

A Winning Proposition? States' Military Effectiveness and the Reliability of Their Allies

Stephen B. Long

Department of Political Science
University of Richmond
slong@richmond.edu

This paper explores the relationship between states' past military performance in interstate wars and the likelihood that their allies will come to their aid when the terms of their alliance require intervention. Following Leeds (Leeds 2003), I argue that decisions to honor or violate alliances depend on changes that have occurred since the signing of the alliance and various factors that affect the expected costs of honoring the alliance. What I add to this approach is an exploration of whether and how a state's performance in prior interstate wars sends signals about its likely military effectiveness in the current war. I argue that superior records of military performance signal to allies that their potential costs for intervention will be lower than they would be with a less effective ally. I use an outcomes-based measure of military effectiveness to predict when states will violate or honor their alliances.

Introduction

Why do states choose to violate formal interstate alliances that they once found useful enough to sign? Alliances involve costs for all members and thus are not likely to be entered into lightly, and yet the existence of alliance violations in world history is indisputable. Formal theoretical work on alliances by Smith (1995, 1996), Morrow (1994), and Fearon (1997) offers models that expect no violation of alliances or very infrequent violation of alliances, even though empirical work has established that formal alliances are violated approximately 25% of the time (Leeds et al 2002).

This mismatch between formal theoretical expectations and the observed frequency of alliance violation may be due to sampling, as suggested by Smith (1995, p. 418), or may be due to factors that change over the course of the alliance (Leeds 2003). It is the latter explanation that is the focus of this paper. Specifically, I adopt the theoretical framework offered by Leeds (Leeds 2003), which separates causes of violation into factors that change over the course of the alliance and factors that influence the costs of violation. Extending the costs side of the theoretical argument, I argue that states consider the potential costs of violating or honoring their alliances and that one way in which they calculate those costs is by referring to the record of military effectiveness of the ally in need of assistance. I test this hypothesis by employing a measure of past military effectiveness among a field of controls derived from the existing empirical literature in a logit analysis of alliance violation.

States deciding whether to honor or violate their alliances when the terms are in effect are standing at the brink of war. Their choice has the potential to change their territorial holdings, wealth, population and productive capacity. In other words, unlike choices to honor or violate many other international agreements, alliance choices have the potential to permanently alter states' security and relative power in the international system. I argue that given the gravity of their decision, states are not driven solely by incentives for violation, but also seek out all relevant information that could help them calculate the likely costs of honoring their alliance. While much of this effort to obtain information takes place through clandestine intelligence collection and thus is beyond the reach of measurement by social scientists, there is also public information that can inform states' decisions. One of the richest sources of information allied potential interveners have about likely costs of intervention is their ally's historical record of military effectiveness in warfare, the major elements of which are observable and known to other states.

Explaining Alliance Reliability

Scholarship on the factors affecting alliance reliability has followed two related paths. The first path employs formal modeling to describe the nature of states' strategic interactions surrounding formal alliances and decisions to honor or violate those alliances, finding formally that alliance violation should occur infrequently or not at all because of the role of alliances as signals. The second path has tackled the empirical existence of moderate levels of violation, exploring the impact of a set of

explanatory factors like regime change, regime type, and changes in the balance of power after signing.

Leeds (2003) sets up the mismatch between formal work on alliances and empirical findings on alliance violation. Morrow (1994) describes alliances as costly signals of common interests among allies. In his model, states with common interests form alliances to signal those interests in the face of uncertainty about whether they would intervene on one another's behalf, and these alliances are expected to be credible because of their significant peacetime costs (Morrow 1994, 272-273).

Smith (1995) offers a model that connects alliance reliability, alliance formation, and the effect of alliances on the occurrence of war. In this model, the expected reliability of the B-C alliance is crucial both for state A's decision to attack and state B's decision to retaliate. Smith's results suggest that a nation is generally more likely to intervene if it has an alliance with one of the others, but also that almost all alliances should be reliable because states will only form alliances that benefit them (Smith 1995, 416). Smith recognizes the mismatch between this prediction and empirical reality and argues that the cause is sampling bias (those alliances that are attacked are those that state A has deemed to be unreliable, inflating the proportion of unreliable alliances that we observe). Fearon (1997), sees alliances as tied hands signals, and as such, he expects them not be subject to bluffing. This is because the availability of high-cost signals (reliable alliances) as a tool makes any partial commitment a non-credible signal by comparison.

