accuracy and Efficiency in Supervised Learning Methods." *Journal of Information Technology & Politics* 93 (3): 298–318.

Cooley, Alexander. 2008. *Base Politics: Democratic Change and the U.S. Military Overseas.* Ithaca, NY: Cornell University Press.

Correlates of War Project. 2017. "State System Membership List, v2016." Accessed August 8, 2019. http://www.correlatesofwar.org/data-sets/state-system-membership.

Crawford, Timothy W. 2003. *Pivotal Deterrence: Third-Party Statecraft and the Pursuit of Peace.* Ithaca, NY: Cornell University Press.

D'Orazio, Vito. 2012. "War Games: North Korea's Reaction to US and South Korean Military Exercises." *Journal of East Asian Studies* 12 (2): 275–294.

———. 2013. "International Military Cooperation: From Concepts to Constructs." PhD diss., Pennsylvania State University.

De Soysa, Indra, and Paul Midford. 2012. "Enter The Dragon! An Empirical Analysis of Chinese versus US Arms Transfers to Autocrats and Violators of Human Rights, 1989-2006." *International Studies Quarterly* 56 (4): 843–856.

Department of Defense Cost Assessment and Program Evaluation Simulation and Analysis Center. 2010. *Security Force Assistance Skills and Capabilities Study.* Washington, D.C.: GPO.

Embassy of the United States of America Singapore. 2019. "Fact Sheet: Defense Cooperation," July. Accessed February 21, 2020. https://sg.usembassy.gov/wp-content/uploads/sites/197/Fact-Sheet-Defense-Cooperation-small.pdf.

Erickson, Jennifer L. 2013. "Market imperative meets normative power: Human rights and European arms transfer policy." *European Journal of International Relations* 19 (2): 209–234.

Farrell, John F. 2009. "Team Spirit: A Case Study on the Value of Military Exercises as a Show of Force in the Aftermath of Combat Operations." *Air and Space Power Journal* 23 (3): 94–106.

Fearon, James D. 1994. "Signaling versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model." *Journal of Conflict Resolution* 38 (2): 236–269.

———. 1997. "Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs." *Journal of Conflict Resolution* 41 (1): 68–90.

———. 2002. "Selection Effects and Deterrence." *International Interactions* 28 (1): 5–29.

Fearon, James D., and Bertel Hansen. 2017. "The Arms Trade, International Alignments, and International Conflict." *Paper presented at the annual meeting of the APSA Annual Meeting & Exhibition, San Francisco.*

Folke Bernadotte Academy. 2017. "Questions and answers about Viking," November 26, 2017. Accessed November 10, 2020. https://fba.se/en/areas-of-expertise/leadership-cooperation/questions-and-answers-about-viking/.

Government of Australia Department of Defense. n.d. "Australia-Singapore Military Training Initiative." Accessed February 21, 2020. https://www.defence.gov.au/Initiatives/ASMTI/.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as data: The promise and pitfalls of automatic content analysis methods for political texts." *Political analysis* 21 (3): 267–297.

Grinberg, Marc Toby. n.d. "Military Capacity Building and the International Transfer of Military Power." *Working Paper.*

Hall, Todd, and Keren Yarhi-Milo. 2012. "The Personal Touch: Leaders' Impressions, Costly Signaling, and Assessments of Sincerity in International Affairs." *International Studies Quarterly* 56 (3): 560–573.

Headquarters, Department of the Army. 2010. *Army Regulation 220-1: Army Unit Status Reporting and Force-Registration — Consolidated Policies.* Washington, D.C.: Government Printing Office.

———. 2016. *FM 7-0: Train to Win in a Complex World.* Washington, D.C.: Government Printing Office.

———. 2017. *Army Regulation 350-1: Army Training and Leader Development.* Washington, D.C.: Government Printing Office.

———. 2019. *ADP 7-0: Training.* Washington, D.C.: Government Printing Office.

Holmes, Marcus. 2013. "The Force of Face-to-Face Diplomacy: Mirror Neurons and the Problem of Intentions." *International Organization* 67 (4): 829–861.

Hopkins, Daniel J., and Gary King. 2010. "A method of automated nonparametric content analysis for social science." *American Journal of Political Science* 54 (1): 229–247.

Hunzeker, Michael Allen, and Alexander Lanoszka. 2016. "Landpower and American Credibility." *Parameters* 45 (4): 17–26.

Huth, Paul. 1988. *Extended Deterrence and the Prevention of War.* New Haven, CT: Yale University Press.

Huth, Paul, and Bruce Russett. 1984. "What Makes Deterrence Work? Cases from 1900 to 1980." *World Politics* 36 (4): 496–526.

Huth, Paul, and Bruce Russett. 1988. "Deterrence Failure and Crisis Escalation." *International Studies Quarterly* 32 (1): 29–45.