This set of formal models lays out a logical foundation that recognizes the strategic nature of the decision to violate or honor alliances, one that is largely supported by the history of what is perhaps the most important interstate alliance of the modern era, the North Atlantic Treaty Organization (NATO). Formed in the face of increasing Soviet power and the militarization of the Cold War, NATO created a credible transatlantic deterrent against Soviet incursion into Western Europe and served as a signal of the unity of American and European understandings of the primary threats to their security. It was precisely the high peacetime costs and strategic integration of national military forces in NATO that made the alliance credible, and as predicted by Morrow, the alliance proved highly successful.

However, the broader history of alliances is not so positive as the case of NATO might suggest. In fact, 25% of alliances are violated when their terms are in effect (Leeds 2003, 802-803)? Smith and Fearon offer competing responses to this mismatch between the expectation of very high reliability of alliances that are formed as credible signals of intent and the empirical record of moderate levels of violation. Smith focuses on the problem of sampling bias, as noted above. Fearon focuses on the possibilities that tying hands may provoke other states, that the need for domestic support may lead to partial commitment, that changes in incentives to fight may take place after the commitment, and that providing a strong commitment may embolden the ally to the point of being more willing to fight (Fearon 1997, 84-85).¹ While the theoretical framework and empirical findings offered by Leeds (2003) respond to Fearon's speculation, another branch of work has developed with

the goal of explaining the violations that we do observe, only partly responding to formal work on the topic.

Empirical work on alliance duration and reliability has developed partly alongside the formal models described above and partly in response to the questions they raise. Much of the focus of work predicting alliance duration, as opposed to alliance violation by one side, has been on the effect of democracy. Consistent with work positing that democracies behave differently than other states in their international commitments (Russett 1993, Fearon 1994, Slaughter 1995, Simmons 1999, Mansfield et al 2002, Mitchell 2002), Gaubatz (1996) argues that several theoretical bases exist for believing that democracies will be distinctive in their ability to uphold commitments. In addition to prior arguments about audience costs, vertical enforcement, and norms, Gaubatz argues that democracies are stable in their preference orderings, have durable foreign policy bureaucracies, and suffer less from dramatic fluctuations in policy after leadership changes. Leadership changes are one potential source of alliance violation pointed out by Morrow (1991) and later tested by Siverson and Starr (1994). Also, economic interdependence between democracies and isolationist tendencies may lead democracies to select their commitments more carefully or avoid situations that make commitment breaking appealing. To test his hypotheses, he uses Doyle's liberal democracy measure (Doyle 1983) and the COW alliance data (Singer and Small, 1966) in a survival analysis that finds that democracies are no more or less likely to form alliances than other states, but alliances between democracies do tend to last substantially longer than mixed or nondemocratic alliances.²

Contradicting Gaubatz's argument, Gartzke and Gleditsch (2004) argue that democracies will actually be less reliable alliance partners than non-democracies. The authors argue that states use alliances when they have problems honoring informal agreements, essentially tying themselves together for fear that they would otherwise separate as circumstances change. This means that alliances should be seen as an indication of commitment problems, not a willingness to commit. This is especially the case for democracies because of the instability of the democratic principal-agent relationship, specifically due to cycling and the influence of interest groups (Gartzke and Gleditsch 2004, 781-782). Predicting the occurrence of intervention (rather than duration of the alliance), the authors find that democracies are, in fact, less reliable allies than non-democracies.

While the empirical literature described here does not purposefully respond to formal theories by Morrow, Smith, and Fearon, it does investigate a set of causal factors that speak to the puzzle that comes out of that formal work. As Fearon and Smith speculate, changes that occur following the signing of an alliance may make violation more appealing. Arguments on regime change (Morrow 1991, Siverson and Starr 1994) and changes in the balance of power (Morrow 1991, Bennett 1997) are clearly consistent with this speculation, since both types of change could easily alter the incentives that states have to seek renegotiation or outright termination of their alliances through violation. Even in the highly successful case of NATO, French vacillation can be seen as evidence of the importance of post-ratification changes in incentives.

Leeds (2003) provides a logical framework that clarifies the connection between the causes of violation established in prior empirical work and the formal arguments made by Morrow, Smith, and Fearon. She argues that the fact that alliances are violated in approximately 25% of the instances in which their terms are in effect generally supports the notion that alliances are reliable agreements, but that the remaining violations are still in need of explanation. She argues that violations can be explained by reference to two forces: bluffing and changes after signing. Bluffing may occur at the time of signing when states expect the costs of violation to be low, meaning that some alliances are at greater risk of violation than others from the beginning. So, factors that lower the costs of violation increase the probability of states forming risky alliances (Leeds 2003, 809-810). On the other hand, as Fearon (1997) suggests, changes may occur after the signing of an alliance that create new incentives for violation. Since these changes are also visible to other states, these unreliable alliances may also be more frequently targeted for challenges (Leeds 2003, 810).

Together, these two arguments help organize and connect existing lines of empirical testing. Since regime change may alter incentives for violation after signing, regime type may affect costs of violation, and changes in power may alter incentives after signing, all of these previously examined factors fit nicely into the logical framework offered by Leeds (Leeds 2003, 813, 816).

Changes, Costs, and Prior Military Effectiveness

While Leeds provides a useful logical framework, she does not explore an extension of that framework that could improve our ability to explain states' decisions to

honor or violate their formal alliances when the terms are in effect. In discussing major and minor powers, Leeds notes that 'If the leader of a minor power knows that he or she would be joining a losing effort, avoiding the conflict might be particularly appealing' (Leeds 2003, 814). The notion that weak states do not wish to join losing coalitions is not only intuitive, but part of a broader argument that should apply to all potential interveners. While there is some randomness in the outcomes of all wars, some states perform better in warfare over time than do other states, and records of military effectiveness that states develop as they fight wars with other states are a source of easily observable information about which coalitions are likely to win or lose wars.

States' records of military effectiveness are not only observable, but highly credible information about the probability of future victories. Winning or losing wars is certainly not 'cheap talk,' as states invest massive sunk costs in wars that are paid regardless of the outcome. So, why would states ever form an alliance with a proven loser in the first place? If states were acting rationally, they would choose to ally with states with records of effectiveness to avoid being pulled into a military defeat. I argue that just as capabilities change over time, however, so does the more directly observable measure of effectiveness, actual war outcomes. As we know, CINC scores and other resource-based measures of power do not perfectly predict war outcomes. In reality, war outcomes are affected by a basket of societal, political, and military advantages that states have to varying degrees. These configurations of factors supportive of military effectiveness vary not only between states, but within states over time. These changes alter states' abilities to conduct effective wars

against other states and ultimately win those wars. This means that even if states form alliances with each other's military effectiveness in mind, that effectiveness could easily change after the signing of the alliance. So, while alliances with strong states are desirable, alliances with states that are weak and growing, or alliances with strong states that become unexpectedly weak, are certainly possible.

I argue that states facing a decision to honor or violate an alliance prefer winning propositions, all else being equal. Being on the winning side of a war brings with it territorial gains or other spoils that constitute payoffs for the victors at the expense of the defeated. These payoffs should be tempting for all alliance members when an opportunity for war arises, but since states can also pay tremendous costs during war, states will be more likely to honor their alliances when the target state has a superior record of military effectiveness than when the target state has an inferior records of military effectiveness. This leads to the key hypothesis to be tested in this study alongside a set of controls from previous work:

[H1:] States are more likely to violate their alliances if the target state has a weaker record of military effectiveness than if it has a stronger record of military effectiveness.