Joachims, Thorsten. 1997. "A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization." In *International Conference on Machine Learning.*

———. 1998. "Text categorization with support vector machines: Learning with many relevant features." In *European Conference on Machine Learning.*

———. 1999. "Transductive inference for text classification using support vector machines." In *International Conference on Machine Learning.*

———. 2001. "A Statistical Learning Model of Text Classification for Support Vector Machines." In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval.*

Johnson, Dave. 2017. "ZAPAD 2017 and Euro-Atlantic security." NATO Review, December 14, 2017. Accessed November 10, 2020. https://www.nato.int/docu/review/articles/2017/12/14/zapad-2017-and-euro-atlantic-security/index.html.

Johnson, Richard A. I., and Spencer L. Willardson. 2018. "Human Rights and Democratic Arms Transfers: Rhetoric Versus Reality with Different Types of Major Weapon Systems." *International Studies Quarterly* 62 (2): 453–464.

Johnson, Richard A.I. 2015. "United States Arms Transfer Decision-Making: Determinants of Sales versus Aid." *Peace Economics, Peace Science and Public Policy* 21 (4): 507–514.

Joint Chiefs of Staff. 2017. *Joint Publication 3-20: Security Cooperation.* Washington, D.C.: Government Printing Office.

Karlin, Mara E. 2018. *Building Militaries in Fragile States: Challenges for the United States.* Philadelphia, PA: University of Pennsylvania Press.

Kim, Tongfi. 2011. "Why Alliances Entangle But Seldom Entrap States." *Security Studies* 20 (3): 350–377.

Kinne, Brandon J. 2018. "Defense Cooperation Agreements and the Emergence of a Global Security Network." *International Organization* 72 (4): 799–837.

———. 2019. "The Defense Cooperation Agreement Dataset (DCAD)." *Journal of Conflict Resolution.*

Kolari, Pranam, Akshay Java, Tim Finin, Tim Oates, and Anupam Joshi. 2006. "Detecting spam blogs: A machine learning approach." In *Proceedings of the national conference on artificial intelligence.* AAAI Press.

Lebovic, James H., and Elizabeth N. Saunders. 2016. "The Diplomatic Core: The Determinants of High-Level US Diplomatic Visits, 1946–2010." *International Studies Quarterly* 60 (1): 107–123.

Leeds, Brett Ashley, Jeffrey M. Ritter, Sara McLaughlin Mitchell, and Andrew G. Long. 2002. "Alliance Treaty Obligations and Provisions, 1815-1944." *International Interactions* 28 (3): 237–260.

Limon, Armando. 2016. "Tiger Balm 2016 Kicks Off on Oahu with Focus on Partnership." U.S. Indo-Pacific Command, July 21, 2016. Accessed November 10, 2020. https://www.pacom.mil/Media/News/News-Article-View/Article/852599/tiger-balm-2016-kicks-off-on-oahu-with-focus-on-partnership/.

Lyle, Amaani. 2014. "DoD Announces Upcoming NATO, Maritime Exercises." U.S. Department of Defense, August 27, 2014. Accessed November 10, 2020. https://www.defense.gov/Explore/News/Article/Article/603139/.

Machain, Carla Martinez, and T. Clifton Morgan. 2013. "The Effect of US Troop Deployment on Host States' Foreign Policy." *Armed Forces & Society* 39 (1): 102–123.

Marshall, Monty G., Ted Robert Gurr, and Keith Jaggers. 2018. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2017. Dataset Users' Manual." Center for Systemic Peace, October 24, 2018. Accessed August 8, 2019. http://www.systemicpeace.org/inscr/p4manualv2017.pdf.

McManus, Roseanne W. 2018. "Making It Personal: The Role of Leader-Specific Signals in Extended Deterrence." *Journal of Politics* 80 (3): 982–995.

McManus, Roseanne W., and Mark David Nieman. 2019. "Identifying the level of major power support signaled for proteges: A latent measure approach." *Journal of Peace Research* 56 (3): 364–378.

McManus, Roseanne W., and Keren Yarhi-Milo. 2017. "The Logic of "Offstage" Signaling: Domestic Politics, Regime Type, and Major Power-Protégé Relations." *International Organization* 71 (4): 701–733.

McNerney, Michael, Angela O'Mahony, Thomas S. Szayna, Derek Eaton, Caroline Baxter, Colin Clarke, Emma Davies, et al. 2014. *Assessing Security Cooperation as a Preventive Tool, RR-350-A*. Santa Monica, CA: RAND.

Ministry of Defence of Georgia. 2015. "International Training Eternity 2014 Completed," May 2, 2015. Accessed November 10, 2020. https://mod.gov.ge/en/news/read/3206/international-training-%E2%80%9Ceternity-2014%E2%80%9D-completed.

Morrow, James D. 1994. "Alliances, Credibility, and Peacetime Costs." *Journal of Conflict Resolution* 38 (2): 270–297.

———. 2000. "Alliances: Why Write Them Down?" *Annual Review of Political Science* 3:63–83.