Thinking about Military Effectiveness

Scholars have been trying for many years to understand the sources of military effectiveness, generally defined as a state's ability to achieve desired military outcomes in times of conflict. Beyond that general definition, the concept of

effectiveness becomes quite complicated. The first issue that could be problematic for any study using the concept of military effectiveness is that it can be observed at the tactical level, the operational level, or the strategic level, and factors that improve effectiveness on one level may not translate directly to another (Brooks 2003, 153). Tactically excellent militaries can still fail to achieve the political goals of the state in a war, and similarly, eventual achievement of political goals in a war (territorial conquest, for instance) may have little to do with being tactically excellent. The Soviet Union certainly won its territorial and political goals in World War II, but few would argue that it showed tactical or operational excellence across the war.

A second issue is that the concept of military effectiveness is closely tied to the concepts of capabilities and state power, and the three are often treated in quantitative research as one and the same. However, military effectiveness differs conceptually and empirically from capabilities as they are commonly defined and measured in international relations research. Capabilities, as they are measured in the Correlates of War project's CINC scores and other variables commonly used in quantitative international relations research, are resource-based and are thus measures of the potential of a state in war than its actual performance in war. Capabilities, when measured as resources, do not necessarily lead to success on the tactical, operational, or strategic levels. As Biddle notes in his broader critique of existing treatments of military power, capabilities as measured by GNP, population, military personnel, military expenditures, or CINC scores, do only a slightly better job at predicting war outcomes than tossing a coin (military personnel actually does

worse) (Biddle 2004, 21). Biddle offers a different framework for understanding real capability in war, arguing that a focus on resources misses other crucial factors that influence the way that states put their resources to use to achieve military effectiveness on the operational level. Specifically, Biddle's focus is on the degree to which states implement the 'modern system' of force employment, a series of tactical and operational advances that first saw widespread implementation late in World War I (see Biddle 2004, ch.3).

In other words, political and territorial victories do not necessarily reflect military effectiveness, and states with greater levels of traditional power indicators (CINC scores, for instance) do not necessarily perform with high military effectiveness. My argument is that states, facing the critical decision to honor or violate their formal alliance with another state when the terms are in effect, rationally seek out available, but credible information about the likely costs of aiding their ally and that past military effectiveness of the ally provides such credible information.

Data

Military Effectiveness Variable

The approach that I take here is a simple one. Since my study focuses on the influence that records of military effectiveness have on allies' assessments of each other in times of crisis, it isn't necessary to represent the full complexity of the process through which states achieve or fail to achieve their military goals on the tactical, operational, and strategic levels. Given that states have different levels of natural resources over time and different opponents, a measure of past military

effectiveness that records generally-defined victories and defeats across time should serve as some indication of the ability of a state to translate its resources and other attributes into military outcomes. Admittedly, this abstracts from operational or tactical-level concepts of military effectiveness to the level of the final war outcome. However, if allies are concerned about their potential costs for intervening in a conflict, effectiveness at this level may be most relevant. Focusing on victory and defeat in past wars rather than operational or tactical success provides a measure of effectiveness that accounts for both resources and the basket of political factors that determine how states employ those resources in war. More importantly, unlike measures based on casualty exchanges, using a simple measure of victories and defeats does not assume that states operate in a complete information environment. While other, more detailed information about military effectiveness comes from battles and wars than recorded here, states' awareness of such information may be limited. Intelligence collection capabilities and other limits force states to operate on imperfect information. Focusing on the broadest outcome of prior conflicts, victory and defeat, reflects what is likely to be of the most interest to potential intervener: the ability of the state to achieve its long-term goal in a war, given the resources it has at the time and against a variety of opponents.

To measure this concept, I rely on data from Reiter and Stam's Win/Lose/Draw coding (1998). For each case in the data set (State A at the time of the war performance opportunity), I refer to the composite list of all prior wars involving State B.³ For each war that ends before the date of the war performance opportunity in the Leeds (2003) data (based on the Correlates of War dates), I note

the outcome code (win/lose/draw). Based on these histories, I create a count of wins for State B up to the point of the war performance opportunity.⁴ Where no combat history exists, I code the observation as missing. This leads to the simple count variable, *B's Prior Wins*.

Dependent Variable: Alliance Violation

I use the population of cases provided by Leeds (2003), which is based on the Alliance Treaty Obligations and Provisions (ATOP) data set (Leeds et al 2002). The version of the data used here contains 143 cases spanning from 1816 to 1944 and is directed, such that each case refers to one member of an alliance at a point in time in which the terms of the alliance are called into effect (what Leeds calls 'war performance opportunities') (Leeds 2003, 811-812). I refer here to the potential violator as State A and the target as State B.

The dependent variable, a state's (State A) decision to violate or honor its alliance given the opportunity, is measured as a dummy variable that takes the value of 1 if the state violates the alliance and 0 if the state honors the alliance, based on the ATOP data and the extension of that coding by Leeds (2003).

Control Variables

I use the five variables employed by Leeds (2003) as controls, which I will briefly describe here. The variable *State A Democratic* is a dummy variable that takes a value of 1 if State A (the potential violator) if the state scores five or higher on the Polity III democracy scale (Jagers and Gurr 1996) and 0 otherwise. This tests the hypothesis that democracies are more constrained by vertical enforcement and other costs of violation than other states (Leeds 2003, 815). The variable *State A*

is Major Power is a dummy variable that takes a value of 1 if State A is a major power according to COW and 0 otherwise, under the hypothesis that major powers are less sensitive to the costs of violation than are other states (Leeds 2003, 815).

The variable *State A Δ in Capabilities* is a dummy variable that takes a value of one if State A experienced a change in its capabilities between the signing of the alliance and the war-performance opportunity, measured as a change of more than 10 percent in its CINC score, and zero otherwise. This tests the hypothesis offered by Morrow and Siverson and Starr noted above (Leeds 2003, 816-817). The variable *State A Regime Change* is a dummy variable that takes a value of one if State A's Polity democracy score changes by two or more points between the signing of the alliance and the war-performance opportunity and zero otherwise. This tests the hypothesis that changes in leadership may also mean changes in priorities and a sense that old obligations are obsolete (Leeds 2003, 816) Finally, the variable *State B Original Target* is a dummy variable that takes a value of one if State B was an original target in the war that led to the war-performance opportunity and zero otherwise, which Leeds sees as capturing what states know about the reliability of alliances they challenge that is not included in the model (Leeds 2003, 818).

Analysis

In my analysis, I present a *logit* model predicting an individual state's decision to honor or violate an alliance of which it is a member at a time at which the terms of the alliance are in effect. Since this is intuitively a dichotomous concept (violation occurs or it does not) and since a measure of the concept in a continuous format does not exist, the *logit* statistical model is an appropriate choice. I control for the

variables described above, as well as clustering on ATOP alliance number to adjust for likely autocorrelation among observations.

One might argue that the decision to honor or violate an alliance is just one step on a longer path of interconnected decisions (alliance formation being an obvious earlier step). If the factors that influence alliance violation decisions are also among the factors that influence decision to form an alliance in the first place, then a two-stage model might be a better option. While I cannot rule this out, I do not adopt a two-stage model here for two reasons. First, I do not have specific expectations as to which of the independent variables in this analysis may have a relationship with earlier stages (like alliance formation). Second, as Leeds argues, while a two-stage model may be a bit more realistic than a single-stage model of reliability, the underlying reality clearly involves more than just the steps of forming an alliance and then deciding whether to honor or violate that alliance. Instead, the opportunity to honor or violate the alliance (as measured here) can't exist without there first being a militarized interstate dispute, followed by escalation to war, so what would be needed would be a four-stage model (Leeds 2003, 818). Given that a two-stage model can't portray the whole process leading to the decision to honor or violate an alliance, I choose the more straightforward approach of a single-stage logit analysis.⁵

TABLE 1 HERE

The Hosmer-Lemeshow's goodness of fit test for Model 1 using nine groups yields a Chi-squared of 8.56 and Prob>Chi-squared=0.286. Overall model fit with this p-value appears good, given the assumption of nine groups.⁶