Palmer, Glenn, Vito D'Orazio, Michael Kenwick, and Matthew Lane. 2015. "The MID4 dataset, 2002–2010: Procedures, coding rules and description." *Conflict Management and Peace Science* 32 (2): 222–242.

Paul, Christopher, Colin P. Clarke, Beth Grill, Stephanie Young, Jennifer D. P. Moroney, Joe Hogler, and Christine Leah. 2013. *What Works Best When Building Partner Capacity and under What Circumstances?, MG-1253/1-OSD.* Santa Monica, CA: RAND.

Perkins, Richard, and Eric Neumayer. 2010. "The organized hypocrisy of ethical foreign policy: Human rights, democracy and Western arms sales." *Geoforum* 41 (2): 247–256.

Pevehouse, Jon C. W., Timothy Nordstron, Roseanne W. McManus, and Anne Spencer Jamison. 2020. "Tracking Organizations in the World: The Correlates of War IGO Version 3.0 datasets." *Journal of Peace Research* 57 (3): 492–503.

Platte, Hendrik, and Dirk Leuffen. 2016. "German Arms Exports: Between Normative Aspirations and Political Reality." *German Politics* 25 (4): 561–580.

Porter, Martin F. 1980. "An algorithm for suffix stripping." *Program* 14 (3): 130–137.

Quinlivan, James T., and Olga Oliker. 2011. *Nuclear Deterrence in Europe: Russian Approaches to a New Environment and Implications for the United States.* Santa Monica, CA: RAND Corporation.

Savage, Jesse Dillon, and Jonathan D Caverley. 2017. "When human capital threatens the Capitol: Foreign aid in the form of military training and coups." *Journal of Peace Research* 54 (4): 542–557.

Schelling, Thomas C. 1966. *Arms and Influence.* New Haven, CT: Yale University Press.

Schroedt, Philip. n.d. "`CountryInfo-1`." GitHub. Accessed February 23, 2020. https://github.com/philip-schrodt/CountryInfo-1.

Selden, Zachary. 2013. "Balancing Against or Balancing With? The Spectrum of Alignment and the Endurance of American Hegemony." *Security Studies* 22 (2): 330–364.

Singer, J. David, Stuart Bremer, and John Stuckey. 1972. "Capability Distribution, Uncertainty, and Major Power War, 1820-1965." In *Peace, War, and Numbers,* edited by Bruce Russett. Beverly Hills: Sage.

"SIPRI Arms Transfers Database." 2020. Stockholm International Peace Research Institute. Accessed June 25, 2020. https://www.sipri.org/databases/armstransfers.

Sislin, John. 1994. "Arms as Influence: The Determinants of Successful Influence." *Journal of Conflict Resolution* 38 (4): 665–689.

Slantchev, Branislav L. 2005. "Military Coercion in Interstate Crises." *American Political Science Review* 99 (4): 533–547.

Smith, Alastair. 1995. "Alliance Formation and War." *International Studies Quarterly* 39 (4): 405–425.

———. 1998. "Extended deterrence and alliance formation." *International Interactions* 24 (4): 315–343.

Snyder, Glenn H. 1961. *Deterrence and Defense.* Princeton, NJ: Princeton University Press.

———. 1997. *Alliance Politics.* Ithaca, NY: Cornell University Press.

Stinnett, Douglas M., Jaroslav Tir, Philip Schafer, Paul F. Diehl, and Charles Gochman. 2002. "The Correlates of War Project Direct Contiguity Data, Version 3." *Conflict Management and Peace Science* 19 (2): 58–66.

The Baltic Times. 2018. "Members of Baltic air forces to practice rescuing at sea in Estonia," August 1, 2018. Accessed November 10, 2020. https://www.baltictimes.com/members_of_baltic_air_forces_to_practice_rescuing_at_sea_in_estonia/.

Watts, Stephen, Caroline Baxter, Molly Dunigan, and Christopher Rizzi. 2012. *The Uses and Limits of Small-Scale Military Interventions, MG-1226-RC.* Santa Monica, CA: RAND.

Weidmann, Nils B., and Kristian Skrede Gleditsch. 2010. "Mapping and Measuring Country Shapes: The `cshapes` Package." *R Journal* 2 (1): 18–24.

Weitsman, Patricia A. 2013. *Waging War: Alliances, Coalitions, and Institutions of Interstate Violence.* Stanford, CA: Stanford University Press.

Wolfley, Kyle J. 2018. "Training Not to Fight: How Major Powers Use Multinational Military Exercises to Manage Strategic Uncertainty." PhD diss., Cornell University.

———. 2021. "Military Statecraft and the Use of Multinational Exercises in World Politics." *Foreign Policy Analysis* 17 (2).

Yung, Christopher D. 2006. *Gators of Neptune: Naval Amphibious Planning for the Normandy Invasion.* Annapolis, MD: Naval Institute Press.