The logit analysis for Model 1 yields several interesting results. The key independent variable measuring State B's past wins is negative as expected and statistically significant at the 0.05 threshold ($p=0.020$). This statistical significance level is not sensitive to the treatment of missing values. If I do not drop observations without combat histories, instead imputing zeros for the 17 observations with missing values and adding a dummy coded one for the same cases, the p-value changes slightly, but still remains under the 0.05 threshold ($p=0.033$).⁷ These results support H1, as states with histories of winning their prior wars are less likely to see their alliances violated by their alliance partners when the terms of the alliance are called into effect by events on the ground.

The control variables largely replicate the findings reported by Leeds (2003). States are less likely to violate an alliance when they are democratic than when they are not. This suggests that democracies are in fact more subject to restraints on alliance violation (regardless of the regime type of the ally) than are other states. States are more likely, however, to violate their alliances when they experience a change in their capabilities (positive or negative) between the signing of the alliance and the war-performance opportunity or when they experience a change in leadership during that time. State B being an original target in the war and State A's major power status are statistically insignificant predictors of violation.

When taken together, these results suggest that democracy works against incentives to violate alliances and changes in power and regime create incentives to violate, but given all of those competing forces, a state's history of winning its wars does affect its chances to retain alliance reliability. Specifically, a state's history of

military victories in past wars increases the chance that its allies will honor their commitments to it, given the operationalization of military effectiveness and other variables used here.

While interpreting signs and significance is useful and gives us a general sense of which factors are important predictors of alliance violation and which are not, a comparison of the effects of changes in the values of these variables on the predicted probability of violation offers a more straightforward interpretation that allows us to judge the comparative magnitude of their effects.⁸

Table 2 presents the effect on the predicted probability of violation of moving each significant variable from its minimum to its maximum while holding all other variables at their means. As the table shows, the strongest effect on the predicted probability of violation comes from changes in the capabilities of the potential violator since the time of signing, followed in magnitude by regime change since the time of signing. Surprisingly, however, democracy is the least influential, with the prior military effectiveness of the ally seeking assistance taking a more prominent role than democracy. So, while the significance of the finding on military effectiveness is moderate and its robustness against changes in operationalization is poor (see below), at least this measure of past effectiveness has a notable effect on the probability of violation that is stronger than the effect of the potential violator's regime type.

TABLE 2 HERE

Conclusions

States deciding whether to assist an ally when the terms of their alliance come into effect are essentially standing at the brink of war. Their decision to become involved or to violate their alliance has the potential to permanently alter their position in the international system, making it one of the most crucial decisions that states can make in international politics. I have argued here that given the gravity of their choice, states actively seek out credible information about the likely effects of their decision, and that historical information about their ally's military performance is both readily available and highly credible. Extending the framework offered by Leeds (2003), I argue that states with histories of superior military performance against various opponents across time demonstrate the ability to use whatever resources they have at the time to best fight whatever opponent they face, even though those resources and opponents change over time. States with superior records of military performance are offering their allies a winning proposition, an opportunity to gain the spoils of war while contributing relatively little to the war effort.

I test this hypothesis among a field of controls, finding support for the hypothesis and confirming much of what previous research has found for the control variables. I find that a simple count of an ally's past military victories is associated with a higher probability that its allies will honor their obligations when war breaks out. Not only is the finding moderately significant, but the magnitude of its effect rivals that of regime type. However, this finding should be viewed with caution for several reasons. First, the measure used here to represent military effectiveness is simple, and the results vary with the other operationalizations

tested. To some extent, the value of the results depends on our assumptions about how much information states have about each other's military engagements. It could be, as tested here, that it is only the most general sense of accumulated victories that other states could be expected to observe with any accuracy, given the scope of history covered and the geographical distance between conflicts. While it's reasonable to assume that other states observed that Russia engaged in a war with Japan and that Japan won, it may not be reasonable to assume knowledge of more detailed information about Russia's effectiveness on, say, the tactical level. However, if we assume that states have a more nuanced way of acquiring and processing information about each other's war performance, more complex measures may be in order.

While the findings presented here are not conclusive, they suggest that further research is warranted on the potential connection between states' histories of military performance and their success in recruiting and retaining allies.

References

Bennett, D. Scott. (1997) Testing alternative models of alliance duration, 1816-1984.

American Journal of Political Science 41 (3): 846-878.

Bennett, D. Scott and Allan C. Stam III. (1998) The declining advantages of democracy: A combined model of war outcomes and duration. *Journal of Conflict Resolution* 42 (3): 344-366.

Brandt, Patrick T. and Christina J. Schneider. (2004) What's the size? Does it have any power? Questions about hypothesis tests in selection models. Paper presented at the Midwest Political Science Association Annual Meeting; 15 April-18 April, Chicago, United States.

Brooks, Risa A. (2003) Making military might: Why do states fail and succeed? A review essay. *International Security* 28 (2): 149-191.

Doyle, Michael W. (1983) Kant, liberal legacies, and foreign affairs. Part 1. *Philosophy and Public Affairs* 12: 205-235.

Fearon, James D. (1994) Domestic political audiences and the escalation of international disputes. *American Political Science Review* 83 (3): 577-592.

Fearon, James D. (1997) Signaling foreign policy interests: Tying hands versus sinking costs. *Journal of Conflict Resolution* 41 (1): 68-90.

Gartzke, Eric and Kristian Skrede Gleditsch. (2004) Why democracies may actually be less reliable allies. *American Journal of Political Science* 48 (4): 775-795.

Gaubatz, Kurt Taylor. (1996) Democratic states and commitment in international relations. *International Organization* 50 (1): 109-139.

Jagers, Keith and Ted Robert Gurr. (1996) POLITY III: REGIME CHANGE AND POLITICAL AUTHORITY, 1800-1994. Computer File. 2nd ICPSR version. Boulder, CO: Keith Jagers/College Park, MD: Ted Robert Gurr [producers], 1995. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].

Leeds, Brett Ashley. (2003) Alliance reliability in times of war: Explaining state decisions to violate treaties. *International Organization* 57: 801-827.

Leeds, Brett Ashley, Jeffrey M. Ritter, Sarah McLaughlin Mitchell and Andrew G. Long. (2002) Alliance treaty obligations and provisions, 1815-1944. *International Interactions* 28: 261-284.

Long, J. Scott and Jeremy Freese. (2003) *Regression Models for Categorical Dependent Variables Using Stata*, Revised Edition. College Station, TX: Stata Press.

Mansfield, Edward D., Helen V. Milner and B. Peter Rosendorff. (2002) Why democracies cooperate more: Electoral control and international trade agreements. *International Organization* 56 (3): 477-513.

Mitchell, Sarah McLaughlin. (2002) A Kantian system? Democracy and third-party conflict resolution. *American Journal of Political Science* 46 (4): 749-759.

Morrow, James D. (1991) Alliances and asymmetry: An alternative to the capability aggregation model of alliances. *American Journal of Political Science* 35: 904-933.

Morrow, James D. (1994) Alliances, credibility, and peacetime costs. *Journal of Conflict Resolution* 38 (2): 270-297.

Reed, William. (1997) Alliance duration and democracy: An extension and cross-validation of 'Democratic states and commitment in international relations'. *American Journal of Political Science* 41 (3): 1072-1078.

Reiter, Dan and Allan C. Stam III. (1998) Democracy, war initiation, and victory. *American Journal of Political Science* 92 (2): 377-389.

Russett, Bruce. (1993) *Grasping the Democratic Peace: Principles for a Post-Cold War World*. Princeton, NJ: Princeton University Press.

Simmons, Beth. (1999) See you in 'court?' The appeal to quasi-judicial legal processes in the settlement of territorial disputes. In *A Road Map to War: Territorial Dimensions of International Conflict*, ed. Paul F. Diehl. Nashville, Tennessee: Vanderbilt University Press.

Singer, J. David and Melvin Small. (1966) Formal alliances, 1815--1939: A quantitative description. *Journal of Peace Research* 1: 1-32.

Siverson, Randolph and Harvey Starr. (1994) Regime change and the restructuring of alliances. *American Journal of Political Science* 38: 145-161.

Slaughter, Anne-Marie. (1995) International law in a world of liberal states. *European Journal of International Law* 6 (4): 503-538.

Smith, Alastair. (1995) Alliance formation and war. *International Studies Quarterly* 39 (4): 405-425.

Smith, Alastair. (1996) To intervene or not to intervene: A biased decision. *Journal of Conflict Resolution* 40 (1): 16-40.

Table 1: Logit Model of Alliance Violation

Variable	
<i>B's Prior Wins</i>	-0.527** (0.227)
<i>State A Democratic</i>	-2.697** (1.084)
<i>State A is Major Power</i>	0.520 (0.657)
<i>State A Δ in Capabilities</i>	2.372*** (0.605)
<i>State A Regime Change</i>	1.554*** (0.644)
<i>State B Original Target</i>	0.189 (0.653)
N	126
Log likelihood	-41.990

Standard errors adjusted for clustering by ATOP ID in ().

***p<= .01, **p<=.05, *p<=.1.

Table 2: Effects on the Predicted Probability of Violation

Covariate	Min	Predicted p	Max	Predicted p	% Change
<i>B's Prior Wins</i>	0	0.3765	1	0.0149	-96%
<i>State A Democratic</i>	0	0.1555	1	0.0123	-92%
<i>State A Δ in Capabilities</i>	0	0.0475	1	0.3481	633%
<i>State A Regime Change</i>	0	0.0813	1	0.2953	263%

Predicted probabilities calculated holding other variables at their means.

¹ Fearon questions the apparent mismatch between formal results and empirical findings because of the data available at the time (Fearon 1997, 85-86). However, the results found by Leeds et al (2002) appear to confirm the violation finding using data that match the approach suggested by Fearon.

² Reed (1997) confirms these findings in a study that extends the time period of Gaubatz's study and employs a different statistical model, discrete-time event history analysis.

³ Because of the nature of their study, Reiter and Stam (1998) do not include wars that ended as draws in their data set. To remedy this, I create a composite data set using Reiter and Stam's coding as the master set and incorporating the draws included in Bennett and Stam (1998). Where other differences appear between Reiter and Stam's (1998) coding and Bennett and Stam's (1998) coding, I rely on the former. This includes any differences in combatant lists, win/lose/draw codings, the inclusion or exclusion of non-draw cases, etc.

⁴ Since Reiter and Stam break up World War I and World War II into major periods or campaigns and COW provides only one set of dates for each combatant (instead of dates for each campaign), multiple outcomes for each combatant (as recorded by Reiter and Stam) are counted, but only after the conclusion of the single war as recorded by COW.

⁵ In addition to these issues, other work has suggested that in studies with small sample sizes (as is the case here), common tests of selection are problematic (Brandt and Schneider 2004, 4).

⁶ Specifying five groups yields a p-value of 0.0685 and specifying 15 groups (13 after ties) yields a p-value of 0.4094. There is some variation by group assumption, but two of the three group assumptions still indicate a good fit overall. Variance inflation factors show no problems with multicollinearity (all VIFs are below two). The linktest yields a marginally significant result (p=0.093), suggesting a possibility of model specification error, but I have no theoretical priors for creating interactions to resolve the diagnostic result, and doing so would be atheoretical tweaking of the model to improve statistical fit.

⁷ Comprehensive measures subtracting losses from wins or normalizing wins-losses by the number of wars experienced yield insignificant results, but also impute a

functional form that makes one win equivalent in importance to one loss, a strong assumption. Measures using Loss-Exchange Ratios, the ratio of one side's casualties to the other's, are also statistically insignificant, so there is some sensitivity to variable operationalization.

⁸ Predicted probabilities are calculated using Long and Freese's (2003) SPost commands for Stata.